International Journal of Innovative Research in Computer Science and Technology (IJIRCST)
ISSN (Online): 2347-5552, Volume-12, Issue-4, July 2024
https://doi.org/10.55524/ijircst.2024.12.4.9
Article ID IJIRD-1344, Pages 52-58
www.ijircst.org

Diffusion Dynamics Applied with Novel Methodologies

Anmol Chauhan[1], Sana Rabbani[2], Prof. (Dr.) Devendra Agarwal[3], Dr. Nikhat Akhtar[4], and Dr. Yusuf Perwej[5]

[1]B.Tech Scholar, Department of Information Technology, Goel Institute of Technology & Management, Lucknow, India
[2]Assistant Professor, Department of Information Technology, Goel Institute of Technology & Management, Lucknow, India
[3]Dean (Academics), Goel Institute of Technology & Management, Lucknow, India
[4]Associate Professor, Department of Information Technology, Goel Institute of Technology & Management, Lucknow, India
[5]Professor, Department of Computer Science & Engineering, Goel Institute of Technology & Management, Lucknow, India

Correspondence should be addressed to Dr. Yusuf Perwej; yusufperwej@gmail.com

Received: 9 June 2024 Revised: 25 June 2024 Accepted: 9 July 2024

ABSTRACT- An in-depth analysis of using stable diffusion models to generate images from text is presented in this research article. Improving generative models' capacity to generate high-quality, contextually appropriate images from textual descriptions is the main focus of this study. By utilizing recent advancements in deep learning, namely in the field of diffusion models, we have created a new system that combines visual and linguistic data to generate aesthetically pleasing and coherent images from given text. To achieve a clear representation that matches the provided textual input, our method employs a stable diffusion process that iteratively reduces a noisy image. This approach differs from conventional generative adversarial networks (GANs) in that it produces more accurate images and has a more consistent training procedure. We use a dual encoder mechanism to successfully record both the structural information needed for picture synthesis and the semantic richness of text. outcomes from extensive trials on benchmark datasets show that our model achieves much better outcomes than current state-of-the-art methods in diversity, text-image alignment, and picture quality. In order to verify the model's efficacy, the article delves into the architectural innovations, training schedule, and assessment criteria used. In addition, we explore other uses for our text-to-image production system, such as for making digital art, content development, and assistive devices for the visually impaired. The research lays the groundwork for future work in this dynamic area by highlighting the technical obstacles faced and the solutions developed. Finally, our text-to-image generation model, which is based on stable diffusion, is a huge step forward for generative models in the field that combines computer vision with natural language processing.

KEYWORDS- Diffusion Model, Image Generation, Machine Learning (ML), Text-to-image, Generative Adversarial Networks (GANs).

I. INTRODUCTION

Many people's minds automatically form mental images when they read or listen to stories. Mental images, often known as "seeing with the mind's eye" or visual mental imaging, are essential for many cognitive [1] processes, including memory, reasoning, and thinking.

A significant step toward user intelligence could be the development of technology that can generate visual representations of written descriptions by recognizing the relationship between vision and words. The development of artificial intelligence and deep learning in recent years has enabled tremendous growth in [3] image-processing techniques and computer vision (CV) applications. The conversion of text into images is one such rapidly expanding area. The process of creating aesthetically pleasing images from text inputs is known as text-to-image. One way to caption images is by text-to-image generation, which is the process of creating written descriptions from images. This is also called image-to-text creation [4]. When using text-to-image generation, a human-written description is fed into the model, which then generates an RGB image that corresponds to the description. The remarkable versatility of text-to-image generation has made it a significant topic of research [5]. Common uses for producing photorealistic images from text include photo-editing, industrial design, portrait drawing, photo-searching, art generating, captioning, and image manipulation [6]. The picture synthesis, picture super-resolution, data augmentation, and image-to-image conversion are four areas where [7] generative adversarial networks (GANs) have proven to excel in recent years. Utilizing advanced neural networks, such as Generative Adversarial Networks (GANs) [8] and Autoregressive Transformers, is fundamental to TIG models. These models can read and comprehend text, and then they can make pictures that match the descriptions. This method has the ability to transform content production in many fields because it learns complex patterns, textures, and contextual features from textual clues [9].

II. RELATED WORK

The area of text-to-image synthesis has seen tremendous growth in the past few years as academics have worked to make created graphics more controllable, realistic, and useful. A number of seminal articles covering a range of topics in this field are summarized in this literature review. A new diffusion model called Corgi is proposed by Zhou et al. [10] to improve text-to-image generation. By bringing together picture and text modalities, Corgi hopes to improve the use of CLIP and other pre-trained models. By functioning in supervised, semi-supervised [11], and language-free environments, the model exhibits adaptability. Corgi is a potential improvement in the realm of text-to-image generation, as demonstrated by extensive studies. By implementing annotation at the dataset level, Park et al. [12] tackle difficulties with unpaired image-to-translation systems. To lessen the burden of per-sample domain labeling, the suggested LANIT architecture makes use of candidate textual domain descriptions to identify target domains. While tackling the difficulties of per-sample domain annotation, LANIT's generator framework, rapid learning, and domain regularization loss help to achieve results that are on par with or even better than those of established approaches. Reed et al. [14] used GANs for text-to-image generation in 2016, while Goodfellow [13] introduced them in 2014 first. With the help of training stabilizing approaches, Salimans et al. [15] improved model performance on the MNIST, CIFAR-10, and SVHN datasets. Zia et al. [17] created the attention-based recurrent neural [16] network. An attention-based auto-encoder learnt word-to-pixel dependencies and an autoregressive-based decoder learned pixel-to-pixel dependencies in their model.

The model ReCo, introduced by Yang et al. [18], integrates the best features of text-based and layout-based models to provide accurate region management during text-to-image generation. Users can now submit region-controlled text together with free-form descriptions and location coordinates using ReCo, which enhances pre-trained models to comprehend spatial coordinate inputs. Extensive tests show that ReCo can handle difficult scenarios with better object classification accuracy and detector precision. The year 19 Converting Text into Images with No Training Required In their work on text-to-image [19] synthesis, Mao and Wang present a technique for controlling the position and size of objects with fine-grained precision. In order to alter the position of individual items without extra training, the suggested approach uses diffusion models to manipulate the values of cross-attention layers [20]. To prove that the approach is successful in producing user-aligned generation, object detector-based evaluation metrics measure the efficacy of object-wise location-guided generation. [21] The BATINeT Network: Synthesizing and Manipulating Background-Aware Text to Images The authors Morita et al. present BATINet, a Background-Aware Text to Image synthesis system, to solve problems with text-to-image synthesis. synthesized network [22]. Generating foreground material that harmonizes with a given background is the goal of BATINet. A Position Detect Network, a Generation Network, and a Harmonization [23] Network make up the architecture. Results from extensive testing on the CUB dataset show that BATINet can produce high-quality images that blend in with any background. Dong et al. [24] used an unsupervised approach to train a model that could generate images from text. There was an emphasis on creating programs that could convert text to images by Berrahal et al. [25]. They generated images of faces from descriptions in text using deep fusion GAN (DF-GAN). To improve the semantic fidelity of images generated from textual descriptions, Zhang et al. [26] suggested the cross-domain feature fusion GAN (CF GAN). In order to generate high-resolution images, current text-to-image algorithms utilize a lot of parameters and do a lot of computations, which makes training unreliable and expensive. The direct synthesis of high-resolution images is made possible by Tao et al. [27]'s DF-GAN, a one-stage text-to-image backbone. Enhancements to semantic consistency and efficient text-image fusion are brought about by the new Deep Text-Image Fusion Block (DFBlock) and the Target-Aware Discriminator. DF-GAN overcomes the shortcomings of state-of-the-art text-to-image synthesis models and produces more lifelike and text-consistent images than competing methods [28]. The model's performance is further improved by implementing a one-way output approach in the discriminator and incorporating DFBlock for deep fusion. A convincing technique for producing high-quality images from textual descriptions, DF-GAN has been proven to be superior through extensive testing and benchmarking on hard datasets. One groundbreaking study that uses natural languages to create graphics is Align DRAW [29], however, the outcomes are unrealistic.

III. STABLE DIFFUSION

One kind of generative AI is Stable Diffusion, which takes input in the form of text and images and uses them to generate original, photorealistic graphics. Its initial release year was 2022. The model is not limited to just photographs; it can also be used to make animations and videos [30]. Diffusion technology and latent space form the basis of the model. This makes the model much more efficient, allowing it to be run on GPU-equipped computers or laptops. With just five photos, you may use transfer learning to fine-tune Stable Diffusion [31] to your individual needs. Several fields, including as physics, economics, and computer science, contributed to the concept of steady diffusion. It explains the gradual stabilization of a system as a result of information or change propagation. Computer scientists and software engineers can put stable diffusion to use in state management applications. Part of this is ensuring that different parts of the application reliably and consistently propagate changes to the application state. Alterations to data in one part of the system will be suitably and reliably reflected in other parts of the system, thanks to constant diffusion. The importance and accessibility of Stable Diffusion cannot be overstated. Graphics cards designed for home use are capable of running it [32]. The model is now freely available for download, making it possible for anybody to create their own images. We may also adjust critical hyperparameters like the amount of noise and the number of denoising steps [33].

A. Stable Diffusion Architecture

Stability AI developed the Stable Diffusion model, a state-of-the-art technique for converting text to images. In order to generate high-quality images from textual descriptions, it employs diffusion models, [34] a type of generative model that refines noisy images iteratively. The idea of Latent Diffusion is the basis for Stable Diffusion, as seen in figure 1. For more information on this topic, see the publication High-Resolution Image Synthesis using Latent Diffusion Models [35]. Because of its adaptability, Stable Diffusion has many potential applications. Multiple parts and models come together to form Stable Diffusion. There isn't just one solid model. Stable Diffusion relies on this ingredient. Much of the improvement in performance over earlier models occurs there. In order to create picture data, this part passes through a series of procedures. The steps parameter, typically set to 50 or 100 by default in Stable Diffusion interfaces and libraries, is this. It is in the picture information space, sometimes called latent space, that the image information producer operates entirely [36]. We'll get deeper into the implications of that later on in the piece. Compared to earlier diffusion models that operated in pixel space, this feature makes it much faster. A scheduling algorithm and a UNet neural network constitute this part, to put it technically. Dispersion best characterizes the process at work here. A high-quality image is ultimately produced (by the subsequent component, the image decoder) by the sequential processing of data.

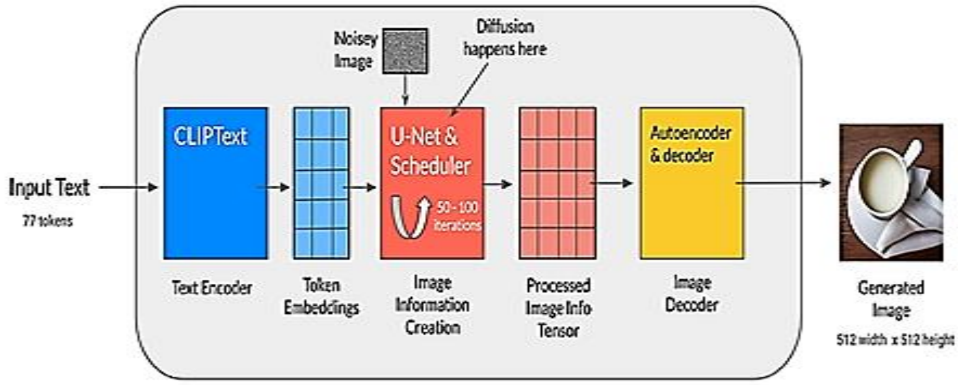


Figure 1: The Stable Diffusion Architecture

The picture decoder uses the data it received from the data producer to create an image. When the procedure is complete, it runs just once to generate the final pixel image. Denoising an object (such as a picture) allows generative models of the diffusion variety to be trained to extract an interesting sample. The model is trained to progressively remove noise from the image until a sample is collected. You can see this process in Figure 2.

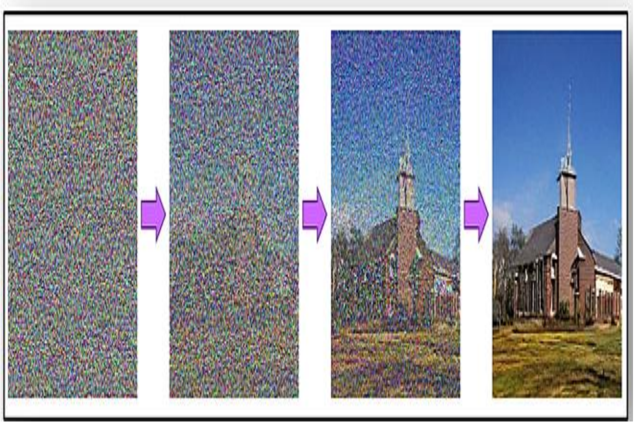


Figure 2: The Denoising Process for Images

These diffusion models have been more popular in recent years, mostly because of the state-of-the-art picture data they generate. However, using diffusion models can be memory-and CPU-intensive.

B. Latent Diffusion Components

When dealing with stable diffusion, one type of diffusion model is latent diffusion. The principle behind it is diffusion, wherein an object (such an image) is denoised and a model is trained to extract a desired sample. In order to reduce complexity and memory usage, latent diffusion extends the diffusion process to a lower dimensional latent space. The Variational Autoencoder (VAE) paradigm consists of an encoder and a decoder [37]. Figure 3 shows the U-Net model in action, with the encoder taking an image as input and the decoder transforming it back into a latent representation. The encoder then uses the picture to construct a low-dimensional latent representation.

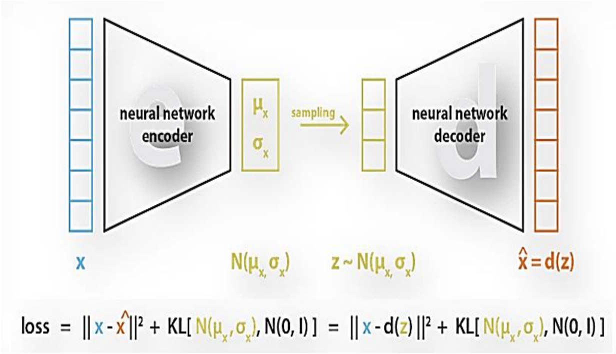

neural network encoder
μ_x
σ_x
sampling
neural network decoder
x
$N(\mu_x, \sigma_x)$
$z \sim N(\mu_x, \sigma_x)$
$\hat{x} = d(z)$
$\text{loss} = ||x - \hat{x}||^2 + KL[N(\mu_x, \sigma_x), N(0, I)] = ||x - d(z)||^2 + KL[N(\mu_x, \sigma_x), N(0, I)]$

Figure 3: The Architecture of the Variational Autoencoder

U-Net is a well-liked convolutional neural network (CNN) that is used for picture segmentation. It also has a decoder and two ResNet block encoders. As seen in Figure 4, an encoder compresses an image [38] into a lower quality image. The decoder, which is supposed to be less noisy, then converts this lower resolution image back to the original, higher resolution image.

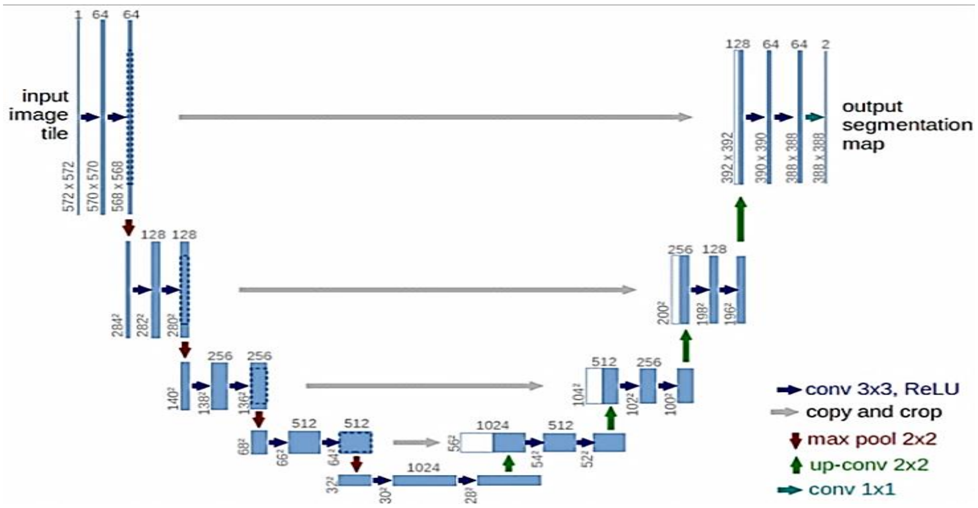


Figure 4: The U-Net Architecture

The text encoder's job is to transform the input text into an embedding space that U-Net can understand. To convert the sequence of input tokens into the latent text embeddings displayed in figure 5, a simple encoder based on transformers is usually employed [39].

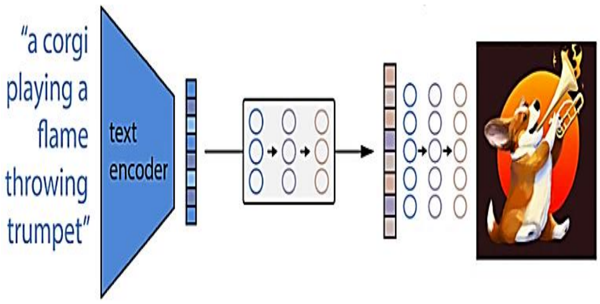


Figure 5: The Working of Text Encoder

IV. SUGGESTED APPROACH

Processing of text prompts, production of associated images, and integration of pre-trained models are all handled by the pipeline. A straightforward interface for picture generation is provided, and the complexity of dealing with diffusion models is abstracted away. The Diffusers package includes a Stable Diffusion Pipeline that simplifies the process of creating visuals from text by utilizing stable diffusion models. Stable Diffusion Pipeline cannot be used until the Diffusers library has been installed. Prior to utilizing the Stable Diffusion Pipeline, it is necessary to install the Diffusers library [40]. In order to accomplish this, you will need to use pip, Python's package manager, to install diffusers. With the library installed, we have access to the stable diffusion models and all of the functionality that goes along with them. The next step, after installing the Diffusers library, is to add the necessary classes and modules to your Python environment. In most cases, this requires bringing in modules for deep learning frameworks like PyTorch and specific classes from the Diffusers library. The provided example makes use of several imports, including Torch for PyTorch, Stable Diffusion Pipeline from diffusers, and picture from PIL for picture visualization.

```
# Library imports
# Importing PyTorch library, for building and training neural networks
import torch
# Importing StableDiffusionPipeline to use pre-trained Stable Diffusion models
from diffusers import StableDiffusionPipeline
# Image is a class for the PIL module to visualize images in a Python Notebook
from PIL import Image
```

A. Establishing an Instance of Stable Diffusion Pipeline

By importing the necessary modules and installing the library, we can create an instance of the Stable Diffusion Pipeline [41]. A pre-trained model, such as "CompVis/stable-diffusion-v1-4," and optional options, such as torch_dtype for datatype settings, are initialized with this instance. To generate images from text using the provided diffusion model, all the necessary functionality is present in the pipeline instance.

```
# Creating pipeline
pipeline = StableDiffusionPipeline.from_pretrained("CompVis/stable-diffusion-v1-4",
torch_dtype=torch.float16)
```

B. Use in the Creation of Image Grids

Prior to beginning to create images, it is helpful to establish a function that can generate an image grid. By passing in a set of images along with the desired row and column dimensions, this function generates a visual grid. This feature is really helpful because it enhances the presentation of generated photos and makes comparison easier.

```
# Defining function for the creation of a grid of images
def image_grid(imgs, rows, cols):
    assert len(imgs) == rows*cols
    w, h = imgs[0].size
    grid = Image.new('RGB', size = (cols*w, rows * w))
    grid_w, grid_h = grid.size
    for i, img in enumerate(imgs):
        grid.paste(img, box = (i%cols*w, i // cols*h))
    return grid
```

C. Transferring Pipeline to GPU

To expedite image processing, especially for neural network inference, it is usual practice to move the pipeline to a GPU. One way to speed up processing is by utilizing PyTorch to transfer models and tensors to GPU [42] devices. For faster text-to-image generation, the given example moves the StableDiffusionPipeline instance to the GPU using uses.to('cuda').

```
# Moving pipeline to GPU
pipeline = pipeline.to('cuda'
```

V. OUTCOME

As seen in Figure 6, stable diffusion models have been effectively used to create images from text. Converted detailed written descriptions into high-quality images, demonstrating the power of stable dissemination in the production of creative content. Researched and understood the function of score-based generative modelling, noise scheduling, and posterior sampling as they pertain to the Stable Diffusion. Improved usability is one outcome of web apps that use stable diffusion to generate images on the fly in reaction to user input (see Figures 7 and 8).

Website designers and developers can increase user engagement and retention with Stable Diffusion by creating aesthetically pleasing content that is personalized to the user's interests or behaviours.

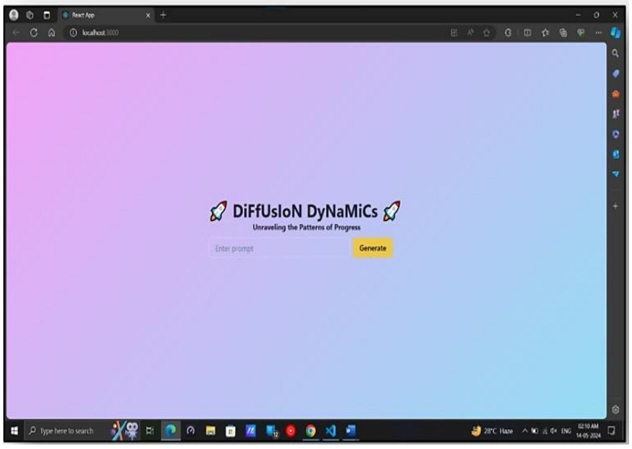
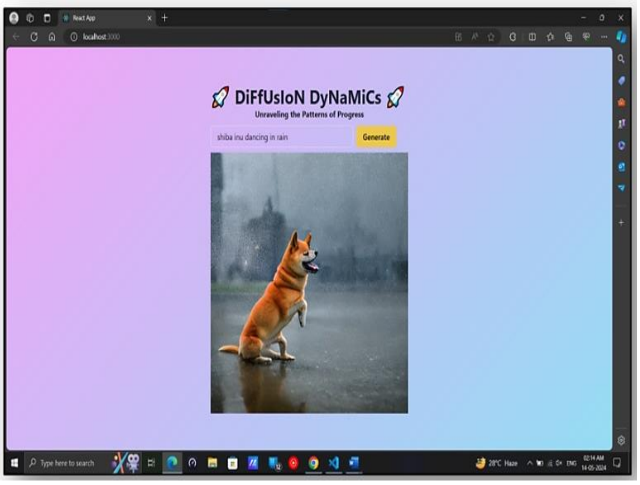


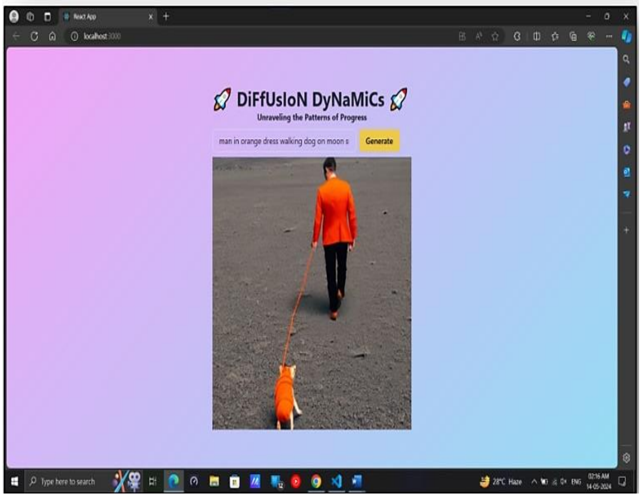
Figure 6: The Home Page

Using Stable Diffusion, web development platforms can automate the design process, resulting in unique and visually appealing layouts (Figures 9 and 10), illustrations, and graphics. Online shoppers can benefit from more accurate product visualizations by employing stable diffusion in conjunction with textual product descriptions. Original and inventive material can be easily created with Stable Diffusion and then integrated into many internet applications for advertising, entertainment, and art, among other uses.



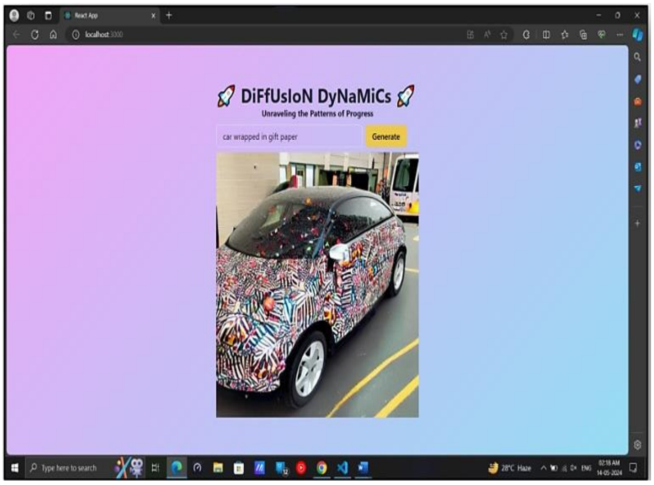
shiba inu dancing in rain

Figure 7: The Shiba Inu Dancing in Rain



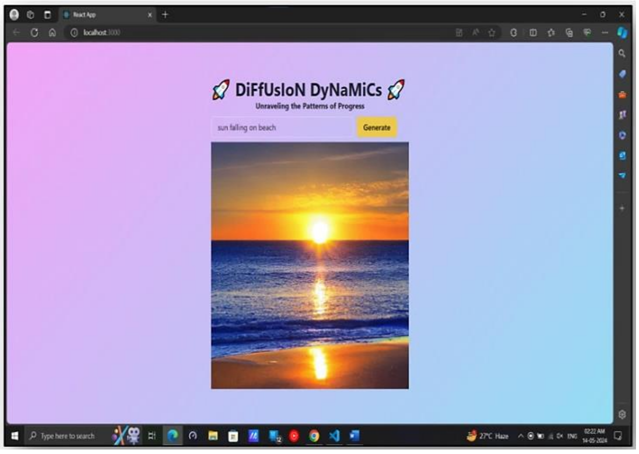
man in orange dress walking dog on moon s

Figure 8: The Man in Orange Dress Walking Dog on Moon Surface



car wrapped in gift paper

Figure 9: Car wrapped in Gift Paper



🚀 DiFfUsIoN DyNaMiCs 🚀

Unraveling the Patterns of Progress

sun falling on beach

Figure 10: The Sun Falling on Beach

VI. CONCLUSION

Nowadays, text-to-image generation is all the rage in computer vision and natural language processing. Producing consistently realistic photographs under specified settings is the most difficult endeavour. The current crop of text-to-image creation algorithms consistently cranks out misaligned images. Building a purpose-built web app is the focus of this article. It might be a social media site, a platform for online business, or even a tool for education. The goal of this research was to investigate whether steady diffusion could be useful for generating high-quality images from textual descriptions. Following an overview of stable diffusion and its salient characteristics, we dove headfirst into the training and application of stable diffusion models for text-to-image generation. This work demonstrated the efficacy of stable diffusion in producing realistic images from text, showcasing its potential across various fields. Although there have been numerous technical advancements, such as systems that aid in face recognition and matching, a straightforward method for generating photos for use in criminal investigations would be text-to-image generation.

CONFLICTS OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

[1] Zhang, C. Zhang, S. Zheng, M. Zhang, M. Qamar, S.-H. Bae, and I. S. Kweon, "A survey on audio diffusion models: Text to speech synthesis and enhancement in generative AI," arXiv preprint arXiv:2303.13336, 2023. Available from: https://doi.org/10.48550/arXiv.2303.13336

[2] S. M. Kosslyn, G. Ganis, and W. L. Thompson, "Neural foundations of imagery," Nat. Rev. Neurosci., vol. 2, pp. 635-642, 2001. Available from: https://doi.org/10.1038/35090055

[3] Y. Perwej, "An Evaluation of Deep Learning Miniature Concerning in Soft Computing," Int. J. Adv. Res. Comput. Commun. Eng., vol. 4, no. 2, pp. 10-16, Feb. 2015. Available from: https://doi.org/10.17148/IJARCCE.2015.4203

[4] A. Karpathy and L. Fei-Fei, "Deep visual-semantic alignments for generating image descriptions," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Boston, MA, USA, Jun. 2015, pp. 3128-3137. Available from: https://doi.org/10.48550/arXiv.1412.2306

[5] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, "Show and tell: A neural image caption generator," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Boston, MA, USA, Jun. 2015, pp. 3156-3164. Available from: https://doi.org/10.48550/arXiv.1411.4555

[6] C. Zhang et al., "A complete survey on generative AI (AIGC): Is chatgpt from GPT-4 to GPT-5 all you need?," arXiv:2303.11717, 2023. Available from: https://doi.org/10.48550/arXiv.2303.11717

[7] K. Xu et al., "Show, attend and tell: Neural image caption generation with visual attention," in Proc. Int. Conf. Mach. Learn., Lille, France, Jul. 2015, pp. 2048-2057. Available from: https://doi.org/10.48550/arXiv.1502.03044

[8] K. Wang, C. Gou, Y. Duan, Y. Lin, X. Zheng, and F.-Y. Wang, "Generative Adversarial Networks: Introduction and Outlook," IEEE/Caa J. Automatica Sinica, vol. 4, no. 4, pp. 588-598, Oct. 2017. Available from: https://doi.org/10.1109/JAS.2017.7510583

[9] I. Goodfellow et al., "Generative adversarial nets," in Advances in Neural Information Processing Systems, 2014, pp. 2672-2680. Available from: https://doi.org/10.48550/arXiv.1406.2661

[10] Y. Zhou et al., "Shifted Diffusion for Text-to-image Generation," in 2023 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Vancouver, BC, Canada, 2023, pp. 10157-10166. Available from: https://doi.org/10.1109/CVPR52729.2023.00979

[11] Y. Perwej, N. Akhtar, and F. Parwej, "The Kingdom of Saudi Arabia Vehicle License Plate Recognition using Learning Vector Quantization Artificial Neural Network," Int. J. Comput. Appl. (IJCA), USA, vol. 98, no. 11, pp. 32-38, 2014. Available from: https://10.5120/17230-7556

[12] J. Park et al., "LANIT: Language-Driven Image-to-Image Translation for Unlabeled Data," in Proc. 2023 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Vancouver, BC, Canada, 2023, pp. 23401-23411. Available from: https://doi.org/10.1109/CVPR52729.2023.02241

[13] I. J. Goodfellow et al., "Generative adversarial networks," Accessed on: Jul. 13, 2024. Available from: https://arxiv.org/abs/1406.2661

[14] S. Reed et al., "Generative adversarial text to image synthesis," Accessed on: Jul. 13, 2024. Available from: https://arxiv.org/abs/1605.05396

[15] T. Salimans et al., "Improved techniques for training GANs," in Proc. Advances in Neural Inf. Process. Syst. 29 (NIPS 2016), Spain, Dec. 5-10, 2016. Available from: https://doi.org/10.48550/arXiv.1606.03498

[16] Y. Perwej, "Recurrent Neural Network Method in Arabic Words Recognition System," Int. J. Comput. Sci. Telecommun. (IJCST), Sysbase Solution (Ltd), UK, London, ISSN 2047-3338, vol. 3, no. 11, pp. 43-48, Nov. 2012. Available from: https://doi.org/10.48550/arXiv.1301.4662

[17] T. Zia et al., "Text-to-Image Generation with Attention Based Recurrent Neural Networks,".Accessed on: Jul. 13, 2024. Available from: https://arxiv.org/abs/2001.06658

[18] Z. Yang et al., "ReCo: Region-Controlled Text-to-Image Generation," in Proc. 2023 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Vancouver, BC, Canada, 2023, pp. 14246-14255. Available from: https://doi.org/10.48550/arXiv.2211.15518

[19] J. Mao and X. Wang, "Training-Free Location-Aware Text-to-Image Synthesis," in Proc. 2023 IEEE Int. Conf. Image Process. (ICIP), Kuala Lumpur, Malaysia, 2023, pp. 995-999. Available from: https://doi.org/10.1109/ICIP49359.2023.10222616

[20] Y. Perwej and F. Parwej, "A Neuroplasticity (Brain Plasticity) Approach to Use in Artificial Neural Network," Int. J. Sci. Eng. Res. (IJSER), France, vol. 3, no. 6, pp. 1-9, Jun. 2012. Available from: https://10.13140/2.1.1693.2808

[21] R. Morita, Z. Zhang, and J. Zhou, "BATINeT: Background-Aware Text to Image Synthesis and Manipulation Network,"

in Proc. 2023 IEEE Int. Conf. Image Process. (ICIP), Kuala Lumpur, Malaysia, 2023, pp. 765-769. Available from: https://10.1109/ICIP49359.2023.10223174

[22] Y. Perwej, F. Parwej, and A. Perwej, "Copyright Protection of Digital Images Using Robust Watermarking Based on Joint DLT and DWT," Int. J. Sci. Eng. Res. (IJSER), France, ISSN 2229-5518, vol. 3, no. 6, pp. 1-9, Jun. 2012. Available from: https://10.13140/2.1.1693.2808

[23] Y. Perwej, A. Perwej, and F. Parwej, "An Adaptive Watermarking Technique for the Copyright of Digital Images and Digital Image Protection," Int. J. Multimedia Its Appl. (IJMA), Academy & Industry Res. Collab. Center (AIRCC), vol. 4, no. 2, pp. 21-38, Apr. 2012. Available from: https://10.5121/ijma.2012.4202

[24] Y. Dong, Y. Zhang, L. Ma, Z. Wang, and J. Luo, "Unsupervised text-to-image synthesis," Pattern Recognit., vol. 110, p. 107573, 2021. Accessed on: Jul. 13, 2024. Available from: https://doi.org/10.1016/j.patcog.2020.107573

[25] M. Berrahal and M. Azizi, "Optimal text-to-image synthesis model for generating portrait images using generative adversarial network techniques," Indones. J. Electr. Eng. Comput. Sci., vol. 25, pp. 972-979, 2022. Accessed on: Jul. 13, 2024. Available from: https://doi.org/10.11591/ijeecs.v25.i3.pp972-979

[26] Y. Zhang, S. Han, Z. Zhang, J. Wang, and H. Bi, "CF-GAN: Cross-domain feature fusion generative adversarial network for text-to-image synthesis," Vis. Comput., pp. 1-11, 2022. Accessed on: Jul. 13, 2024. Available from: https://doi.org/10.1007/s00371-022-02175-4

[27] M. Tao, H. Tang, F. Wu, X. Jing, B.-K. Bao, and C. Xu, "DF-GAN: A Simple and Effective Baseline for Text-to-Image Synthesis," in Proc. 2022 IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), New Orleans, LA, USA, 2022, pp. 16494-16504. Available from: https://doi.org/10.48550/arXiv.2008.05865

[28] Y. Perwej, "An Optimal Approach to Edge Detection Using Fuzzy Rule and Sobel Method," Int. J. Adv. Res. Electr. Electron. Instrum. Eng. (IJAREEIE), ISSN 2320-3765 (Print), ISSN 2278-8875 (Online), vol. 4, no. 11, pp. 9161-9179, 2015. Available from: https://10.15662/IJAREEIE.2015.0411054

[29] E. Mansimov, E. Parisotto, J. L. Ba, and R. Salakhutdinov, "Generating images from captions with attention," ICLR, 2016. Accessed on: Jul. 13, 2024. Available from: https://arxiv.org/abs/1511.02793

[30] L. Maiano et al., "Human Versus Machine: A Comparative Analysis in Detecting Artificial Intelligence-Generated Images," IEEE Secur. Privacy, vol. 22, no. 3, pp. 77-86, 2024. Available from: https://10.1109/MSP.2023.3062239002E

[31] S. Banerjee, G. Mittal, A. Joshi, C. Hegde, and N. Memon, "Identity-Preserving Aging of Face Images via Latent Diffusion Models," in Proc. 2023 IEEE Int. Joint Conf. Biometrics (IJCB), 2023, pp. 1-10. Available from: https://doi.org/10.48550/arXiv.2307.08585

[32] A. O. Levin and Y. S. Belov, "A Study on the Application of Using Hypernetwork and Low Rank Adaptation for Text-to-Image Generation Based on Diffusion Models," in Proc. 2024 6th Int. Youth Conf. Radio Electron. Electr. Power Eng. (REEPE), 2024, pp. 1-5. Available from: https://doi.org/10.1109/REEPE60449.2024.10479561

[33] D. Husain, Y. Perwej, S. K. Vishwakarma, Prof. (Dr.) S. Rastogi, V. Singh, and N. Akhtar, "Implementation and Statistical Analysis of De-noising Techniques for Standard Image," Int. J. Multidiscip. Educ. Res. (IJMER), ISSN: 2277-7881, vol. 11, no. 10 (4), pp. 69-78, 2022. Available from: https://10.IJMER/2022/11.10.72

[34] S. Siemens, M. Kastner, and E. Reithmeier, "Synthetically generated microscope images of microtopographies using stable diffusion," in Proc. Automated Visual Inspection Machine Vision V, 2023, p. 7. Available from: https://doi.org/10.1117/12.2673643

[35] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2022. Available from: https://doi.org/10.48550/arXiv.2112.10752

[36] A. Dwivedi, Dr. B. B. Dumka, N. Akhtar, F. Shan, and Y. Perwej, "Tropical Convolutional Neural Networks (TCNNs) Based Methods for Breast Cancer Diagnosis," Int. J. Sci. Res. Sci. Technol. (IJSRST), Print ISSN: 2395-6011, Online ISSN: 2395-602X, vol. 10, no. 3, pp. 1100-1116, May-Jun. 2023. Available from: https://10.32628/IJSRST523103183

[37] A. Blattmann, R. Rombach, H. Ling, T. Dockhorn, S. W. Kim, S. Fidler et al., "Align your latents: High-resolution video synthesis with latent diffusion models," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), 2023. Available from: https://doi.org/10.48550/arXiv.2304.08818

[38] N. Akhtar, Dr. H. Pant, A. Dwivedi, V. Jain, and Y. Perwej, "A Breast Cancer Diagnosis Framework Based on Machine Learning," Int. J. Sci. Res. Sci. Eng. Technol. (IJSRSET), Print ISSN: 2395-1990, vol. 10, no. 3, pp. 118-132, 2023. Available from: https://10.32628/IJSRSET2310375

[39] L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in Proc. Eur. Conf. Comput. Vis. (ECCV), 2018. Available from: https://doi.org/10.48550/arXiv.1802.02611

[40] J. Ho, C. Saharia, W. Chan, D. J. Fleet, M. Norouzi, and T. Salimans, "Cascaded diffusion models for high fidelity image generation," J. Mach. Learn. Res., vol. 23, pp. 47-1, 2022. Available from: https://doi.org/10.48550/arXiv.2106.15282

[41] A. Nichol, P. Dhariwal, A. Ramesh, P. Shyam, P. Mishkin, B. McGrew, I. Sutskever, and M. Chen, "Glide: Towards photorealistic image generation and editing with text-guided diffusion models,". Accessed on: Jul. 13, 2024. Available from: https://arxiv.org/abs/2112.10741

[42] J. Nickolls and W. J. Dally, "The GPU computing era," IEEE Micro, vol. 30, no. 2, pp. 56-69, Mar./Apr. 2010. Available from: https://10.1109/MM.2010.41