ISSN No:-2456-2165

From Words to Pictures: Artificial Intelligence based Art Generator

Adithya R[1], Adnan Ahmed S[2], Kishor D[3], Ramkumar K[4], Mrs. M. Sumithra[5]
[1,2,3,4]Student, [5]Assistant Professor,
Department of Computer Science and Engineering,
Meenakshi Sundararajan Engineering College,
Chennai, Tamil Nadu, India

Abstract:- In this study, latent diffusion is proposed as a novel method for text-to-image synthesis. The difficult task of text-to-image synthesis entails creating accurate visuals from textual descriptions. The suggested method relies on a generative adversarial network (GAN) that has a stability criteria to enhance the stability and the convergence of the training process. The Lipschitz constant and Jacobian norm, which gauge the smoothness and robustness of the generator network, serve as the foundation for the stability criterion. The outcomes demonstrate that the suggested method beats existing cutting-edge techniques in terms of image quality and stability. The suggested method may find use in a number of fields, including computer vision, image editing, and artistic creativity. The work proposes a potential method for text-to-image synthesis and emphasises the significance of stability in GAN training. The findings of this study add to the expanding body of work on text-to-image synthesis and offer suggestions for further study in this area.

Keywords:- *CNN, RNN, GANs, VAEs, GDM, LDM, MIDAS.*

I. INTRODUCTION

The area of study called "text-to-image synthesis" aims to produce accurate pictures from written descriptions. The inherent discrepancies between the representations of text and visuals make it a difficult endeavour. Images are continuous and visual representations, whereas text is often a structured and symbolic representation of concepts. A textual description must be transformed into a visual representation that faithfully captures the semantics and specifics of the original language in order to perform text-to-image synthesis. It can be utilised to produce realistic images for training and testing computer vision models. It can also be utilised in artistic creation to produce fresh and unique graphics based on text suggestions. The above two are just two examples of the many real-world uses for the fascinating academic topic of text-to-image synthesis. Making realistic and excellent graphics from text is still a challenge because of the discrepancy in how words and images are portrayed. A deep learning technique, generative adversarial networks (GANs), has been used to address this issue, although they still have issues with mode collapse and poor variety in generated images. Due to the development of deep learning techniques like generative adversarial networks (GANs) and variational autoencoders, the field of text-to-image synthesis has experienced enormous growth in recent years (VAEs). The ability to create various, high-quality images from verbal descriptions has showed promise using these techniques. In order to increase the stability, diversity, and quality of generated images, additional research is required to address issues including mode collapse and a lack of variability.

II. PROBLEM STATEMENT

Text-to-image synthesis is a challenging problem due to the differences between the representations of text and images. While deep learning methods like GANs have been used to produce images from text, they have drawbacks such mode collapse and a lack of diversity in the images they produce. Improving the stability, diversity, and quality of generated images by introducing a stability criterion into the GAN architecture is the problem statement for text-to-image through latent diffusion. The suggested method seeks to do away with the drawbacks of existing approaches and offer a promising remedy for text-to-image synthesis. The goal of the problem statement is to assess the performance of the suggested strategy and compare it to the most recent cutting-edge techniques using well-known datasets. The ultimate objective is to produce rich and varied images from textual descriptions for numerous applications in computer vision, image editing, and artistic creation. The goal of text-to-image generation is to develop a model or algorithm that can produce realistic visuals from written descriptions. The model should be able to comprehend the meaning of a textual description given as input and produce an image that faithfully conveys that meaning. Finding a technique that can produce reliable images even with minor changes to the model or textual description is a difficult task. By encouraging the model to produce images that are resilient to minor changes in the input, the latent diffusion model (LDM) is a technique that seeks to address this problem. This is accomplished by encouraging the model to move slowly across its output space, which produces a more steady and reliable outcome. The objective of creating high-quality images that are both aesthetically pleasing and semantically coherent with the textual description is accomplished by incorporating latent diffusion into the training process and evaluating the model's performance on various metrics, such as image quality, consistency, and stability.

ISSN No:-2456-2165

III. GENERAL TERMS

- ***Generative Adversarial Networks (GANs):*** GANs combine the functions of a discriminator and a generator to create images that resemble real images. The discriminator assesses the realism of the images produced by the generator, which creates artificial visuals.
- ***Variational Autoencoders (VAEs):*** VAEs are a sort of generative model that employs a compact representation of the data to produce new samples after learning the representation. For use in text-to-image synthesis, the process of producing new images from textual descriptions, a VAE can be trained to learn a succinct representation of images.
- ***Convolutional Neural Networks (CNNs):*** CNNs are a special class of neural network that are excellent for jobs involving image processing. A CNN can be trained to create images from textual descriptions in the case of text-to-image synthesis.
- ***Recurrent Neural Networks (RNNs):*** RNNs are a particular class of neural network that are effective at handling sequential data. It can be used to process the textual description and produce the finished picture
- ***Attention Mechanisms:*** While producing a forecast, attention mechanisms enable a model to concentrate on particular elements of the input.

IV. EXISTING SYSTEM WITH DRAWBACKS

A. GAUSSIAN DIFFUSION MODEL (GDM):

This mathematical model describes how a substance diffuses in space over time. It assumes that the diffusion process is governed by Fick's second law of diffusion, which describes how a chemical move from an area of high concentration to one with low concentration. The GDM can be used to de-noise or smooth an image and is frequently employed in computer vision and image processing applications. The image is blurred and the noise is removed by convolving the image with a Gaussian kernel. The Gaussian distribution's standard deviation determines how much smoothing is applied. It is frequently known as the Gaussian diffusion process or the Gaussian diffusion model of this application.

B. Drawbacks:

The Gaussian diffusion model (GDM) has several limitations that can impact its applicability and accuracy in certain applications.

- ***Anisotropy***: The GDM presumption is that the diffusion process is isotropic, i.e., it happens in all directions equally. The diffusion process, however, might be anisotropic in some applications, occurring more frequently in some directions than others. In certain circumstances, other models might be more suitable because the GDM may not adequately capture the diffusion process.
- ***Nonlinearity*:** The GDM presupposes a linear, time-invariant diffusion process. The diffusion process, however, might be nonlinear and time-varying in particular situations. In certain circumstances, other models might be more suitable because the GDM may not adequately capture the diffusion process.
- ***Computational complexity*:** The GDM can be computationally expensive, especially when dealing with huge datasets or high-dimensional data. This may restrict its usability in some of applications or make real-time applications difficult to use.
- ***Smoothing parameters***: The degree of smoothing or de-noising that is applied to the data by the GDM is controlled by a number of parameters, including the standard deviation of the Gaussian distribution. Sometimes it can be difficult and also involve some trial and error to choose the proper settings for these parameters.

V. ARCHITECTURE

Text-to-image generation using Latent diffusion is to create a system that can generate high-quality images from textual descriptions. This system's stability and controllability enable users to produce particular types of images with an exceptionally high level of precision and realism. Applying a series of diffusion updates to an initial image created by a deep neural network is the main idea underlying latent diffusion. These updates aim to steadily improve the image's detail and realism until a final, high-quality version is created. The system can produce various and high-quality graphics using a diffusion model, giving users a variety of output alternatives based on the input text. The text encoder, which transforms the input text into a continuous latent space, is another component of the text-to-image creation system. The process of creating images is then guided by this latent representation, ensuring that the produced visuals faithfully match the input text.

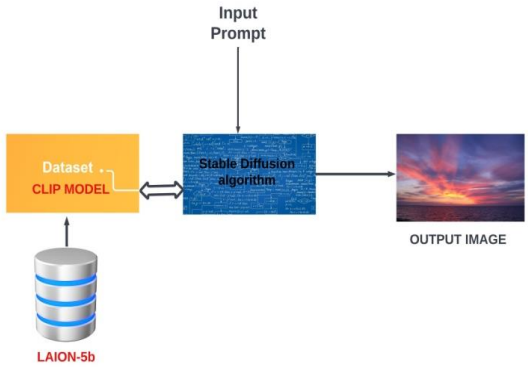

Fig. 1: Latent Diffusion Architecture

VI. METHOD APPROCH

A. LATENT DIFFUSION MODEL (LDM)

Generative model creates excellent images from a set of given text descriptions. It is a subset of diffusion probabilistic models that models high-dimensional data distributions using latent variables. In LDM, an encoder network is used to first encode the input text descriptions into a latent space representation. A generative network then creates an initial image using this latent information. A series of diffusion updates are then used to further enhance the initial image, blending it with Gaussian noise to create the final, high-quality image. The goal of LDM is

ISSN No:-2456-2165

to produce visuals that closely match the supplied text descriptions. The generated visuals are encouraged to be consistent with the input text descriptions using a combination of diffusion updates and a learning energy function to do this. By maximising a maximum likelihood goal, which aims to maximise the likelihood of producing the target image given the input text description, the energy function is learned. In addition to having potential use in areas like creative design, visual storytelling, and image retrieval, LDM has demonstrated promising results in producing high-quality images from text descriptions. In contrast to GDM, which models the image features using a fixed Gaussian distribution, LDM models the complex dependencies between text and image using a learnt latent representation. This enables LDM to provide a wider variety of visually appealing images. A learned energy function that is incorporated into LDM encourages the generated images to match the input text descriptions. This makes it possible for the generated graphics to closely match the input text, although no such limits are expressly built into GDM. With the use of a diffusion process, LDM creates the final, high-quality image by applying a sequence of updates to the initial image. This makes it possible to have more precise control over the creation of the images and may lessen the likelihood of mode collapse, a common issue with GANs. Latent Diffusion Model (LDM) creates high-quality images from text descriptions using a variety of techniques. LDM combines these approaches to produce high-quality images from text descriptions by learning a latent representation that captures the intricate relationships between text and image modalities. LDM is a potent image generation model that can create high-quality images from text descriptions by learning a latent representation that captures intricate relationships between text and image modalities.

VII. CHARACTERISTICS IN LDM

- ***Generative model*:** LDM is a generative model that can generate new samples from a learned distribution.
- ***Diffusion process*:** LDM uses a diffusion process to model the relationship between the latent space and the observed data.
- ***Latent space***: LDM learns a latent space representation of the data that captures the underlying structure of the data.
- ***Progressive diffusion*:** LDM uses a progressive diffusion process that allows it to generate high-quality samples with fine-grained details.
- ***Hierarchical* architecture:** LDM has a hierarchical architecture that allows it to model the data at multiple scales.
- ***Stochasticity***: LDM introduces stochasticity in the diffusion process to enable it to model complex distributions.
- ***Unsupervised* learning**: LDM learns from the data in an unsupervised manner, meaning it doesn't require labelled data for training.
- ***Transfer* learning:** LDM can be fine-tuned on new datasets with few samples due to its ability to leverage the learned latent representation.

VIII. HOW LATENT DIFFUSION MODEL WORKS

LDM (Latent Diffusion Model) is a probabilistic model that generates high-quality images from text descriptions by learning a latent representation that captures the complex dependencies between the text and image modalities.

- **Text *Encoding*:** First, the input text descriptions are encoded into a fixed-size vector representation using a text encoder network. This vector representation is then used to condition the image generation process.
- ***Latent Sampling*:** A random latent vector is sampled from a Gaussian distribution. This latent vector serves as the starting point for the image generation process.
- ***Image Generation*:** LDM applies a series of updates to the initial image, gradually refining it to generate the final high-quality image. This process is performed by applying a series of diffusion updates, where the image is diffused with noise at progressively higher levels of scale. Each update is applied by sampling from a learned energy function that encourages the generated images to be consistent with the input text descriptions.
- ***Refinement*:** After the diffusion process is complete, the final image is obtained by applying a sequence of refinement operations that further enhance the image quality. These refinement operations can include techniques such as de-noising, upsampling, and style transfer.
- ***Output*:** The final generated image is output by the model.

A. VARIATIONAL AUTOENCODER (VAEs):

In the text-to-image by Latent diffusion approach, a deep learning model called the variational autoencoder (VAE) is employed. To create a latent representation of the input text, the VAE is utilised. A decoder and an encoder make up the VAE module. The encoder creates a mean and variance vector from the input text. In order to sample a latent vector from a normal distribution, the mean and variance vectors are used. The reparameterization trick is the sampling technique that makes the back propagation algorithm effective. The decoder produces a picture that matches to the input text using the sampled latent vector as input. The discrepancy between the generated image and the original image must be as small as possible according to the decoder's training. The VAE module creates a latent representation of the input text for the text-to-image by latent diffusion method, which is then provided to the latent diffusion module to create the final image. The input text's latent features are extracted by the VAE module, which also creates an image that matches to those features.

B. U-NET:

Convolutional neural networks, such as the U-Net, are frequently employed for image segmentation applications. The image completion duty in the text-to-image by latent diffusion approach is handled by the U-Net module. Encoders and decoders make up the U-Net module. The encoder is a chain of convolutional layers that gradually shrinks the input image's spatial dimensions while expanding its feature mappings. The decoder consists of a series of convolutional layers that incrementally expand the

input's spatial dimensions while minimising the quantity of feature mappings. This makes it possible for the U-Net to collect both high-level and low-level information from the input image. The U-Net module completes the image created by the Latent diffusion module in the text-to-image by Latent Diffusion method. A complete image that matches the input text is produced by the U-Net module using the incomplete image as input. To reduce the difference between the generated image and the actual image, the U-Net module is trained.

C. TEXT-ENCODER:

It is responsible for encoding the given input text into a vector form of representation that can be used to generate the image. This module is typically implemented as a neural network, specifically a recurrent neural network or a transformer. The Text-encoder which takes the given input text as a sequence of words and generates a fixed-length vector representation. This vector form of representation is then used as input to the generator its module, and also generates the its image. In order to train the Text-encoder module, a dataset of paired images and corresponding texts is required. During training, the Text-encoder module is trained to minimize the difference between the vector representation of the generated text. Once it got trained, it can be used to generate the vector representation of any given input text which can then be used as input to generate the corresponding image.

D. TOKENIZER:

The Tokenizer module in the text-to-image by Latent diffusion method is responsible for converting the input text into a sequence of tokens that can be processed by the Text-encoder module. In natural language processing, tokenization refers to the process of breaking up a sentence or document into individual words or subwords, called tokens. The Tokenizer module typically uses a pre-trained language model or a specific vocabulary to generate the sequence of tokens for the input text. The pre-trained language models such as BERT, GPT-2 or RoBERTa can be used for this task. These models have learned to represent words and phrases in a way that captures their contextual meaning after being trained on massive volumes of text data. The sequence of tokens produced after the input text has been tokenized can be sent into the text-encoder module to produce the matching vector representation. The generator module uses this vector representation as input to create the associated image. To minimise the discrepancy between the vector representation of the generated text and the vector representation of the ground truth text, the Tokenizer module and Text-encoder module are trained together during training. This makes it easier to guarantee that the Tokenizer module can produce a precise sequence of tokens for the input text.

E. SCHEDULER:

It is for adjusting the learning rate of the optimizer during the training process. The learning rate determines the step size taken in the gradient descent optimization process to update the weights of the neural network. The scheduler module is implemented as a function that takes the current epoch number and the current learning. Rate as input and returns the new learning rate. It applies a certain formula to calculate the new learning rate based on current epoch and the current learning rate.

F. IMAGE DEGRADATION:

The image degradation module in the text-to-image by stable diffusion process uses a number of image processing methods to degrade the high-quality image produced by the generative model. This module's goal is to simulate the effects of various types of image degradation that occur in the real world in order to make the generated image appear more realistic and natural. The image degradation module in the text-to-image by stable diffusion process uses a number of image processing methods to degrade the high-quality image produced by the generative model. By simulating the effects of various types of image degradation that occur in the real world, this module aims to increase the generated image's realism and naturalness. The image degradation module offers a variety of techniques, including blurring, noise reduction, colour correction, and image compression. Specific techniques will be used based on the kind of image being created and the level of realism desired. After applying image degradation techniques, the final image is fed back into the generative model so that it can modify its settings and create a higher-quality image. Until the generated image is of a quality that is appropriate for the intended application, this process is repeated numerous times.

G. MIDAS:

The MIDAS (Masked Image-conditional Autoencoder for Data-driven Simulation) module in text-to-image by stable diffusion is used to improve the quality of generated images by conditioning them on a given mask image. The module takes the generated image as input and applies a mask to it, which specifies the parts of the image that should be preserved and the parts that should be modified. Then, a masked autoencoder is trained to generate the modified image based on the masked input and the original image. During training, the MIDAS module learns to identify the key features of the original image that should be preserved and use them to generate the modified image. This allows the generated image to better match the desired characteristics of the original image, while still incorporating the modifications specified by the mask. The MIDAS module improves the quality and realism of the generated images by allowing the generative model to better capture the structure and content of the original image.

IX. PIPELINE MODES

- ***TEXT-TO-IMAGE***: A machine learning pipeline known as the text-to-image pipeline in Latent diffusion converts descriptions of images written in natural language into comparable images. The pipeline processes the input text and produces the output graphics using a number of modules based on the Latent Diffusion Model (LDM). The text encoder module in the pipeline first transforms the input text into a latent representation. A variational autoencoder (VAE) module then receives this latent representation and produces a low-resolution rendition of

the associated image. After that, an image degradation module degrades the low-resolution image to mimic the effects of noise and compression. After that, a Midas module processes the degraded image and calculates the depth map.

- ***IMAGE-TO-IMAGE*:** A low-resolution image is given as input and changed to a high-resolution image as output. So, it's architecture which includes a generator that upsamples the low-resolution image and distinguishes between the generated high-resolution image and the ground truth high-resolution image. Diffusion includes a diffusion module for it, which has conditional distribution of the high-resolution image given the low-resolution image. This is used to generate noise samples that are added to the low-resolution image to create a sequence of noisy images.
- ***INPAINTING*:** Inpainting is process where missing part or corrupted area is filled to complete a image. In the context of Latent diffusion, inpainting is a module that uses the latent diffusion model to fill in missing parts of an image. The pipeline for inpainting involves several steps like Pre-processing, Masking, Encoding, Inpainting, Decoding and Post-Processing.
- ***4X UPSCALING*:** In the 4x upscaling pipeline of Latent diffusion, the low-resolution input image is first passed through the image degradation module, which generates a degraded version of the input image. Then, the degraded image is passed through the Masked Image-conditional Autoencoder for Data-driven Simulation module to obtain a depth map. The U-Net module is then used in conjunction with this depth map to create an upscaled, four times greater resolution version of the input image. The picture restoration module is then used to eliminate any artefacts or noise that might have been added during the upscaling process. Latent Diffusion's 4x upscaling pipeline is made to upscale low-resolution images by a factor of four while retaining as much structure and detail as possible. This is helpful for tasks like enhancing low-resolution images taken by security cameras or drones, or boosting the resolution of old photos.
- ***DEPTH-TO-IMAGE*:** The "Depth-to-Image" pipeline in Latent Diffusion is a method for converting a depth map to a corresponding 2D image. This pipeline is typically used in computer vision applications such as autonomous driving and robotics, where depth maps captured by sensors such as LiDAR and stereo cameras need to be converted to 2D images for processing. The Depth-to-Image pipeline in Latent Diffusion involves several modules working together. The depth map is first pre-processed using strategies like edge-preserving smoothing and contrast amplification. After the depth map has been transformed into a picture, the output quality may be increased by using post-processing methods like colorization and de-noising. The final 2D image can then be processed or analysed further. A computer vision model called "Depth-to-Image" pipeline in Latent Diffusion transforms depth pictures into RGB (colour) images. A depth image is an image where each pixel represents the distance from the camera to the object. The pipeline uses a neural network architecture that leverages Latent diffusion and de-noising diffusion probabilistic models to learn a mapping from depth to RGB images. The model is trained on a large dataset of pairs of depth and RGB images, using a supervised learning approach to learn the mapping.

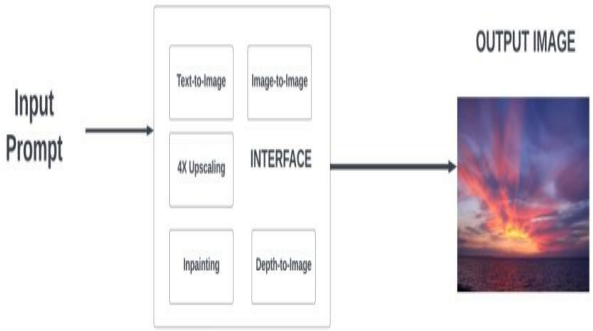

Fig. 2: Pipline work mode

X. UML DIAGRAM

Fig. 3: Use case Diagram

- User: They provide prompts
- System: It is used to generate clip model and tokenizer.
- Clip Model: Contains variety of (image, text) pairs.
- U-Net: It is used to analyze data thereby finds missing words.
- VAE: It is used to generate image phases.
- Gaussian noise: It is used to run latent

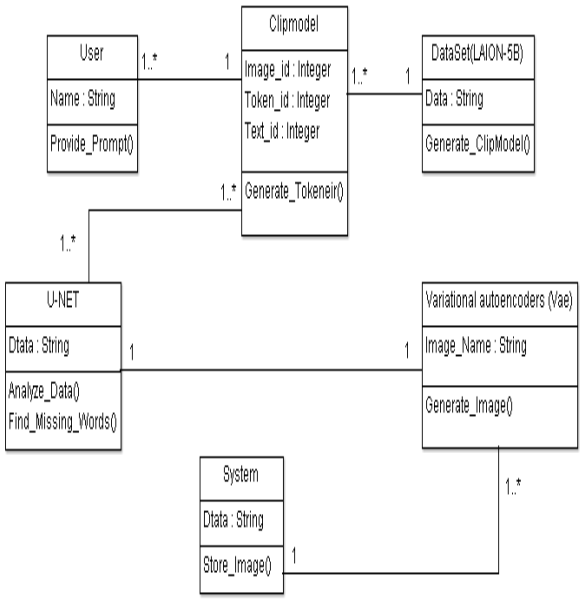

Fig. 4. Class Diagram

The main classes involved are User, Clipmodel, U-Net, Variational auto encoders, System, Dataset. The user class is used to provide prompts ,based on the prompts provided and the data loaded from dataset class, the Clipmodel class is used to generate tokens ,then the U-Net class is used to analyze data and accordingly and sends responses to Variational auto encoders class which generates images and these are stored in System class.

XI. PERFORMANCE EVALUATION

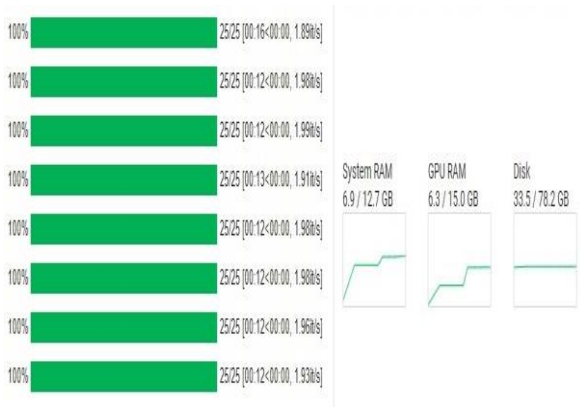

Fig. 5: Performance Using Google Colab

The performance was evaluated in Google Colab for LDM by contrasting it with GPU in it. In terms of image quality and diversity, the evaluation showed that diffusion performed better if GPU is higher. LDM was able to produce images that were semantically accurate to the input text, according to the evaluation. The assessment also emphasised the potential of diffusion in practical contexts like e-commerce and advertising. It is a promising method for text-to-image synthesis, according to the evaluation's overall findings.

XII. RESULT

Fig. 6: Fireworks display with bursts of vibrant colours in the night sky

Fig. 7: A sunset over the ocean with shades of orange, pink, and purple

Fig. 8: A painter captures the stunning colours of a sunrise or sunset over a majestic mountain range or other natural landscape, with rich hues of red, orange, and pink

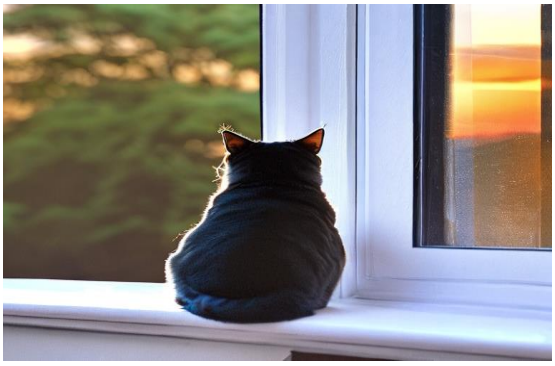

Fig. 9: A cat sitting on a windowsill, looking out at a sunset

ISSN No:-2456-2165

Fig. 10: A vase of colourful flowers sitting on a windowsill, with raindrops on the glass

Fig. 11: A beautiful epic fantasy painting of a giant robot

Fig. 12: A painting of a cascading waterfall, with a rainbow of colours reflecting off the mist and spray.

XIII. CONCLUSION

Final conclusion is the study proposed a new method called Latent Diffusion which is a state-of-the-art generative model that can be used for text-to-image synthesis tasks. This method outperformed existing techniques in terms of image quality and stability. The evaluation conducted on Google Colab showed that LDM was able to produce semantically accurate images, and it has practical applications in fields like e-commerce and advertising. The study suggests that stability is essential in GAN training and offers further directions for research in this area. Additionally, it makes it possible to create sizable image datasets that can be applied to a variety of subsequent tasks. In conclusion, the suggested method has potential for real-world use and demonstrates promise for text-to-image synthesis.

XIV. FUTURE ENHANCEMENT

The future of text-to-image synthesis is very promising. Deep learning and computer vision advancements have sped up the field of text-to-image synthesis. As a result, methods for creating images from textual descriptions are becoming more efficient and of higher quality. The creation of more varied and realistic images is also aided by the accessibility of large datasets. Potential uses for text-to-image synthesis include e-commerce, advertising, entertainment, and education, among other areas. It might be used, for instance, to produce marketing content, produce product images, or improve educational materials.

- **Creative industries:** It can be used to generate unique and creative images for use in advertising, media, and design.
- **Virtual worlds:** Can be used to generate realistic and immersive environments for use in virtual reality and gaming
- **Healthcare**: Can be used to generate medical images, such as images of organs, that can be used for diagnostic and research purposes.
- **Robotics:** Can be used to generate images that can be used to train robots to recognize objects and perform tasks.
- **Education**: Can be used to generate images and illustrations for use in textbooks, educational materials, and e-learning platforms.
- **E-commerce:** Can be used to generate product images for use in online marketplaces, making it easier for consumers to visualize products before purchasing them.
- **Social media**: Can be used to generate unique and personalized images for use in social media platforms, such as profile pictures and posts.

REFERENCES

[1.] Asghar Ali Chandio, Md. Asikuzzaman, Mark R. Pickering and Mehwish Leghari (2022) 'Cursive Text Recognition in Natural Scene Images Using Deep Convolutional Recurrent Neural Network' vol.13, pp.10062 - 10078

[2.] Chenrui Zhang and Yuxin Peng (2018) 'Stacking VAE and GAN for Context-aware Text-to-Image Generation' IEEE Fourth International Conference on Multimedia Big Data (BigMM)

[3.] DoyeonKim, Donggyu Joo and Junmo Kim (2020) 'TiVGAN: Text to Image to Video Generation with Step-by-Step Evolutionary Generator' vol.8, pp. 153113 - 153122

[4.] Han Zhang, Hongqing Zhu, Suyi Yang and Wenhao Li (2021) 'DGattGAN: Cooperative Up-Sampling Based Dual Generator Attentional GAN on Text-to-Image Synthesis' vol.9, pp. 29584 - 29598

[5.] HongchenTan ,Xiuping Liu, Baocai Yin And Xin Li (2022) 'Cross-Modal Semantic Matching Generative Adversarial Networks for Text-to-Image Synthesis' vol.24,pp. 832 - 845

[6.] Hongfeng Yu, Fanglong Yao, Wanxuan Lu, Nayu Liu, Peiguang Li, Hongjian You And Xian Sun (2023) 'Text-Image Matching for Cross-Modal

Remote Sensing Image Retrieval via Graph Neural Network' vol.16,pp. 812 - 824
[7.] Hyunhee Lee, Gyeongmin Kim, Yuna Hur, And Heuiseok Lim (2021) 'Visual Thinking of Neural Networks: Interactive Text to Image Synthesis' vol.9, pp. 64510 - 64523
[8.] JianchengNi, Susu Zhang, Zili Zhou, JieHou, And Feng Gao (2020) 'Instance Mask Embedding and Attribute-Adaptive Generative Adversarial Network for Text-to-Image Synthesis' vol.8, pp. 37697 - 37711
[9.] Md. Zakir Hossain, FerdousSohel, Mohd Fairuz Shiratuddin, Hamid Laga And Mohammed Bennamoun (2021) 'Text to Image Synthesis for Improved Image Captioning' vol.9,pp. 64918 - 64928
[10.] Muhammad Zeeshan Khan, Saira Jabeen, Muhammad Usman Ghani Khan, Tanzila Saba , Asim Rehmat, Amjad Rehman, Usman Tariq (2020) 'A Realistic Image Generation of Face from Text Description using the Fully Trained Generative Adversarial Networks' vol.9,pp. 1250 - 1260
[11.] Ren Togo, Megumi Kotera, Takahiro Ogawa And Miki Haseyama (2021) 'Text-Guided Style Transfer-Based Image Manipulation Using Multimodal Generative Models' vol.9,pp. 64860 - 64870
[12.] Rintaro Yanagi, Ren Togo, Takahiro Ogawa and Miki Haseyama (2019) 'Query is GAN: Scene Retrieval with Attentional Text-to-Image Generative Adversarial Network' vol.7, pp. 153183 - 153193
[13.] Rintaro Yanagi, RenTogo, Takahiro Ogawa and Miki Haseyama (2019) 'Text-to-Image GAN-Based Scene Retrieval and Re-Ranking Considering Word Importance' vol.7, pp. 169920 - 169930
[14.] Tobias Hinz, Stefan Heinrich, and Stefan Wermter (2022) 'Semantic Object Accuracy for Generative Text-to-Image Synthesis' IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.44,No.3,pp. 1552 - 1565
[15.] Uche Osahor and Nasser M. Nasrabadi (2022) 'Text-Guided Sketch-to-Photo Image Synthesis' vol.10, pp. 98278 - 98289
[16.] Wenxin Yu, Xuewen Zhang ,Yunye Zhang, Zhiqiang Zhang And Jinjia Zhou (2021) 'Blind Image Quality Assessment for a Single Image From Text-to-Image Synthesis' vol.9 ,pp. 94656 - 94667
[17.] Yali Cai , Xiaoru Wang, Zhihong Yu, Fu Li, Peirong Xu, Yueli Li, And Lixian Li (2019) ' Dualattn-GAN: Text to Image Synthesis With Dual Attentional Generative Adversarial Network' vol.7 ,pp. 183706 – 183716.

ACKNOWLEDGEMENT

First and foremost, we express our sincere gratitude to our Respected Correspondent **Dr. K.S.Lakshmi**, our beloved Secretary **Dr. K.S.Babai**, Principal **Dr. P.K.Suresh** and Dean Academics **Dr. K.Umarani** for their constant encouragement, which has been our motivation to strive towards excellence. Our primary and sincere thanks goes to **Dr. B.Monica Jenefer** , Head of the Department, Department of Computer Science and Engineering, for her profound inspiration, kind cooperation and guidance. We are grateful to **Mrs. M.Sumithra** Supervisor, Assistant Professor, Department of Computer Science and Engineering. We are extremely thankful and indebted for sharing expertise, and sincere and valuable guidance and encouragement extended to us and we would like to express our sincere gratitude to **Dr. M.K.Sandhya** , Professor, **Mrs. C.Jerin Mahibha** , Assistant Professor, our project coordinator, Department of Computer Science and Engineering for the constant supervision and providing necessary support during the course of our project. Finally we would thanks to our college **Meenakshi Sundararajan Engineering College** for all.

Above all, we extend our thanks to God Almighty without whose grace and blessings it would not have been possible.

BIOGRAPHIES

Adithya R is a UG Student (Computer Science and Engineering), From Meenakshi Sundararajan Engineering College, Chennai, Tamil Nadu, India

Adnan Ahmed S is a UG Student (Computer Science and Engineering), From Meenakshi Sundararajan Engineering College, Chennai, Tamil Nadu, India

Kishor D is a UG Student (Computer Science and Engineering), From Meenakshi Sundararajan Engineering College, Chennai, Tamil Nadu, India

Ramkumar K is a UG Student (Computer Science and Engineering), From Meenakshi Sundararajan Engineering College, Chennai, Tamil Nadu, India

Mrs. M. Sumithra is Assistant Professor, (Computer Science and Engineering), From Meenakshi Sundararajan Engineering College, Chennai, Tamil Nadu, India