Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 1, January : 2024

AI IMAGE GENERATION USING STABLE DIFFUSION MODEL

Syed Faisal Assistant Professor Dept. of Computer Engineering Sandip Institute of Technology and Research Centre, Nashik, Maharashtra: syed.faisal@sitrc.org

Sunil Kale Assistant Professor Dept. of Computer Engineering Sandip Institute of Technology and Research Centre, Nashik, Maharashtra

Harsh Chaudhari Department of CSE Sandip Institute of Technology and Research Centre, Nashik, Maharashtra Savitribai Phule Pune University. harshchaudhari0110@gmail.com

Harshal Chaudhari Department of CSE Sandip Institute of Technology and Research Centre, Nashik, Maharashtra Savitribai Phule Pune University. theguyharshal.@gmail.com

Parag Pal Department of CSE Sandip Institute of Technology and Research Centre, Nashik, Maharashtra Savitribai Phule Pune University. Paalparag007@gmail.com

Gitesh Makwane Department of CSE Sandip Institute of Technology and Research Centre, Nashik, Maharashtra Savitribai Phule Pune University. Giteshmakwane2002@gmail.com

Abstract

The emergence of AI Image Generator represents a ground-breaking advancement in technology, revolutionizing the landscape of content creation. This innovative system harnesses the power of advanced machine learning algorithms to automate the intricate process of transforming textual descriptions and keywords into high-quality, visually appealing images. Catering to the diverse needs of content producers and designers, this technology not only meets but exceeds industry standards for image quality and content relevance. At its core, the AI Image Generator is defined by its user-centric approach, offering a remarkably intuitive and user-friendly interface. Designed to accommodate users with varying technical backgrounds, the system provides extensive customization options. Users are empowered to tailor image styles, content, and parameters according to their specific requirements and creative visions. This level of flexibility ensures that the generated images align seamlessly with the unique demands of each user, fostering a sense of creative freedom and expression. Crucially, the system is engineered for scalability and efficiency, capable of handling multiple image requests with unparalleled ease. Through rigorous testing and validation processes, its performance and quality are meticulously verified, guaranteeing reliability and consistency in every generated image. Moreover, the system's scalability ensures its applicability across a wide range of contexts, from individual users to large-scale commercial projects.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 1, January : 2024

1. Introduction:

Image generation is a technology that enhances creativity by providing tools for creating unique visuals. It streamlines the creation process, saving time and effort by automating tasks that would be timeconsuming for humans. The goal is to maintain consistent style and quality in generated images, reducing errors and variability. AI image generation allows for personalized content creation, serving various needs. It is cost-effective, especially for repetitive or high-volume image production tasks. The technology is scalable, meeting the demands of content generation for websites, marketing, and more. It simplifies the image creation process, making visual content accessible to individuals with varying design expertise.

AI image generation also helps in data visualization, turning complex data into visually appealing charts, graphs, and infographics. It encourages innovation in art, design, and various industries by providing new tools and possibilities for creative expression. AI models are continuously improving with more data and user feedback. AI image generators are trained on an extensive amount of data, which comprises large datasets of images. Through the training process, the algorithms learn different aspects and characteristics of the images within the datasets. As a result, they become capable of generating new images that bear similarities in style and content to those found in the training data. There is a wide variety of AI image generators, each with its own unique capabilities.

Notable among these are the neural style transfer technique, which enables the imposition of one image's style onto another; Generative Adversarial Networks (GANs), which employ a duo of neural networks to train to produce realistic images that resemble the ones in the training dataset; and diffusion models, which generate images through a process that simulates the diffusion of particles, progressively transforming noise into structured images.

Challenges:

1. Recognizing Descriptions Making AI systems understand complex and context-specific textual descriptions is one of the major issues. Because language is complex, AI must be able to understand the hidden meaning of words in order to provide relevant visuals.

2. AI systems have to be careful not to always produce the same image for inputs with similar text. Some changes is necessary to prevent the AI from creating the same or very similar images for related textual prompts, particularly when the input descriptions differ slightly.

3. Keeping Clarity and Realism in Mind Clear and realistic photos should be produced. Rather than being blurry or abstract, they should at least resemble real-world photos. One of the biggest challenges in AI image synthesis is producing generated images with realistic details and clarity.

4. Producing Images with Greater Resolution Complexity is increased by producing highresolution photographs. More complex features are needed at higher resolutions, and complex techniques and processing capacity are needed to maintain accuracy and clarity at greater image sizes.

5. A Wide Range of Sufficient Training Data for AI systems to learn efficiently, a wide variety of training datasets are needed. Giving the AI a variety of examples helps it comprehend a broad range of scenarios, which improves its capacity to produce a variety of visuals from textual inputs.

Key Components of the System:

1. **Natural Language Processing (NLP):** - NLP is used to interpret and understand textual descriptions or prompts provided by users. It helps in converting human language into a format that the AI model can work.
2. **Generative Models:** - Generative models, such as Generative Adversarial Networks (GANs) and Variational Autoencoders

(VAEs), are at the core of AI image generation. These models learn to generate images that match the textual descriptions by optimizing their parameters through training

3. **Text-to-Image Generation Models:-** Specific models designed for text-to-image generation, like DALL·E and CLIP, are crucial. These models combine NLP and computer vision techniques to create images based on textual inputs.
4. **Software:** Software is the set of instructions and programs that dictate how the system operates. It includes the operating system, application software, firmware, and other software components that control the system's behaviour. Software plays a crucial role in instructing the processing unit on how to process input data and produce the desired output.

Key Features and Benefits:

1. **Generative Models:** AI-powered image generation relies on sophisticated generative models, such as Generative Adversarial Networks (GANs) and Variational Autoencoders (VAEs). These models are capable of creating images from random noise or existing data. 2. High-Quality Output: AI-generated images are known for their high quality, often indistinguishable from humancaptured photographs. They offer exceptional detail, vibrant colors, and realistic textures.

2. **Accountability:** With transparent records and automated processes, it becomes easier to hold government officials and agencies accountable for their financial decisions

3. Image to Text Conversion: One of the key features is written-to-Image conversion, which allows users to write down thoughts, concepts, or scenarios, and the AI system converts these written descriptions into visually appealing visuals. It is an invaluable tool for creative professionals and organizations trying to speed content development since it acts as a bridge between text and pictures.

4. Various Outputs: AI picture generating systems are prevented from continually producing the same image for textual inputs that are identical by having diverse outputs. Rather, they provide a variety of pictures that represent various readings of the given information, encouraging imagination and investigation.

Benefits of AI-Powered Image Generation:

1. **Creativity Empowerment:** AI image generation tools empower artists, designers, and creative professionals with powerful aids for their work. They can use AI to spark and enhance their creative ideas**.**
2. **Efficiency and Time Savings**: Automation in image generation can significantly reduce the time and effort required for content creation. This is particularly useful for marketing and advertising campaigns where large volumes of visual content are needed.
3. **Scalability:** AI-powered image generation is highly scalable, making it suitable for projects that require vast amounts of visual content, such as e-commerce websites.
4. **Customization**: AI enables the creation of tailored images, aligning with specific preferences or requirements. This is particularly valuable in personalized marketing and content creation.

2. Literature survey:

[1] Goodfellow, I., et al. (2014). "Generative Adversarial Nets." In Advances in Neural Information Processing Systems. GANs are foundational to AI-powered image generation. This paperintroduced GANs, a framework where two neural networks (generator and discriminator) compete to produce realistic images.One of the central ideas in information

theory is entropy, which measures the level of uncertainty or randomness in a random variable.

[2] Gatys, L. A., et al. (2016). "Image Style Transfer Using Convolutional Neural Networks." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.Style transfer techniques using neural networks have been instrumental in creating artistic and stylized images from existing one core concept of the paper involves using deep neural networks, specifically CNNs, to separate and manipulate the content and style of images. By extracting feature representations at different layers of the network, the authors demonstrated that it is possible to separate the content (the underlying structure and objects in an image) from the style (the textures, colors, and artistic patterns) of an image

[3] Kingma, D. P., & Welling, M. (2013). "Auto-Encoding Variational Bayes." In International Conference on Learning Representations (ICLR).VAEs are used for image generation and reconstruction. This paper introduces the concept of variational autoencoders for generative tasks. The paper by Kingma and Welling, "Auto-Encoding Variational Bayes," published in 2013, marks a significant advancement in the field of generative models and has had a profound impact on image generation and reconstruction funds using blockchain technology. The current system, of tracking is very problematic and provides needy people with a service that is sometimes difficult to track, depriving them. In this case, we use blockchain cryptography and transaction security at every stage while maintaining transparency so that every transaction is backed up with proof of its authenticity.tasks. It introduces the concept of Variational Autoencoders (VAEs), a probabilistic generative framework that combines deep learning with Bayesian inference.

[4] Mirza, M., & Osindero, S. (2014). "Conditional Generative Adversarial Nets." arXiv preprint arXiv:1411.1784. GANs allow for controlling the generated images based on specific inputs or attributes, which is valuable in various applications. The paper presents a loss function that combines the content loss, style loss, and total variation loss. The content loss ensures that the generated image maintains the same content as the reference image, while the style loss captures the statistical properties of the reference style image. The total variation loss helps reduce noise in the generated image.

[5]Ienca, M., & Vayena, E. (2019). "The Global Landscape of AI Ethics Guidelines." Nature Machine Intelligence, 1(9), 389- 399.As AI-generated images raise ethical concerns, thispaper provides insights into the ethical guidelines and considerations surrounding AI technologies.In their paper, "The Global Landscape of AI Ethics Guidelines," Ienca and Vayena provide a comprehensive overview of the ethical guidelines and considerations that surround the rapidly evolving field of artificial intelligence (AI). The authors analyse a wide array of guidelines and recommendations from various sources, shedding light on the global landscape of AI ethics. The central theme of the paper is the critical importance of ethical considerations in the development and deployment of AI technologies. It highlights the need for responsible AI development that takes into account potential societal, legal, and moral implications.

[6]Esteva,A.,etal.(2017). "Dermatologist-level classification of skin cancer with deep neural networks." Nature, 542(7639), 115-118.AI-generated images have practical applications in medical diagnosis, such as the classification of skin cancer using deep neural networks.

3. Existing System:

Current AI-driven image generation systems show an interesting combination of creativity and artificial intelligence. These systems are made to convert written descriptions into creative and colourful visual content. They use modern machine learning methods to bring the ideas and concepts contained in words to

life, such as generative models like GANs (Generative Adversarial Networks). These systems promise to simplify the creative process and provide new levels of interactive features, available to a wide range of applications from e-commerce and content creation to art and design. These current systems have a lot of promise, but they also have problems with reality, context interpretation, ethical issues, and other things. Some existing systems are:-

1. **DALL-E:-** Developed by OpenAI, DALL-E is an innovative model capable of generating images from textual descriptions. It extends the capabilities of GPT-3 to create images that correspond to natural language inputs, producing imaginative and contextually relevant visuals.

2. **CLIP:-** Another creation from OpenAI, CLIP is a model that understands images and text together. It can search and generate images based on textual prompts and is versatile in tasks.

3.**GAN-based Models** : -Generative Adversarial Networks (GANs) have been widely used in AI image generation. Models like BigGAN and StyleGAN2 have demonstrated the ability to create high-resolution and realistic images from textual prompts.

Runway ML : -Runway ML is a platform that offers various creative AI tools, including text-to-image synthesis. It enables artists and designers to experiment with AI-generated visuals and integrate them into their projects.

5.Artbreeder : Artbreeder is an online platform that uses GAN technology to allow users to create and explore AI-generated art. Users can blend images and apply textual prompts to generate custom artworks.

Problems with existing systems: -

1. Data privacy and security concerns arise due to the substantial data access required by image generation systems.
2. It's still difficult to achieve realism and consistency, particularly in complex scenes and pictures with a high resolution
3. AI systems find it difficult to understand the confusion, details, and context of textual descriptions.
4. AI image generation is less accessible to people with low hardware resources due to its high computational requirements.
5. Increasing both speed and complexity for large-scale, real-time applications like e-commerce and video games.
6. Overfitting problems may limit creativity and result in the same responses to similar text prompts.

4. Proposed System

User Module: In this system, the user will send a request to the system in the form of text to generate the image that he/she requires .The system then process this text understands it with the help of natural language processing (NLP) .It translates the textual data into a machine-friendly language — numerical representations. The field is divided into three parts:-

1. Speech Recognition –Translation of spoken language into text.
2. Natural Language Understanding (NLU) - The computer's ability to understand what we say.
3. Natural Language Generation (NLG). The generation of natural language by a computer. The AI image generator uses this numerical representation as a directions map.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 1, January : 2024

Figure 5.1. ER Diagram

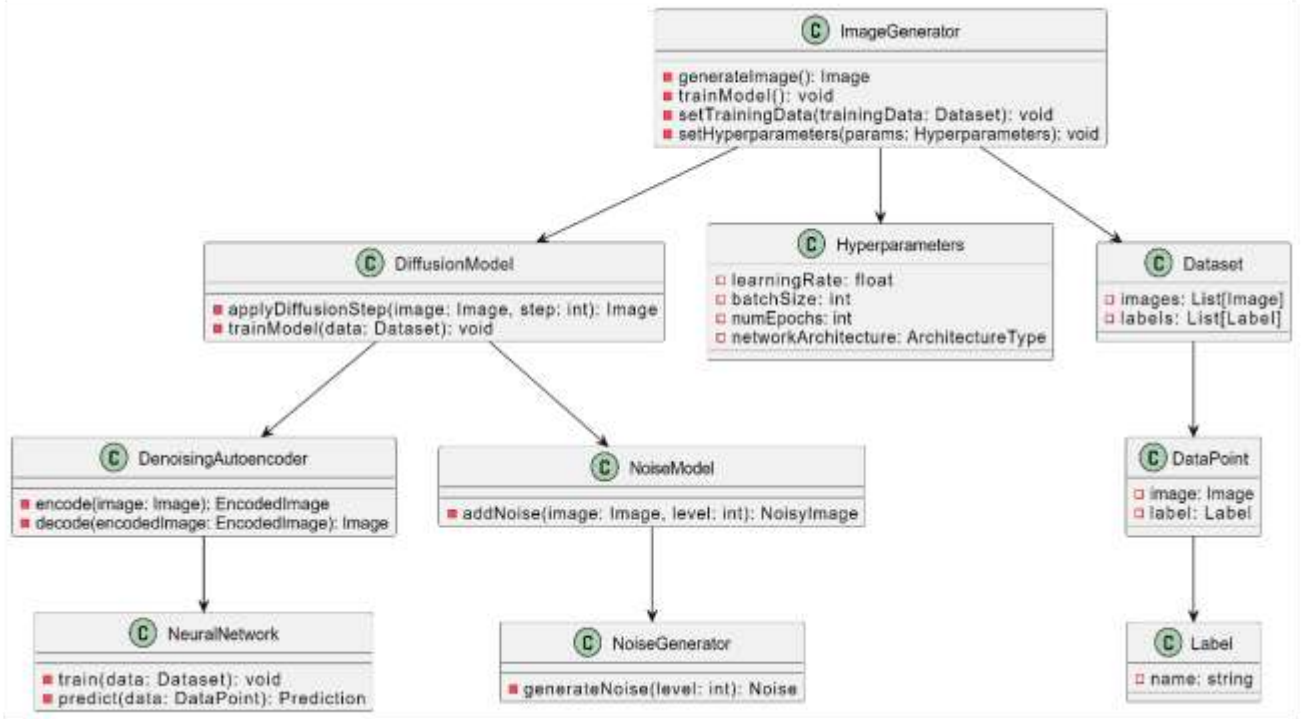


Figure 5.2. Class Diagram

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 1, January : 2024

image while it is being created. It acts as a guide that tells the AI on what elements to include in the picture and how they should work together. It then uses the algorithm of Generative Adversarial Networks, commonly called as GANs this concept was derived by Ian Goodfellow .Two fundamental parts, or sub models, make up GANs: Fake samples are produced by the generator neural network. It creates fake input data by using a random input vector, which is a list of mathematical variables with unknown values. As a binary classifier, the discriminator neural network performs its task. It determines whether a sample is generated by the generator or real by using it as input. The generator is updated to improve performance whenever the discriminator correctly classifies a sample, declaring it the winner. On the other hand, the generator is considered the winner and the discriminator is modified if it is successful in fooling the discriminator. When the generator creates a convincing sample that not only fools the discriminator but is also challenging for humans to distinguish, the procedure is considered successful. Labelled data is useful because it provides the discriminator with a reference for genuine images, which is necessary for it to analyse the generated images effectively. The discriminator is fed real images and images produced by the generator (designated as fake) during training. The "ground truth" that allows a feedback loop is this marked dataset. The feedback loop makes it easier for the discriminator to learn how to tell genuine pictures from fake ones. The generator gets feedback on how successfully it tricked the discriminator at the same time, and it utilizes this information to enhance image generation. The cycle of the game never ends as the discriminator becomes more adept at spotting fakes. there are various modules like government, users, and various types of departments. In our system, there are 2 main modules i.e., Admin (Government) and User. Admin (Government) Module: The government provides the requested funds to the user. User Module: In this system, the user will request the funds according to their needs and also, and they can check their transaction 3. Develop the smart contracts that will automate fund disbursement based on predefined rules and conditions.

Successfully it tricked the discriminator at the same time, and it utilizes this information to enhance image generation.

The cycle of the game never ends as the discriminator becomes more adept at spotting fakes.

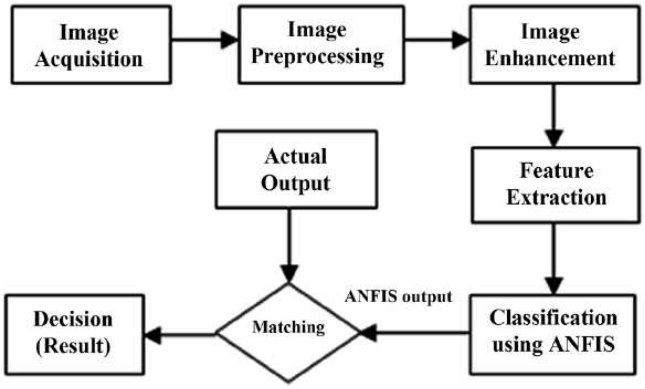


Figure 4.1

5. System Design

5.1 System Architecture:

Based on the research papers and the context provided, it seems you are interested in a system architecture for AI image generation using a stable diffusion model. Here's a high-level system architecture that combines concepts from the mentioned papers to create an AI-powered image generation system using a stable diffusion model

A Stable Diffusion model is a type of diffusion model that generates images by iteratively de-noising a latent image representation. The data pipeline can be implemented as a distributed system using a variety of tools and technologies. For example, Apache Spark can be used to collect and clean the image-text pairs, while Apache Hadoop can be used to store the pre-processed data.The data pipeline should also include a pre-processor that converts the image-text pairs into a format

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 1, January : 2024

Sequence Diagram

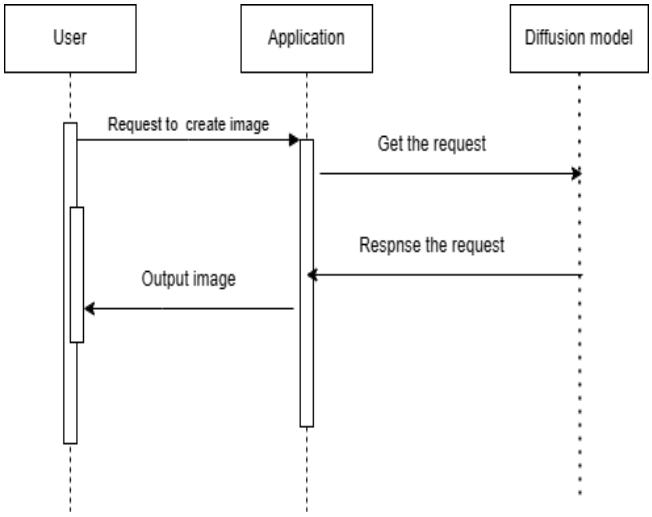


Figure 5.2.2: Sequence Diagram

6. Algorithm Detail:

Algorithm 1 Generative Adversarial Networks (GANs): GANs consist of two neural networks, the generator and the discriminator, engaged in a competitive process. The generator creates synthetic data, in this case, images, from random noise, aiming to produce images that are indistinguishable from real ones. The discriminator, on the other hand, learns to differentiate between real images from the dataset and the fake images produced by the generator. Through adversarial training, the generator refines its ability to create increasingly realistic images, while the discriminator improves its capacity to correctly classify real and generated images. This iterative process continues until the generator generates images that are convincing enough to deceive the discriminator.

Input: Random noise vector z

Training Procedure:

1. Initialize generator G and discriminator D networks with random weights.
2. **Generator Training:**
   - Generate fake samples G(z) from random noise z.
   - Compute generator loss using log(1 - D(G(z))).
   - Update generator weights to minimize the loss.
3. **Discriminator Training:**
   - Train D to correctly classify real and generated samples.
   - Compute discriminator loss using -(log D(x) + log(1 - D(G(z)))).
   - Update discriminator weights to minimize the loss.
4. Repeat steps 2-3 until convergence.

Output: Trained generator G capable of generating realistic data samples.

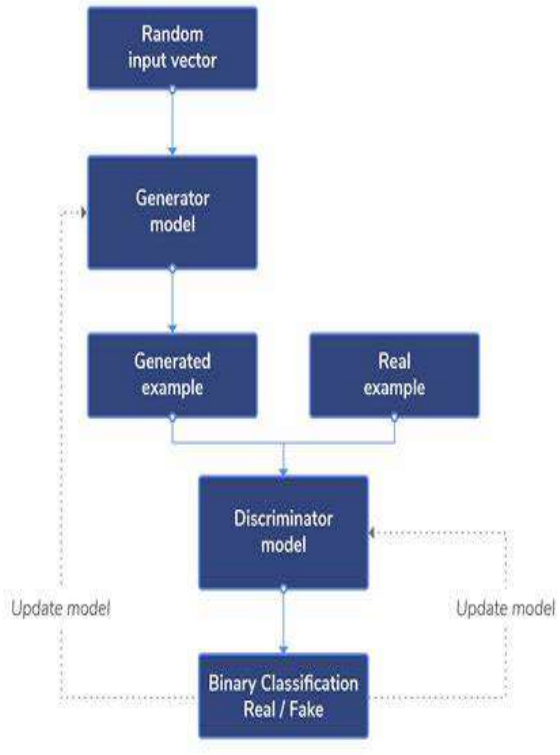


Fig 6.1: flow chart for Algorithm 1 Generative Adversarial Networks

Algorithm 2 Variational Autoencoders (VAEs): VAEs are generative models that consist of two main components: an encoder and a decoder. The encoder maps input images to a probabilistic latent space, where each point represents a compressed representation of

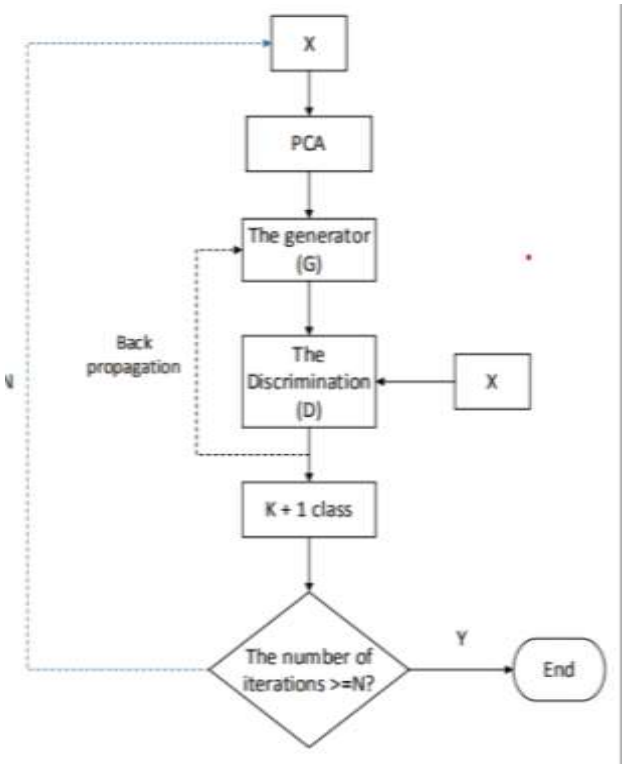
the input image. Latent space sampling involves selecting points from this learned distribution. The decoder then reconstructs the input images from these sampled points. VAEs are trained by minimizing the reconstruction error, ensuring that the generated images are similar to the input data, and by encouraging the latent space to follow a specific probability distribution, typically a Gaussian distribution. This approach allows for both image generation and latent space interpolation, enabling smooth transitions between different image styles.

Input: Input data x

Training Procedure:

1. Initialize encoder E and decoder D networks with random weights.
2. **Encoder Training:**
   - Encode input x to obtain mean (μ)and variance (σ) in latent space.
   - Sample latent vector z from N (μ , σ ^2)).
3. **Decoder Training:**
   - Reconstruct input x from sampled z using D(z).
   - Compute reconstruction loss using a suitable metric (e.g., mean squared error).
   - Regularize latent space using KL divergence.
4. Update encoder and decoder weights to minimize the total loss.
5. Repeat steps 2-4 until convergence.

Output: Trained encoder E and decoder D for encoding and reconstructing input data.



Algorithm 3 Deep Convolutional GANs (DCGANs): DCGANs are a variant of GANs that use convolutional neural networks for both the generator and discriminator architectures. Unlike traditional GANs, DCGANs replace fully connected layers with convolutional layers, enabling the networks to process spatial information effectively. Additionally, DCGANs incorporate batch normalization, which normalizes the inputs of each layer, stabilizing and accelerating the training process. They also avoid max-pooling layers, using strided convolutions for down sampling, and employ leaky ReLU activation functions to prevent the "dying ReLU" problem, ensuring that gradients flow even for negative inputs. These modifications enhance the stability and convergence of the training process, enabling the generation of high-quality images.

Input: Random noise vector (z)

Training Procedure:

1. Initialize DCGAN generator G and discriminator D networks with convolutional layers.
2. **Generator Training:**

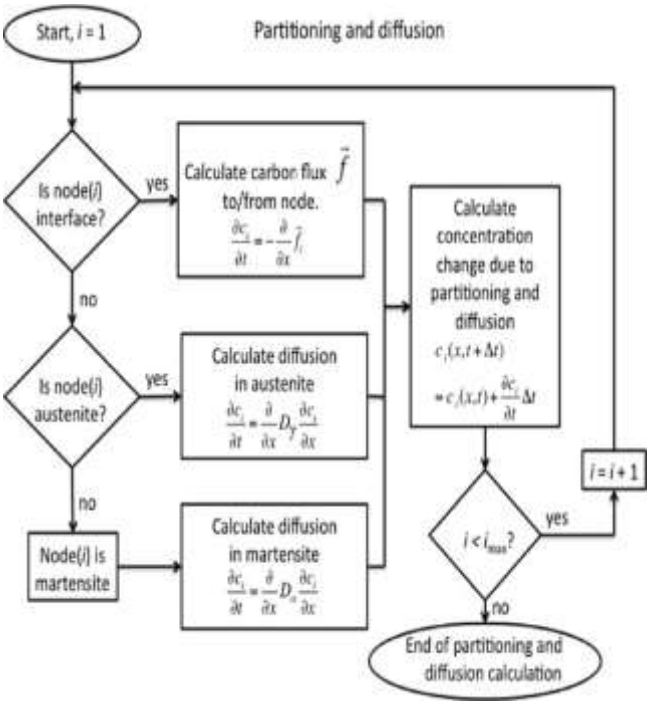
Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 1, January : 2024

- Generate synthetic samples G(z) from random noise z .
- Compute generator loss using suitable objective function.
- Update generator weights using backpropagation.

3. **Discriminator Training:**
- Train D to classify real and generated samples accurately.
- Compute discriminator loss using appropriate objective function.
- Update discriminator weights using backpropagation.

4. Use techniques like batch normalization and leaky ReLU for stability.
5. Repeat steps 2-4 until convergence.

Output: Trained DCGAN capable of generating high-quality images from random noise.



7. Conclusion:

A new era of creativity and innovation could be brought in by AI-powered image generation utilizing text. It provides a dynamic link between written descriptions and attractive visual content, with a wide range of industry applications. However, a number of obstacles prevent this technology from reaching its full potential. Ongoing challenges include the need for highquality training data, legal steps to prevent the creation of inappropriate or harmful content, and the goal of achieving the highest level of realistic and context understanding in generated images. Achieving a perfect balance between control over the user and simplicity of use is crucial for allowing a wide range of users to effectively utilize the technology. It promises to be an important instrument for professionals in a variety of fields as it develops further, building creativity, efficiency, and interaction in the design and creation of content. For this technology to be responsibly and economically applied in a variety of industries, it will be important that the problems and moral issues surrounding it are addressed.

8. Future Scope:

1. AI-powered image generation will provide designers and artists with the means to more quickly and easily realize their imaginative concepts.
2. Companies can use AI to create personalized visual content based on user preferences, which improves customer engagement and personalization.
3. The creation of lifelike environments and objects in virtual and augmented reality experiences will be greatly aided by AIgenerated images.
4. AI-generated medical images have the potential to advance medical research and diagnosis, treatment planning, and clinical practice.
5. AI-generated assets will continue to benefit the gaming industry, boosting gameplay and improving graphics.

9. References:

[1] Goodfellow, I., et al. (2014). "Generative Adversarial Nets." In Advances in Neural Information Processing Systems.

[2] Gatys, L. A., et al. (2016). "Image Style Transfer Using Convolutional Neural Networks." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition

[3] Kingma, D. P., & Welling, M. (2013). "AutoEncoding Variational Bayes." In International Conference on Learning Representations (ICLR).

[4] Mirza, M., & Osindero, S. (2014). "Conditional Generative Adversarial Nets." arXiv preprint arXiv:1411.1784.

[5] Jobin, A., Ienca, M., & Varena, E. (2019). "The Global Landscape of AI Ethics Guidelines." Nature Machine Intelligence, 1(9), 389-399.

[6] Yu, J. Malaeb and W. Ma, "Architectural Facade Recognition and Generation through Generative Adversarial Networks," 2020 International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE).

[7] Z. Zhao and X. Ma, "A Compensation Method of Two-Stage Image Generation for Human AI Collaborated In-Situ Fashion Design in Augmented Reality Environment," 2018 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), Taichung, Taiwan, 2018, pp. 76-83, doi: 10.1109/AIVR.2018.00018.

[8] Elgammal, A., Liu, B., Elhoseiny, M., & Mazzone, M. (2017). "CAN: Creative Adversarial Networks, Generating 'Art' by Learning About Styles and Deviating from Style Norms." arXiv preprint arXiv:1706.07068.

[9] Zhu, J. Y., et al. (2017). "Unpaired Image-to-Image Translation Using CycleConsistent Adversarial Networks." In Proceedings of the IEEE International Conference on Computer Vision.

[10] VZhang, H., et al. (2018). "The unreasonable effectiveness of deep features as a perceptual metric." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

[11] Russakovsky, O., et al. (2015). "ImageNet Large Scale Visual Recognition Challenge." International Journal of Computer Vision, 115(3), 211-252.

[12] Z. Tan et al., "Efficient Semantic Image Synthesis via Class-Adaptive Normalization," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 44, no. 9, pp. 4852-4866, 1 Sept. 2022, doi: 10.1109/TPAMI.2021.3076487.

[13] M. Z. Khan et al., "A Realistic Image Generation of Face From Text Description Using the Fully Trained Generative Adversarial Networks," in IEEE Access, vol. 9, pp. 1250-1260, 2021, doi: 10.1109/ACCESS.2020.3015656. [14] Selvaraju, R. R., et al. (2017). "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization." In Proceedings of the IEEE International Conference on Computer Vision.

[15] Chen, C. L., & Zhang, L. (2018). "Imagebased Fruit Detection with Deep Convolutional 21 Neural Networks." Computers and Electronics in Agriculture, 151, 275-282.

[16] Jin, H., et al. (2018). "Towards Automated Driving in Cities Using Close-to-Market Sensors: An Overview of the V-Charge Project." IEEE Transactions on Intelligent Vehicles, 3(1), 4-19.

[17]FineGAN: Unsupervised Hierarchical Disentanglement for Fine-grained Object

Generation and DiscoveryKrishna Kumar Singh*, Utkarsh Ojha*, Yong Jae Lee CVPR 2019

[18] C. Li, J. Cao and X. Zhang, "Design and Implementation of an Infrared Image Generative Model," 2020 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA), Dalian, China, 2020, pp. 1338-1345, doi:10.1109/ICAICA50127.2020.918256.

[19] Z. Ji, W. Wang, B. Chen and X. Han, "Textto-Image Generation via Semi-Supervised Training," 2020 IEEE International Conference on Visual Communications and Image Processing (VCIP), Macau, China, 2020, pp. 265-268, doi: 10.1109/VCIP49819.2020.9301888.

[20] Neural Discrete Representation Learning A. van den Oord, O. Vinyals, K. Kavukcuoglu 2017.

[21] Scaling Autoregressive Models for Content Rich Text-to-Image Generation (2017).