Survey on Image synthesis using GANs and Diffusion Models

|  |  |  |  |
| --- | --- | --- | --- |
| Abhishek S A and Sahana N D | Namitha K, Puneeth P and Prof. Rashmi | Namitha K, Puneeth P and Prof. Rashmi |  |
| *Department of Artificial Intelligence and Machine Learning*  *Dayananda Sagar College of Engineering*  *Bengaluru, Karnataka, India*  {abhishek.s.a2201 & sahanarajgowda1801}@gmail.com | *Department of Artificial Intelligence and Machine Learning*  *Dayananda Sagar College of Engineering*  *Bengaluru, Karnataka, India*  {12namithadsi & puneeth2417 & rashmitheertha25}@gmail.com | *Bengaluru, Karnataka, India*  {12namithadsi & puneeth2417 & rashmitheertha25}@gmail.com |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

***Abstract* – *Image synthesis has been a fundamental challenge in computer vision for decades. Significant progress in the field of image synthesis has been observed with the emergence of Generative Adversarial networks (GANs) and various diffusion models.***

***Index Terms -* Generative Adversarial Networks (GANs), Diffusion models**

I. Introduction

Image synthesis is a field of computer vision that aims to generate high-quality realistic images from different inputs, such as random noise, low-resolution image, or incomplete image. Generative Adversarial Networks (GANs) and diffusion models are some powerful techniques that have revolutionised the field of image synthesis to generate realistic images that are indistinguishable from real images.

Generative Adversarial Networks (GANs) are a generative model based on deep learning, which consists of two neural networks, a generator and a discriminator, that are trained in an adversarial manner. Diffusion models are probabilistic models that estimated the likelihood of generating an image from the input noise distribution.

Diffusion models differ from GANs with respect to how the two different models are trained, as diffusion models do not use an adversarial training process, but a simple sampling method that can generate high-resolution images.

In this literature survey, we discuss GANs and diffusion models and the various techniques used to improve the quality of these generated images, along with the recent works utilising GANs and diffusion models for various applications.

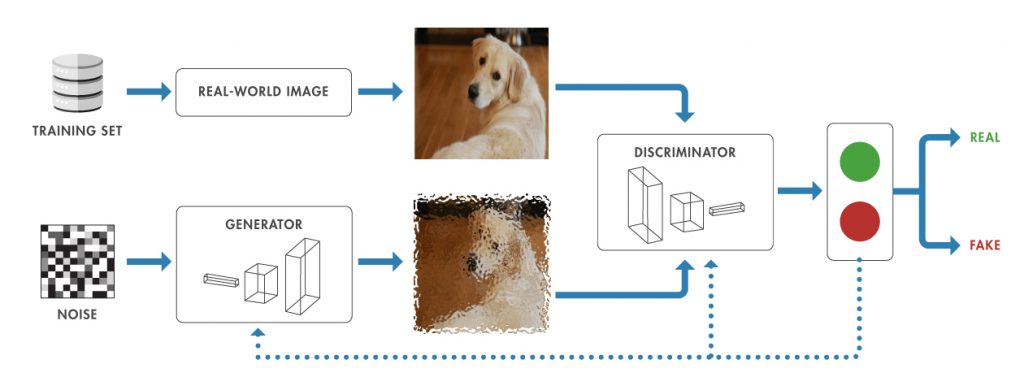
Lastly, we discuss the challenges one might face while synthesising of the images using GANs and diffusion models and their future scope.

Through this literature Survey, we hope to provide valuable insights into the current state-of-the-art image synthesis techniques and promote further research in this rapidly growing field.

II. Generative Adversarial Networks(GANs)

Generative Adversarial Networks (GANs) is a deep learning model that can generate new and previously unseen images and data by learning from existing data. GANs consist of two neural networks, the generator and the discriminator. The Generator generates images and the discriminator tries to discriminate between the generated images and the real images. Through this adversarial training process, the generator learns to generate realistic and high-quality images which can deceive the discriminator, while at the same time the discriminator get better at distinguishing between the synthetic and real images.

GANs can produce images that are higher in quality with rich details and texture when compared to traditional generative models. The applications of GANs are not just limited to image generation, they can also generate different types of data formats like video and music.



There are different types of GANs that have been developed, each with its own strengths and weaknesses, some of them are:

1. *Vanilla GANs:*

This is the basic GAN architecture, where the generator and the discriminator are feed-forward neural networks.

While they are easy to implement, they can suffer from instability during training and may not generate high-quality images.

1. *DCGANs (Deep Convolutional GANs):*

This architecture of GAN use convolutional neural networks for both the generator and the discriminator, which results in better image quality. DCGANs also have better stability during training

1. *WGANs (Wasserstein GANs):*

This architecture is a modification of the GANs that use the Wasserstein distance instead go the traditional binary cross entropy loss function, which results in more stable training and better-quality image.

1. *CycleGANs:*

This architecture is used for image-to-image translation, where the model learns to map an image from one domain to another, without paired data. For example, they can convert an image of a horse to an image of a zebra or a summer landscape to a winter landscape.

The main disadvantages of GANs is the requirement of large amounts of training data and high computational power or resources. They can also suffer mode collapse, which means that the generator starts producing a limited set of output images that are repeated. Despite the challenges faced in utilising GANs, they have the potential to revolutionised the field of computer vision and transform various industries such as the entertainment, healthcare and manufacturing.

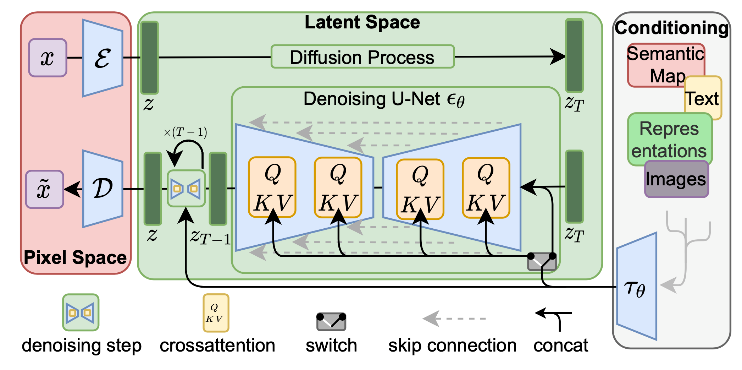
III. diffusion models

A probabilistic diffusion model is another name for a diffusion model. The primary objective of the diffusion model is to produce samples that, after a predetermined time, correspond to the data. Similar to probabilistic models, which examine and forecast the behaviour of systems that change over time, these models are taught via variational inference.

The mathematical framework used to represent how systems change between distinct states is the parameterized Markov chain. The system's present state will determine the probabilistic transitioning of particular states. Variational interference is used in the model's training process, along with intricate computations involving probability distributions, to identify the Markov chain's ideal parameters in relation to the observed data.

The diffusion models can produce samples that are close to the observed data after training. These samples show potential outcomes that the system might experience over a specific time period with a given probability.

The model generates a variety of samples and evaluates their likelihood before predicting how the system will behave in the future.



Three basic mathematical structures make up the diffusion models. These frameworks operate on the idea that fresh samples are created by first adding noise and then taking it away.

1. *Denoising Diffusion Probabilistic Models (DDPMs):*

DDPMs are generative models that emphasise high-quality sample production and noise removal from the data. They employ Markov chains taught with variational inference to simulate diffusion processes. By gradually converting noisy inputs into clean outputs, DDPMs provide realistic sample images that resemble the underlying data distribution.

1. *Noise-Conditioned Score-Based Generative Models SGMs):*

SGMs are used for the generation of new samples from the existing given distribution. They incorporate noise inputs for conditioning and modelling the score function for high-quality sample generation. SGMs generate flexible samples by capturing complex data distribution.

1. *Stochastic Differential Equations (SDEs):*

The description of system evolution in the presence of random noise is done through SDEs. It is a differential equation whose coefficients are functions of variables or random values. Since SDEs permit the coefficients of equations to depend on stochastic processes (also known as random variables) they are regarded as suited for the description of external noise.

Applications for the diffusion model can be found in the industrial, educational, and medical sectors. Diffusion models offer ways to insert artificial intelligence (AI)-generated frames between genuine frames in order to create realistic high-resolution films. Its multimodal generation applications, such as text-to-image generation, scene graph-to-image generation, text-to-3D generation, text-to-motion generation, and others, have been expanded.

It is also used in a variety of sectors, including drug design, life science (to improve the production of molecules or proteins), material design, medical image reconstruction, and other areas.

IV. TECHNIQUES FOR IMPROVING IMAGE

SYNTHESIS:

The task of image Synthesis for generative models has been a difficult challenge because of its complicative visual information. The various techniques developed to subdue the limitations of conventional generative models and to enhance the image quality are as follows:

1. *Progressive growing:*

In this technique the image quality is upgraded by permitting the model to learn the features at lower level prior to complex features. This requires cautious increase from low resolution to high resolution of generated images during the training process.

1. *Self-attention:*

In this methodology, during image generation, the model focuses on applicable regions of the input. This technique allows the model to learn spatial relationships between different parts of the image, thus enhancing the generated image quality.

3. *Style transfer:*

Style transfer is a mechanism of blending the content image and style reference image while retaining the core elements of the content image in the output. It is employed to obtain high-quality images with particular styles.

4. *GAN inversion:*

The GAN inversion method inverts the networks of the generator and discriminator aiming to rebuild an input image from the generator's latent space. This technique generates high-quality images from low-resolution inputs.

5. *Data augmentation:*

Data augmentation is a method of increasing the number of data present in the existing dataset by performing techniques like scaling, cropping, flipping, padding or rotating. This increase's the model's accuracy and ability to generalize by preventing it from overfitting.

6. *Conditional GANs:*

Conditional GANs permit the conditioning of the network to create images based on certain conditions, where both the generator and discriminator have conditioned networks.These specifications may also include class labels and input image descriptions. Some of the applications of CGANs are text-to-image synthesis and video generation based on input description.

These techniques have expanded the scope of generative models and significantly improved the quality of image synthesis. They have also enabled the development of new applications.

V. APPLICATIONS OF IMAGE SYNTHESIS

In a number of industries, including entertainment, medicine, education, product creation, and the arts, image synthesis is crucial. Given here are some of the applications of image synthesis in the above mentioned fields:

1. *Gaming:*

In the video game and entertainment industries, image synthesis is utilised to generate realistic settings, characters, and objects. It lets game designers to produce top-notch graphics, animations, and special effects, creating games that are more immersive and entertaining.

1. *Virtual and augmented reality:*

Realistic 3D models and environments are created via image synthesis in virtual and augmented reality apps, enabling users to interact with virtual objects and situations in a more organic and intuitive manner.

1. *Healthcare:*

Medical imaging applications use image synthesis to create synthetic images that can help with diagnosis, treatment planning, and surgical simulation. For instance, artificial images can be created to model how certain medical procedures will affect a patient's anatomy.

1. *Education:*

Interactive learning materials like virtual labs and simulations can be made using image synthesis in the classroom. It makes it easier for pupils to interact and visualise difficult ideas.

1. *Manufacturing:*

To simulate and optimise production processes, such as assembly line operations and product design, image synthesis is utilised in the manufacturing industry. Before actually adopting them, it enables manufacturers to pinpoint potential problems and improve processes, which lowers repair costs and boosts productivity.

1. *Art and content creation:*

Image synthesis is used to create fresh and distinctive visual resources in creative fields like art and content development. It gives designers and artists the freedom to experiment with various aesthetics and production methods, resulting in more varied and creative material.

1. *Security and surveillance:*

To produce realistic images and videos that may be utilised for training and testing, image synthesis is employed in security and surveillance applications. For example, synthetic images can be generated to simulate different security scenarios and test the effectiveness of security systems.

In general, there are many uses for image synthesis, and more fields are anticipated to be affected by it in the future. Its capacity to provide realistic and varied visual content will keep inspiring innovation and creativity across a range of industries.

VI. CHALLENGES FACED

Despite major advancements in image synthesis, there are still a number of issues that must be resolved before generative models may be made more effective and of higher quality. The main difficulties encountered during picture synthesis are listed below:

1. *Stability and mode collapse*:

One of the key difficulties with GANs is maintaining stability throughout training and preventing mode collapse, in which the generator only produces a small number of outputs. Poor diversity and quality of the generated photos may result from this.

1. *Generalization:*

A further difficulty is generalising the technique to produce excellent photos in a variety of domains and styles. Current models are frequently restricted to certain domains, and it might be difficult to generalise them to new domains.

1. *Latent space exploration:*

This crucial component of generative models enables users to direct and control the generation process. The latent space can, however, be difficult to comprehend and manage, leading to outputs that are unexpected and unrealistic.

1. *Computational cost:*

Image synthesis is computationally expensive, making it difficult to scale up to big datasets and complicated models. It's difficult to lower the computational expense of generative models without sacrificing output quality.

VII. FUTURE SCOPE

To address these challenges, several directions for future research have been proposed, including:

1. *Developing more stable and efficient training algorithms:*

Research efforts are concentrated on creating more reliable and effective training algorithms in order to enhance the functionality and convergence of GANs.

1. *Enhancing the quality and diversity of generated images:*

New architectures, loss functions, and optimisation methods are being investigated in order to enhance the quality and diversity of generated images.

1. *Incorporating domain knowledge:*

Domain knowledge can help generative models produce high-quality images across a range of domains and aesthetics.

1. *Developing more interpretable models*:

It is possible to increase the control and interpretability of the generation process by creating more interpretable models, which will make it simpler for users to modify and control the output.

1. *Reducing computational cost:*

Future research will also concentrate on ways to make generative models less computationally expensive without sacrificing output quality. This entails creating architectures with greater efficiency and researching hardware acceleration strategies.

CONCLUSION

The use of GANs and diffusion models for picture synthesis has advanced significantly in recent years, with cutting-edge models producing images of amazing quality. Both GANs and diffusion models have particular advantages and disadvantages, and how they are used depends on the use case and requirements. Progressive growth, attention mechanisms, and self-attention mechanisms are only a few of the methods that have been created to enhance the effectiveness and quality of image synthesis. These methods have produced more diversified and lifelike images with encouraging outcomes.

Applications for image synthesis can be found in a variety of fields, including security, manufacturing, healthcare, education, and entertainment. The ability to provide realistic and varied visual content will keep inspiring innovation and creativity across a range of industries.

The stability and mode collapse, generalisation, latent space exploration, and computing cost issues still exist in image synthesis despite its advancements.

Future studies will concentrate on creating training algorithms that are more reliable and effective, enhancing the quality and variety of the generated images, incorporating domain expertise, creating more interpretable models, and lowering the computational cost of generative models.All things considered, picture synthesis using GANs and diffusion models is a fascinating and quickly developing topic with lots of room for further study and invention.

References

1. Goodfellow, Ian J, Pouget-Abadie, Jean Mirza, Mehdi Xu, Bing Warde-Farley, David Ozair, Sherjil Courville, Aaron Bengio, Yoshua. “Generative Adversarial Networks” – 2014.
2. Mirza, Mehdi Osindero, Simon. “Conditional Generative Adversarial Nets” – 2014.
3. Radford, Alec Metz, Luke Chintala, Soumith. “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks” – 2015.
4. Isola, Phillip Zhu, Jun-Yan Zhou, Tinghui Efros, Alexei A. “Image-to-Image Translation with Conditional Adversarial Networks” – 2016.
5. Salimans, Tim Goodfellow, Ian Zaremba, Wojciech Cheung, Vicki Radford, Alec Chen, Xi. “Improved Techniques for Training GANs” – 2016.
6. Zhang, Han Xu, Tao Li, Hongsheng Zhang, Shaoting Wang, Xiaogang Huang, Xiaolei Metaxas, Dimitris. “StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks” – 2016.
7. Zhu, Jun-Yan Park, Taesung Isola, Phillip Efros, Alexei A. “Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks” – 2017.
8. Karras, Tero Aila, Timo Laine, Samuli Lehtinen, Jaakko. “Progressive Growing of GANs for Improved Quality, Stability, and Variation” – 2017.
9. Brock, Andrew Donahue, Jeff Simonyan, Karen. “Large Scale GAN Training for High Fidelity Natural Image Synthesis” – 2018.
10. Karras, Tero Laine, Samuli Aila, Timo. “A Style-Based Generator Architecture for Generative Adversarial Networks” – 2018.
11. Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, Björn Ommer, “High-Resolution Image Synthesis with Latent Diffusion Models” – Revised edition 2022.
12. Haziqa Sajid. “Diffusion Models in AI – Everything You Need to Know” – 2023.