e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

IMAGE GENERATION USING GAN

Shubham Kaware[*1], Rushikesh Varkale[*2], Abhishek Bhosale[*3], Aniket Wagh[*4], Prof. Deepali Lavate[*5]

[*1,2,3,4,5]Department Of Computer Engineering Smt. Kashibai Navale College Of Engineering Vadgaon (SPPU- Pune), India.

DOI : https://www.doi.org/10.56726/IRJMETS45822

ABSTRACT

This project delves into the intricate realm of generating images through the convergence of textual descriptions and existing images. Employing the prowess of Generative Adversarial Networks (GANs), the endeavor addresses the quintessential challenges in AI-fueled image synthesis. This venture bears profound significance across the landscape of computer vision, advertising, and entertainment. The primary challenges encompass scarcities in dataset availability,forging meaningful semantic bridges between text and images, and the art of rendering realism into generated images. Our mission is meticulously honed: fashioning a GAN-centric model that orchestrates the fusion of text and images, yielding high-caliber, contextually accurate visual marvels. Beyond catapulting creative workflows and automating image inception, the project propels a wave of innovation across a diverse spectrum of industries.

The methodology orchestrates the meticulous training of a generator network to conjure images and a discriminator network to discern authenticity from generated renditions. Guided by iterative training and bolstered by preprocessing techniques, the system acquires the art of fabricating images imbued with coherent narratives and aesthetic authenticity. This innovation holds the potential to reframe the contours of image creation, charting a pioneering path within AI-driven image synthesis. Keywords: GANs (Generative Adversarial Networks), Computer Vision, Generator Network, Discriminator Network, High-Quality Images.

I. INTRODUCTION

In today's digital world, a big challenge we face is combining written descriptions and pictures seamlessly. This challenge is fascinating and very important because visuals are becoming more and more crucial in how we share information. The reason this challenge is so crucial is that it helps connect what we say with what we show, making communication much better in many different areas. The main problem is making pictures that really match what we describe in words, and this is a tricky task that needs creative solutions and precise methods.

Our goal is to use advanced technology called Generative Adversarial Networks (GANs) to turn words into pictures that look incredibly real. We want to create a smart system that can make images from text in a way that is almost like looking at real photographs. This could change how we make and understand images in a big way. To achieve this, we're using a flexible and efficient way of developing our software. It allows us to make constant improvements and test our ideas quickly. This approach is essential because making images with artificial intelligence involves complex challenges.

This chapter is just the beginning of our project. In the following chapters, we'll explore why we're doing this, the specific problems we're solving, the smart solutions we're coming up with, and how all of this can make a real difference in various industries. We're laying the foundation here for a deeper understanding of what we're trying to achieve and how it could change the way we communicate visually.

II. LITERATURE SURVEY

[1]-Volume: 08 Issue: 12 e-ISSN: 2395-0056]-Deep learning based image caption generator. By M. Sailaja, K. Harika, B. Sridhar, Rajan Singh, V. Charitha, Koppula Srinivas Rao at IEEE

In the realm of artificial intelligence, an extraordinary breakthrough has been achieved through the development of an Image Caption Generator that seamlessly marries visual understanding with linguistic prowess. This cutting-edge model, rooted in the sophisticated architecture of Convolutional Neural Networks (CNN) and Long Short-Term Memory networks (LSTM), ushers in a new era of image analysis and interpretation. With a keen focus on object recognition, the CNN component meticulously dissects intricate

visual features, enabling the system to recognize objects within images with unparalleled precision. Meanwhile, the LSTM network delves into the sequential nature of these features, capturing nuanced contextual relationships and patterns. What sets this innovation apart is its ability to generate descriptive and contextually relevant captions without relying on textual prompts. Instead, it harnesses the power of images, allowing it to interpret visual content and craft eloquent, detailed, and tailored captions, marking a transformative milestone in the fusion of computer vision and natural language understanding.

At the heart of this groundbreaking system lies a synergy between artificial neural networks and human-like cognitive processes. By integrating the advanced capabilities of CNNs for feature extraction with the sequential learning prowess of LSTMs, this model transcends the limitations of traditional image captioning methods. It comprehensively understands the complex interplay of objects within images and the intricate relationships they share. The model's ability to generate contextually rich and linguistically nuanced captions solely based on visual inputs represents a quantum leap in artificial intelligence. Not only does it herald a significant stride in object recognition, but it also opens doors to a myriad of applications, from aiding the visually impaired to enhancing content understanding for diverse industries. This sophisticated fusion of computer vision and natural language processing stands as a testament to the boundless possibilities when technology and human-like cognitive functions converge in harmony.

[2]- New Style Image Generation using Neural Style Transfer By AKHIL SINGH 1 , VAIBHAV JAISWAL 1 , GAURAV JOSHI 1 , ADITH SANJEEVE 1 , SHILPA GITE 1,2, AND KETAN KOTECHA2,3 at IEEE

In recent years, significant advancements have been made in the field of Neural Style Transfer (NST), a deep learning technique that enables the transformation of images and videos into artistic styles. One fundamental concept in NST is the use of Generative Adversarial Networks (GANs), which consist of two neural networks, a Generator and a Discriminator, working in tandem to produce realistic and stylized images.

Traditional GANs, introduced by Goodfellow et al., form the basis of many NST models. These GANs generate data based on learned patterns and regularities, allowing for the creation of unique images. Deep Convolutional GANs (DCGANs) further enhance image generation by incorporating convolutional neural networks, enabling the synthesis of high- resolution images. CycleGANs, on the other hand, focus on unpaired image-to-image translation, allowing style transfer without the need for corresponding pairs of images.

In the realm of style transfer, Conditional GANs (cGANs) have been pivotal. By incorporating structured loss functions, cGANs enable the conditioning of image generation on specific input images, leading to more controlled and structured stylization. Another crucial innovation is the integration of CNN-based parametric texture models, allowing for the separation and recombination of content and style features in images. This approach utilizes the feature representations obtained from deep neural networks, such as VGG networks, to create a bridge between content and style, enabling the generation of visually appealing and artistic images.

Additionally, advancements in NST have been applied to various domains, including anime character creation and cartoonization of real-world images. Researchers have developed specialized GAN architectures, such as CartoonGAN, tailored for transforming real-world scenery images into cartoon-style representations. These innovations have not only enhanced the quality of stylized images but also expanded the scope of applications, making NST a versatile tool in the realm of computer vision and artistic expression.

[3]-Text-to-Image Generation Using Deep Learning By Sachin Tendulkar, SHIVANI SHIVANI, Tribhuvan Singh Rathore, Priyanka Mishra at IEEE

In the presented research, the authors tackle the intricate task of text-to-image synthesis employing Generative Adversarial Networks (GANs). They introduce a novel architecture called Residual GAN (RGAN) designed specifically to generate images from detailed textual descriptions. To achieve this, the researchers harness the power of Convolutional Neural Networks (CNNs) to encode images and Recurrent Neural Networks (RNNs) to process textual data. Notably, the RGAN model incorporates skip connections and learning rate decay mechanisms, ensuring stability during the training process and facilitating convergence to produce visually plausible images. The study's experimentation revolves around the Oxford-102 flowers dataset, where the model is trained using a limited number of captions to strike a balance between computational resources and image quality. The obtained results showcase the effectiveness of the RGAN architecture, demonstrating its capability to generate images that accurately correspond to specific textual queries. Despite the achievements,

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

the study acknowledges challenges and limitations, emphasizing the necessity for future research to enhance GANs' controllability and further improve the quality and realism of the generated images.

[4]A Realistic Image Generation of Face From Text Description Using the Fully Trained Generative Adversarial Networks By MUHAMMAD ZEESHAN KHAN1, SAIRA JABEEN1, MUHAMMAD USMAN GHANI KHAN2, TANZILA SABA 3,ASIM REHMAT2, AMJAD REHMAN3, USMAN TARIQ at IEEE

In the study titled "A Realistic Image Generation of Face From Text Description Using the Fully Trained Generative Adversarial Networks," the authors propose an innovative approach to text-to-face synthesis, a subset of text-to-image synthesis. This area holds significant importance in machine learning, involving complex challenges in handling language modalities, managing incomplete and ambiguous information, and integrating natural language processing techniques with computer vision algorithms. The task aims to generate natural and realistic face images based on textual descriptions, holding applications in art, design, image retrieval, and public safety, particularly in criminal investigations.

To support their methodology, the authors create a dataset by amalgamating data from LFW (Labeled Faces in the Wild), CelebA, and locally prepared datasets. The proposed GAN architecture includes a text encoder and an image decoder. The text encoder utilizes bidirectional LSTM (Long Short-Term Memory) networks to extract semantic features from input sentences, ensuring a comprehensive understanding of textual descriptions. The encoded features, along with noise vectors, are fed into the image decoder, which consists of three blocks of deconvolution layers. These layers upsample the feature maps to generate high-resolution face images.

[5] -Image Generation Using Generative Adversarial Network By Yuusuke Kataoka, Takashi Matsubara, Kuniaki Uehara at IEEE

The presented paper discusses a novel approach to image generation using deep neural networks with an attention mechanism trained using the Generative Adversarial Networks (GANs) approach. The study addresses the challenge of generating detailed images by proposing a network that attends to specific parts of the image and generates images step by step. This attention mechanism is implemented using recurrent neural networks and 2D Gaussian filters. By combining the attention mechanism with GANs, the authors aim to generate more realistic images without issues like blurriness or noise.

Overall, the paper presents a promising technique for generating detailed and realistic images, showcasing the potential of combining attention mechanisms with GANs in the field of image generation.

III. GAP ANALYSIS

Paper Name	Algorithm used	Use of Text	Use of Image	Use of Both (Text & Image)
Deep learning basedimage captiongenerator.	CNN and LSTM Algorithm	No	Yes	No
Neural Style Transfer:A Critical Review	Neural style transfer	Yes	No	No
Text-to-Image Generation UsingDeep Learning	recurrent convolutional generative adversarial network	Yes	No	No
A Realistic Image Generation of Face From Text Description Using theFully TrainedGenerative Adversarial Networks	GAN(Generative Adversarial Networks)	Yes	No	No
Image Generation Using Generative Adversarial Network	DCGAN,	No	Yes	No

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:05/Issue:10/October-2023 Impact Factor- 7.868 www.irjmets.com

IV. CONCLUSION

The emergence of advanced Generative Adversarial Networks (GAN) has ushered in a new era of AI creativity. These sophisticated models effortlessly translate textual descriptions into intricate visual representations, demonstrating the remarkable creative potential of artificial intelligence. This synthesis of AI innovation not only showcases the technology's prowess but also revolutionizes the way we perceive and interact with visual content.

One of the key strengths lies in its ability to empower creativity swiftly and efficiently. By seamlessly merging text and images, this technology accelerates creative processes, fostering innovation across various professions. Its practicality has been proven in real-world applications, especially in digital content creation and marketing. These successes validate the model's effectiveness and underline its wide-ranging applicability in today's fast-paced digital landscape.

V. FUTURE WORK

Looking ahead, the future of this technology is brimming with exciting possibilities. Enhancing training data stands out as a crucial area for development. By refining methods to curate diverse, high-quality datasets, the model's capabilities can be significantly expanded, leading to even more impressive outcomes.

Innovative applications are on the horizon, especially in critical fields like healthcare. Visual representations of complex medical data have the potential to aid diagnosis and enhance understanding, revolutionizing the way we approach healthcare challenges.

Additionally, there is a growing focus on real-time interaction. Developing systems that allow users to dynamically modify text and visualize changes instantaneously opens doors to interactive applications previously unexplored.

Ethics and responsible AI usage remain paramount. Ongoing efforts are directed towards refining algorithms, ensuring that generated content remains ethical, unbiased, and culturally sensitive. This emphasis on ethical AI fosters responsible usage, mitigating potential challenges.

Collaboration is another avenue of exploration. Integrating this technology into collaborative platforms allows multiple users to contribute text collaboratively, leading to a collective creation of visuals. Such collaborative endeavors hold immense potential in various creative and professional spheres.

Lastly, AI education is a crucial aspect of the future landscape. Developing educational tools that leverage this technology to teach AI concepts visually is on the horizon. This approach promotes deeper understanding among students and enthusiasts, democratizing AI education and making it more accessible to all.

VI. REFERENCES

[1] Deep learning based image caption generator. By M. Sailaja, K. Harika, B. Sridhar, Rajan Singh, V. Charitha, Koppula Srinivas Rao at IEEE.

[2] New Style Image Generation using Neural Style Transfer By AKHIL SINGH 1 , VAIBHAV JAISWAL 1, GAURAV JOSHI 1 , ADITH SANJEEVE 1 , SHILPA GITE 1,2, AND KETAN KOTECHA2,3 at IEEE.

[3] Text-to-Image Generation Using Deep Learning By Sachin Tendulkar, SHIVANI SHIVANI, Tribhuvan Singh Rathore, Priyanka Mishra at IEEE.

[4] A Realistic Image Generation of Face From Text Description Using the Fully Trained Generative Adversarial Networks By MUHAMMAD ZEESHAN KHAN1, SAIRA JABEEN1, MUHAMMAD USMAN GHANI KHAN2, TANZILA SABA 3,ASIMREHMAT2, AMJAD REHMAN3, USMAN TARIQ at IEEE.

[5] Image Generation Using Generative Adversarial Network By Yuusuke Kataoka, Takashi Matsubara, Kuniaki Uehara at IEEE.