E-ISSN: 2708-3977
P-ISSN: 2708-3969
IJEDC 2024; 5(1): 01-06

www.datacomjournal.com
Received: 04-11-2023
Accepted: 06-12-2023

Chinni Mohith
Department of CSE, Apex institute of Technology, Punjab, India

Jaya Venkatsh
Department of CSE, Apex institute of Technology, Punjab, India

Corresponding Author:
Chinni Mohith
Department of CSE, Apex institute of Technology, Punjab, India

Imagining the Unseen: Text-driven realism in artificial image generation

Chinni Mohith and Jaya Venkatsh

DOI: https://doi.org/10.22271/27083969.2024.v5.i1a.36

Abstract
This project employs Generative Adversarial Networks (GANs) to tackle the task of generating realistic images from textual descriptions. GANs consist of a generator and a discriminator network engaged in a competitive learning process, enabling the creation of high-quality images. By incorporating natural language processing techniques, we connect textual input to the generator, allowing for the synthesis of images that align closely with provided descriptions. Our methodology involves training the GAN on diverse datasets, optimizing for both visual fidelity and semantic coherence. Through extensive experimentation and evaluation, we showcase the model's effectiveness in transforming text into visually convincing images. This research contributes to the evolving landscape of text-to-image synthesis, demonstrating the potential of GANs in bridging the gap between language and visual representation.

Keywords: Generative Adversarial Networks (GANs), Image Super-Resolution, Deep Learning, Convolutional Neural Networks (CNNs), High-Resolution Imaging, Low-Resolution to High-Resolution

Introduction
In the realm of artificial intelligence and computer vision, the convergence of natural language processing (NLP) and image generation has yielded remarkable advancements, opening new frontiers for applications in content creation, virtual environments, and augmented reality[1] One particularly intriguing avenue of exploration within this intersection is the task of generating realistic images from textual descriptions. This research delves into this burgeoning field, leveraging the power of Generative Adversarial Networks (GANs) to bridge the gap between the expressive richness of human language and the vivid visual representation captured in images. The essence of this research lies in the marriage of two potent technologies: GANs, known for their capacity to generate authentic-looking images, and natural language processing, which enables machines to comprehend and generate human-like textual descriptions. [2] The fusion of these technologies holds the promise of transforming a textual prompt into a tangible visual output, bringing us closer to a future where machines can interpret and manifest the imaginative world conveyed through words. Generative Adversarial Networks, introduced by Ian Goodfellow and his colleagues in 2014, have since become a cornerstone in the field of deep learning. The fundamental idea behind GANs involves training a generator network to create synthetic data, such as images, while concurrently training a discriminator network to distinguish between real and generated data. [3]This adversarial training process propels the generator to refine its output iteratively until it becomes indistinguishable from real data. By incorporating GANs into the realm of text-to-image synthesis, we aim to harness their creative potential in conjuring images that resonate with textual descriptions. The textual descriptions, sourced from a diverse range of datasets, present both an opportunity and a challenge. On one hand, they offer a nuanced and rich source of information, allowing the model to learn the intricate relationships between words and visual features. [4] On the other hand, they introduce complexities associated with ambiguity, variability, and context dependence. Addressing these challenges requires a robust fusion of NLP techniques and image generation architectures within the GAN framework, pushing the boundaries of what is currently achievable. [5]

As we embark on this exploration, the potential applications of successfully generating realistic images from textual descriptions unfold across numerous domains. [6] Content creators can benefit from an intelligent tool that transforms narrative ideas into visual representations, aiding in the rapid prototyping of scenes for movies, video games, or graphic design. [7] Virtual environments can be enriched by dynamically responding to textual cues, enhancing user experiences and interactions. Moreover, augmented reality applications can leverage this technology to seamlessly integrate virtual elements into the real world, creating immersive and contextually relevant overlays. [12] This research aims not only to contribute to the growing body of knowledge in the field of text-to-image synthesis but also to underscore the transformative potential of GANs in realizing the convergence of language and visual creativity. The subsequent sections will delve into the methodology, experimentation, and evaluation processes, providing insights into the intricacies of training GANs to generate realistic images from textual descriptions and the implications of such advancements in reshaping human-computer interactions. [8]

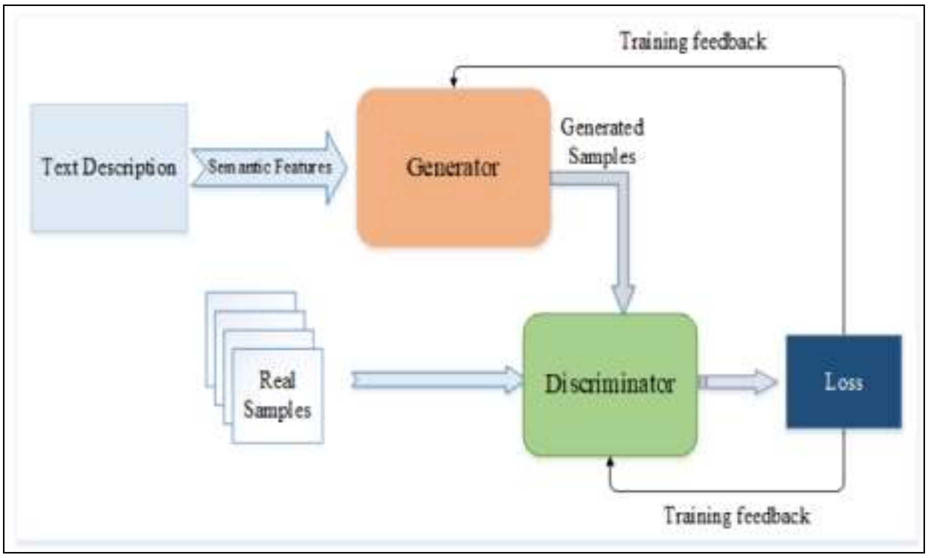


Fig 1

Literature review

image enhancement is a crucial and complex technique in image processing aimed at improving the visual quality of various types of images, such as medical, satellite, aerial, and real-life photographs. It addresses issues like poor contrast and noise, which can significantly degrade image quality. [9] Researchers and scientists have developed numerous techniques, often based on transform domain methods, to enhance digital images. However, it's important to note that some of these methods may introduce artifacts that could potentially reduce image integrity.Image enhancement is a critical technique in image processing, focusing on enhancing visual quality for Computer Vision Algorithms. [10] This paper explores its applications across various image types, including grayscale, color, infrared, and videos. [26] The primary goal is to shed light on the limitations of current image enhancement methods. By addressing these drawbacks, researchers and practitioners can work towards more effective and reliable image enhancement techniques, ultimately benefiting fields such as medical imaging, satellite analysis, and general computer vision applications, where image quality enhancement is paramount for accurate and meaningful data interpretation and analysisunderwater image degradation due to light scattering and absorption poses challenges such as reduced colors, low brightness, and indistinguishable objects. To address these issues, our proposed fusion-based underwater image enhancement technique employs contrast stretching and Auto White Balance. [11] This straightforward approach effectively enhances contrast and color in underwater images, significantly improving their visibility. By mitigating the adverse effects of water on image quality, our method offers a simple yet valuable solution for enhancing underwater imagery, with potential applications in marine research, underwater exploration, and various fields where visual clarity in aquatic environments is essential. [28]

In conclusion, the quest for an objective image quality metric that aligns with subjective perception remains a formidable challenge. Our proposed full reference image quality metric, leveraging features extracted from Convolutional Neural Networks (CNNs), presents a promising solution. By utilizing a pre-trained AlexNet model to extract and compare feature maps from test and reference images across multiple layers, we achieve a comprehensive assessment of image quality. [27] Empirical evaluations on four prominent image quality databases demonstrate that our metric either matches or surpasses the performance of ten other state-of-the-art metrics. This underscores the superiority of CNN-based features, particularly in capturing perceptual nuances that handcrafted features often miss, marking a significant advancement in image quality assessment. [13]

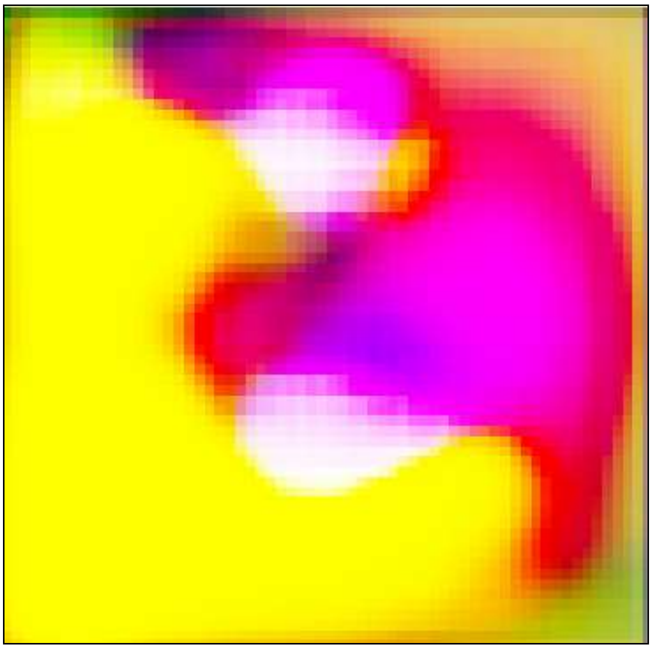
Methodology

The code loads bird images and their corresponding textual descriptions from the CUB-200-2011 dataset.It preprocesses the data, including cropping and resizing images.Model Architecture:The architecture consists of a Generator (stage1_generator), a Discriminator (stage1_discriminator), a Conditioning Augmentation Network (ca_network), and an Embedding Compressor.The Generator generates realistic images from random noise and conditioned text embeddings.The Discriminator evaluates whether an image is real or generated, taking into account the spatially replicated text embeddings.The Conditioning Augmentation Network conditions the generator by transforming the text embeddings. [38] The Embedding Compressor compresses the textual embeddings.Training:The model is trained in an adversarial manner, where the Generator aims to generate realistic images that deceive the Discriminator, and the Discriminator aims to distinguish between real and generated images. [23] The training involves multiple iterations (epochs) over the dataset.Adversarial loss functions are used to optimize the performance of both the Generator and Discriminator. [22] Loss Functions:Binary cross-entropy loss is used for the Discriminator to distinguish between real and generated images.Mean squared error loss is used for the Generator to improve the quality of generated images. [21] Adversarial loss is employed to guide the training of the Generator by considering both the image generation and conditioning augmentation.Checkpoints:The code includes a mechanism for saving model weights at regular intervals during training.Visualization:the code has provisions for visualizing the progress of the Generator using TensorBoard. [20] Testing and Saving:During training, the Generator's progress is periodically evaluated on a test set, and sample images are saved for visual inspection.Model weights are saved at specific intervals for future use or fine-tuning. [18] Final Model Save:At the end of training, the final weights of the Generator and Discriminator models are saved.This methodology represents the foundational steps involved in training the first stage of a StackGAN for generating images from textual descriptions. [19] Keep in mind that the StackGAN architecture typically involves multiple stages for progressively refining the generated images. [24]

This section describes the training details of deep learning-based generative models.Conditional GANs were used with recurrent neural networks (RNNs) and convolutionalneural networks (CNNs) for generating meaningful images from a textual description. [25] Thedataset used consisted of images of flowers and their relevant textual descriptions. Forgenerating plausible images from text using a GAN, preprocessing of textual data andimage resizing was performed. [14] We took textual descriptions from the dataset, preprocessedthese caption sentences, and created a list of their vocabulary. Then, these captions werestored with their respective ids in the list. The images were loaded and resized to a fixeddimension. [15] These data were then given as input to our proposed model. RNN was usedfor capturing the contextual information of text sequences by defining the relationshipbetween words at altered time stamps. Text-to-image mapping was performed using anRNN and a CNN. The CNN recognized useful characteristics from the images without theneed for human intervention. An input sequence was given to the RNN, which convertedthe textual descriptions into word embeddings with a size of 256. These word embeddingswere concatenated with a 512-dimensional noise vector. [16] To train our model, we took a batchsize of 64 with gated-feedback 128 and fed the input noise and text input to a generator. [17]

Results

Text-to-image processing refers to the task of generating images from textual descriptions. It is a challenging area in the field of artificial intelligence and computer vision. As of my last knowledge update in January 2022, here are some general insights into text-to-image processing GANs have been widely used in text-to-image synthesis. [31] They consist of a generator and a discriminator, where the generator creates images from text descriptions, and the discriminator evaluates the realism of the generated images. [28] The two networks are trained adversarially to improve the quality of generated images Conditional GANs take an additional input, such as a textual description, to guide the image generation process. [29] This helps in generating more specific and contextually relevant images based on the provided text.: Large datasets containing pairs of text descriptions and corresponding images are crucial for training text-to-image models effectively. Datasets like MS COCO (Common Objects in Context) and the Visual Genome dataset have been commonly used for this purpose. [30]



Conclusion

Text-to-image processing is a fascinating area within artificial intelligence and computer vision that focuses on generating images from textual descriptions. Leveraging techniques like Generative Adversarial Networks (GANs), researchers have made significant strides in this field. GANs, with their generator-discriminator architecture, are instrumental in creating realistic images based on textual prompts. Conditional GANs, a variation, enhance the process by incorporating additional inputs such as textual descriptions, enabling more contextually relevant image generation. [32] Critical to the development of robust text-to-image models is the availability of large datasets containing pairs of text and corresponding images. Datasets like MS COCO and Visual Genome play a pivotal role in training these models effectively. [33] However, evaluating the quality of generated images poses a challenge. Metrics like Inception Score, Frechet Inception Distance (FID), and Perceptual Similarity Index (PSI) aim to assess aspects such as image quality, diversity, and similarity to real images. [34] Despite notable progress, challenges persist. Achieving a balance between generating diverse and high-quality images that align accurately with textual descriptions remains an ongoing pursuit. [35] Researchers are actively addressing these challenges, exploring novel architectures, and refining training methodologies to enhance the performance of text-to-image models. [36] In conclusion, text-to-image processing has witnessed remarkable advancements, driven by the adoption of GANs and the availability of comprehensive datasets. [37] The field holds promise for applications ranging from content creation to virtual environments. As researchers continue to refine techniques and overcome challenges, the future of text-to-image processing appears dynamic, with the potential to revolutionize how we interact with and generate visual content based on textual input

References

1. Reed S, Akata Z, Yan X, Logeswaran L, Schiele B, Lee H. Generative Adversarial Text-to-Image Synthesis. arXiv preprint arXiv:1605.05396. 2016. Available from: https://arxiv.org/abs/1605.05396
2. Zhang H, Xu T, Li H, Zhang S, Huang X, Wang X, *et al*. StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV); c2017.
3. Xu J, Chao H, Wang D, Liu C. Structured Generative Adversarial Networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); c2018.
4. Zhu JY, Park T, Isola P, Efros AA. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV); c2017.
5. Johnson J, Karpathy A, Fei-Fei L. DenseCap: Fully Convolutional Localization Networks for Dense Captioning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); c2016.
6. Chen X, Mishra A, Rohaninejad M, Abbeel P. Image-

to-Image Translation with Conditional Adversarial Networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2018.
7. Han X, Wu Z, Wang Y, Cheng E. StackGAN++: Realistic Image Synthesis with Stacked Generative Adversarial Networks. IEEE Transactions on Multimedia. 2018;20(8):1943-1956.
8. Nguyen DT, Choi H, Lee KM. A Stochastic Encoder-Decoder Model for Text-to-Image Generation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); c2017.
9. Zhang Z, Xu C, Li J, Zhang X. StackGAN for Conditional Image Generation. In: Proceedings of the European Conference on Computer Vision (ECCV); c2018.
10. Xu T, Zhang P, Huang Q, Zhang H, Metaxas DN. Attngan: Fine-Grained Text to Image Generation with Attentional Generative Adversarial Networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); c2018.
11. Zhang H, Goodfellow I, Metaxas D, Odena A. Self-Attention Generative Adversarial Networks. In: Proceedings of the International Conference on Machine Learning (ICML); 2018.
12. Ssola P, Zhu JY, Zhou T, Efros AA. Image-to-Image Translation with Conditional Adversarial Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); c2017.
13. Sharma, S., Tyagi, A., Kumar, S., & Kaushik, P. (2022). Additive manufacturing process based EOQ model under the effect of pandemic COVID-19 on non-instantaneous deteriorating items with price dependent demand. In A. Editor & B. Editor (Eds.), Additive Manufacturing in Industry 4.0 (1st ed.). CRC Press.
14. Balamurugan, A., Krishna, M.V., Bhattacharya, R., Mohammed, S., Haralayya, B. & Kaushik, P. (2022). Robotic Process Automation (RPA) in Accounting and Auditing of Business and Financial Information. The British Journal of Administrative Management, 58 (157), 127-142.
15. sola P, Zhu JY, Zhou T, Efros AA. Image-to-Image Translation with Conditional Adversarial Networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2017.
16. Elgammal A, Liu B, Elhoseiny M, Marecek J. CAN: Creative Adversarial Networks, Generating" Art" by Learning About Styles and Deviating from Style Norms. arXiv preprint arXiv:1706.07068. 2017. Available from: https://arxiv.org/abs/1706.07068
17. Wang TC, Liu MY, Zhu JY, Tao A, Kautz J, Catanzaro B. High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2018.
18. Li Y, Fang C, Yang J, Wang Z, Lu X. StackGAN-v2: Realistic Image Synthesis with Stacked Generative Adversarial Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2019;41(8):1947-1962.
19. Chopra Y, Kaushik P, Rathore SPS, Kaur P. Uncovering Semantic Inconsistencies and Deceptive Language in False News Using Deep Learning and NLP Techniques for Effective Management. International Journal on Recent and Innovation Trends in Computing and Communication. 2023;11(8s):681-692. https://doi.org/10.17762/ijritcc.v11i8s.7256
20. Kaushik P. Role and Application of Artificial Intelligence in Business Analytics: A Critical Evaluation. International Journal for Global Academic & Scientific Research. 2022;1(3):01-11. https://doi.org/10.55938/ijgasr.v1i3.15
21. Kaushik P. Deep Learning Unveils Hidden Insights: Advancing Brain Tumor Diagnosis. International Journal for Global Academic & Scientific Research. 2023;2(2):01-22. https://doi.org/10.55938/ijgasr.v2i2.45
22. Kaushik P. Unleashing the Power of Multi-Agent Deep Learning: Cyber-Attack Detection in IoT. International Journal for Global Academic & Scientific Research. 2023;2(2):23-45. https://doi.org/10.55938/ijgasr.v2i2.46
23. Kaushik P, Rathore SPS, Miglani S, Shandilya I, Singh A, Saini D, Singh A. HR Functions Productivity Boost by using AI. International Journal on Recent and Innovation Trends in Computing and Communication. 2023;11(8s):701-713. https://doi.org/10.17762/ijritcc.v11i8s.7672
24. Kaushik P, Singh Rathore SPS, Kaur P, Kumar H, Tyagi N. Leveraging Multiscale Adaptive Object Detection and Contrastive Feature Learning for Customer Behavior Analysis in Retail Settings. International Journal on Recent and Innovation Trends in Computing and Communication. 2023;11(6s):326-343. https://doi.org/10.17762/ijritcc.v11i6s.6938
25. Kaushik P, Yadav R. Reliability design protocol and blockchain locating technique for mobile agent. Journal of Advances in Science and Technology (JAST). 2017;14(1):136-141. https://doi.org/10.29070/JAST
26. Kaushik P, Yadav R. Deployment of Location Management Protocol and Fault Tolerant Technique for Mobile Agents. Journal of Advances and Scholarly Researches in Allied Education (JASRAE). 2018;15(6):590-595. https://doi.org/10.29070/JASRAE
27. Kaushik P, Yadav R. Mobile Image Vision and Image Processing Reliability Design for Fault-Free Tolerance in Traffic Jam. Journal of Advances and Scholarly Researches in Allied Education (JASRAE). 2018;15(6):606-611. https://doi.org/10.29070/JASRAE
28. Kaushik P, Yadav R. Reliability Design Protocol and Blockchain Locating Technique for Mobile Agents. Journal of Advances and Scholarly Researches in Allied Education (JASRAE). 2018;15(6):590-595. https://doi.org/10.29070/JASRAE
29. Kaushik P, Yadav R. Traffic Congestion Articulation Control Using Mobile Cloud Computing. Journal of Advances and Scholarly Researches in Allied Education (JASRAE). 2018;15(1):1439-1442. https://doi.org/10.29070/JASRAE
30. Pratap Singh Rathore S. Analysing the efficacy of training strategies in enhancing productivity and advancement in profession: theoretical analysis in Indian context. International Journal for Global Academic & Scientific Research. 2023;2(2):56-77. https://doi.org/10.55938/ijgasr.v2i2.49
31. Pratap Singh Rathore S. The Impact of AI on Recruitment and Selection Processes: Analysing the role of AI in automating and enhancing recruitment and selection procedures. International Journal for Global Academic & Scientific Research. 2023;2(2):78-93.

https://doi.org/10.55938/ijgasr.v2i2.50
32. Rachna Rathore. Application of Assignment Problem and Traffic Intensity in Minimization of Traffic Congestion. IJRST. 2021;11(3):25-34. DOI: http://doi.org/10.37648/ijrst.v11i03.003
33. Rathore R. A Review on Study of application of queueing models in Hospital sector. International Journal for Global Academic & Scientific Research. 2022;1(2):01–05. https://doi.org/10.55938/ijgasr.v1i2.11
34. Rathore R. A Study on Application of Stochastic Queuing Models for Control of Congestion and Crowding. International Journal for Global Academic & Scientific Research. 2022;1(1):01–07. https://doi.org/10.55938/ijgasr.v1i1.6
35. Rathore R. A Study Of Bed Occupancy Management In The Healthcare System Using The M/M/C Queue And Probability. International Journal for Global Academic & Scientific Research. 2023;2(1):01–09. https://doi.org/10.55938/ijgasr.v2i1.36
36. Sharma T, Kaushik P. Leveraging Sentiment Analysis for Twitter Data to Uncover User Opinions and Emotions. International Journal on Recent and Innovation Trends in Computing and Communication. 2023;11(8s):162–169. https://doi.org/10.17762/ijritcc.v11i8s.7186
37. Sharma V. A Study on Data Scaling Methods for Machine Learning. International Journal for Global Academic & Scientific Research. 2022;1(1):23–33. https://doi.org/10.55938/ijgasr.v1i1.4
38. Yadav M, Kakkar M, Kaushik P. Harnessing Artificial Intelligence to Empower HR Processes and Drive Enhanced Efficiency in the Workplace to Boost Productivity. International Journal on Recent and Innovation Trends in Computing and Communication. 2023;11(8s):381–390. https://doi.org/10.17762/ijritcc.v11i8s.7218