iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** III **Month of publication:** March 2024

DOI: https://doi.org/10.22214/ijraset.2024.59396

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue III Mar 2024- Available at www.ijraset.com

An Efficient Deep Neural Network Based Image Enhancement with Improved Filtering Technics

Jariya Begum[1], Siranganayaki. S[2], Steffy. D[3]
[1]*Assitant Professor, Dept of Artificial Intelligence and Data Science, Meenakshi Sundaraarajan Engineering College, Chennai, Tamil Nadu*
[2, 3]*Dept of Artificial Intelligence and Data Science, Meenakshi Sundaraarajan Engineering College, Chennai, Tamil Nadu*

Abstract: ***Paper explores the challenging problem of text-to-image synthesis within the domain of generative modeling. The task involves generating coherent and realistic images from textual descriptions, a critical aspect of automatic learning for applications such as image creation, modification, analysis, and optimization. In response to existing limitations in scene understanding, especially in synthesizing coherent structures in complex scenes, we introduce a novel model called CapGAN. CapGAN utilizes skip-thought vectors to encode given text statements into vector representations, serving as inputs for image synthesis through an adversarial process involving a generator (G) and discriminator (D). The distinguishing feature of CapGAN lies in the integration of capsules at the discriminator level, enriching the model's comprehension of orientational and relative spatial relationships within different entities of an object in an image.***
The proposed model exhibits exceptional proficiency in creating visually coherent structures, marking a significant contribution to the field of generative modeling and image synthesis.

I. INTRODUCTION

In the field of computer-aided design (CAD), automatic art generation (AAG), and various other applications, the problem of generating images or illustrations from a single sentence is a significant challenge. This problem is known as text-to-image synthesis. The goal is to automatically generate images from a given sentence that accurately represents the described scene or object.
One approach to solving this problem is to use machine learning techniques, specifically deep learning models like Generative Adversarial
Networks (GANs) and Variational Autoencoders (VAEs).These models have shown promising results in generating realistic images from textual descriptions.
For example, a GAN model can be trained to generate images from a dataset of sentences and their corresponding images. The model learns to generate images that are similar to the training data. When given a new sentence, the model generates an image that is consistent with the description provided in the sentence.

Another approach is to use a VAE model, which learns to generate images that are similar to the training data. The VAE model learns to generate images by maximizing the likelihood of the training data and minimizing the divergence between the generated images and the training data.
In both cases, the models learn to generate images that are visually similar to the training data. This means that the generated images will have the same general structure and features as the training data.

However, generating images from text can be challenging for complex scenes, where objects are composed of multiple entities with different colors and shapes. In such cases, the models may struggle to generate realistic images that accurately represent the described scene or object.

To address this challenge, we have explore the following strategies:

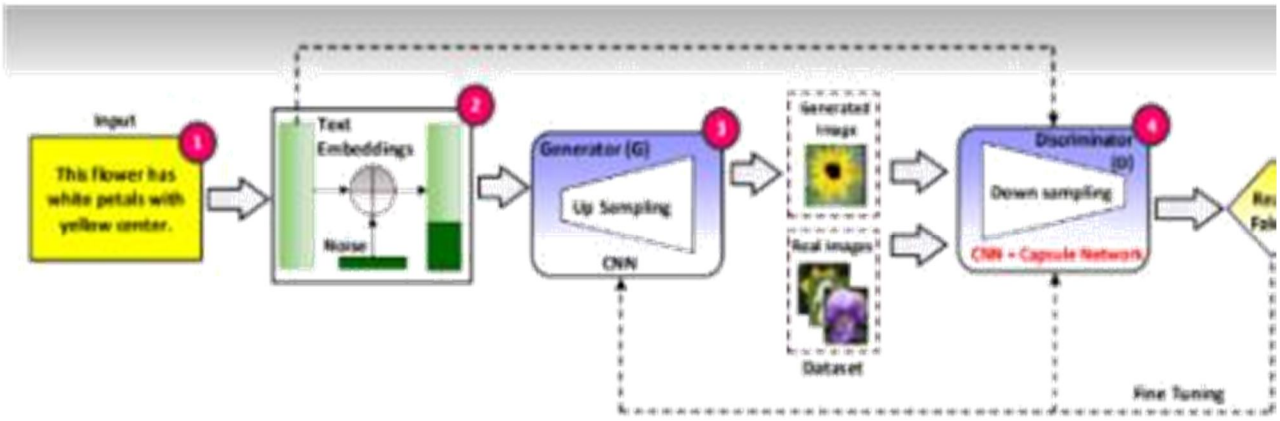
1) *In Corporate Additional Information:* Models can be trained on datasets that include additional information, such as object bounding boxes, part labels, or depth maps. This additional information can help the models generate more accurate images.
2) *Use Pre-trained Modles:* Models can be pre-trained on large datasets, such as ImageNet, and hen fine-tuned on the specific task of generating images from textual descriptions. This approach can help the models learn more general features from the pre-training data and then apply these features to the specific task of generating images from textual descriptions.
3) *Improve the Quality of the Training Data:* Researchers can invest more effort in collecting high-quality training data, which can help the models generate more accurate images.
4) *Explore New Model Architectures:* We can explore new model architectures that are specifically designed for generating images from textual descriptions. These architectures may be more suitable for generating realistic images from complex scenes

In conclusion, the problem of generating images from textual descriptions is a challenging but important problem in the field of computer-aided design, automatic art generation, and various other applications. By leveraging advanced machine learning techniques and incorporating additional information, researchers can develop models that can generate realistic images from textual descriptions. This will enable the automatic learning process and improve the overall quality of generated images.

II. BACKGROUND

Capsule networks have indeed shown promise in various computer vision tasks, including object detection, segmentation, and classification. However, their application to the domain of generating images from textual descriptions remains largely unexplored. This is because the traditional deep learning models, such as GANs, CNNs, and RNNs, have been primarily used for this task. These models have limitations, such as generating low-resolution images, blurred images, and not capturing global coherent structures.

To address these limitations, we have proposed various architectures and techniques. One approach is to use capsule networks with GANs. Capsule networks have the potential to capture hierarchical features and model spatial relationships.



To explore this avenue further,we can consider the following:

1) Investigate the use of capsule networks with GANs for generating images from textual descriptions. This can be done by training a capsule network on a dataset of images and their corresponding textual descriptions. The capsule network can then be used as a generator in a GAN architecture.

2) Experiment with different architectures for the capsule network. These architectures can include variations in the number of layers, the number of capsules in each layer, and the type of capsule layers.

3) Investigate the use of attention mechanisms in conjunction with capsule networks. This can be done by incorporating attention mechanisms into the capsule network architecture. Attention mechanisms can help the network focus on the most relevant words in the text description when generating the image.

4) Evaluate the performance of the proposed capsule network-based GAN architecture on various datasets, such as the MS COCO dataset, the Caltech UCSD birds dataset, and the COMPLEX dataset. These datasets can provide insights into the model's ability to generate realistic images, capture global coherent structures, and handle variations in viewpoints and object relationships.

5) Investigate the use of pre-trained models, such as BERT or GPT, as a source of textual features for the capsule network. This can help the network learn more robust representations of the textual descriptions.

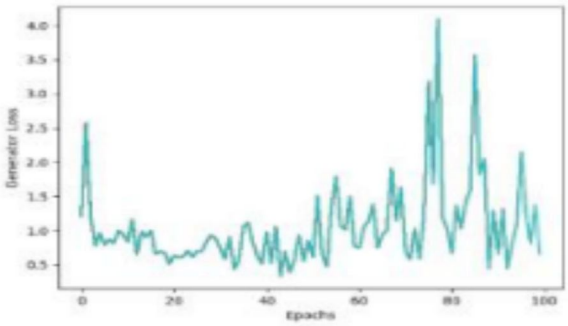
6) Explore the use of multi-modal learning techniques, such as multi-task learning or transfer learning, to improve the performance of the proposed capsule network-based GAN architecture. These techniques can help the model learn more effectively from the textual descriptions and the images. We contribute to the advancement of text-to-image synthesis and the development of more robust and realistic models. These models can potentially be used in various applications, such as virtual reality, digital art, and image search engines.

In conclusion, the use of capsule networks with GANs for generating images from textual descriptions hold promise for revolutionizing the field of text-to-image synthesis. Further research in this area can lead to the development of more robust and realistic models that can capture global coherent structures and handle variations in viewpoints and object relationship.

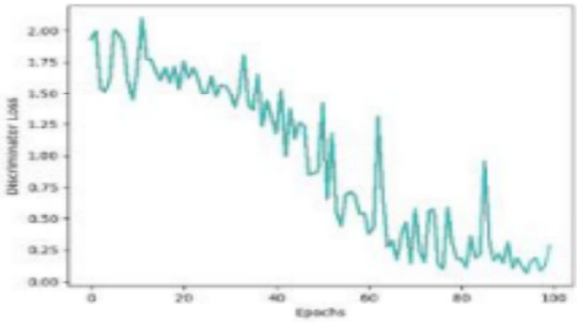
III. METHODOLOGY

Text-to-image synthesis is a challenging problem in the field of computer vision and natural language processing. The goal is to generate an image that corresponds to a given textual description. This involves converting the text input into a meaningful representation, such as a feature vector, and then using this representation to generate an image that matches the description.

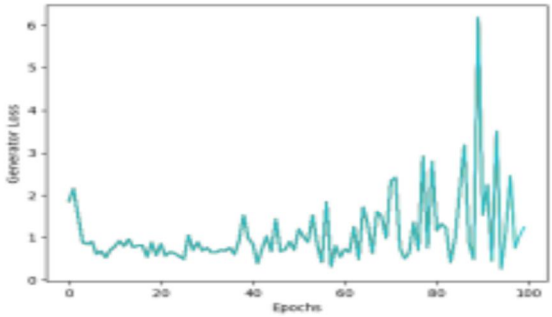
There have been several approaches to text-to-image synthesis, including using convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to generate images from text.



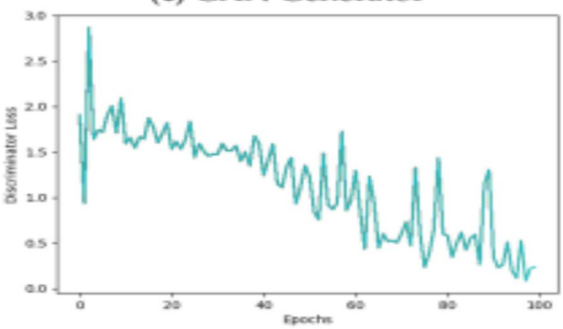
(a) CapGAN Generator



(b) CapGAN Discriminator



(c) GAN Generator

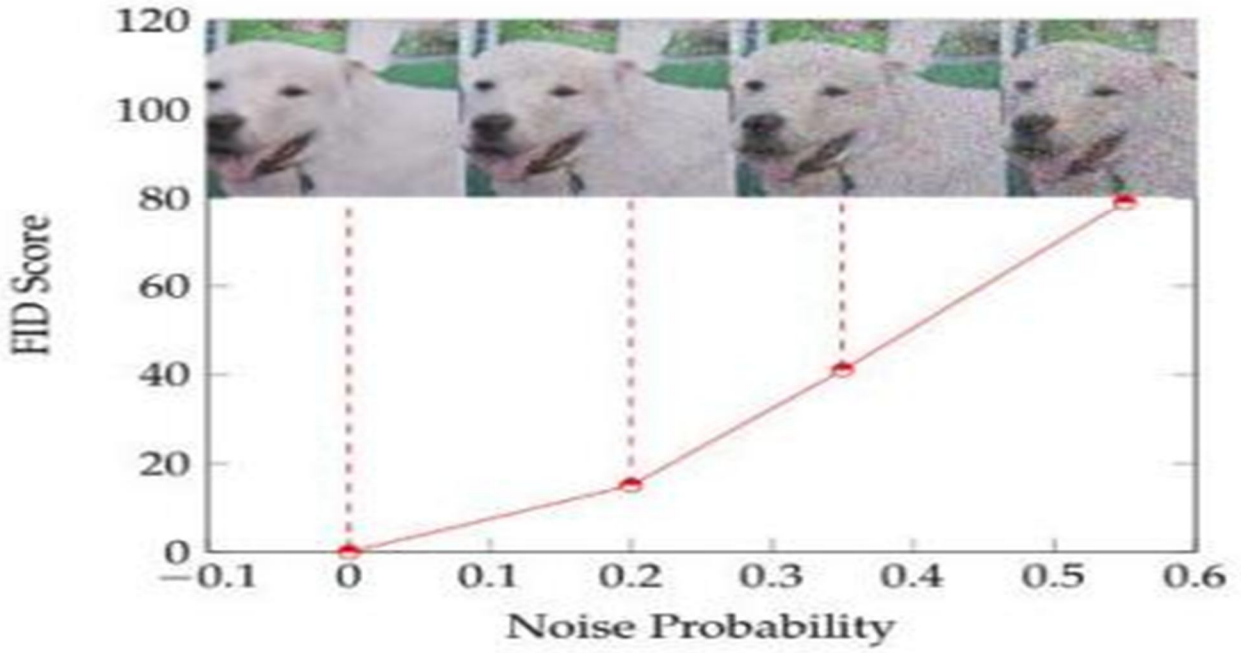


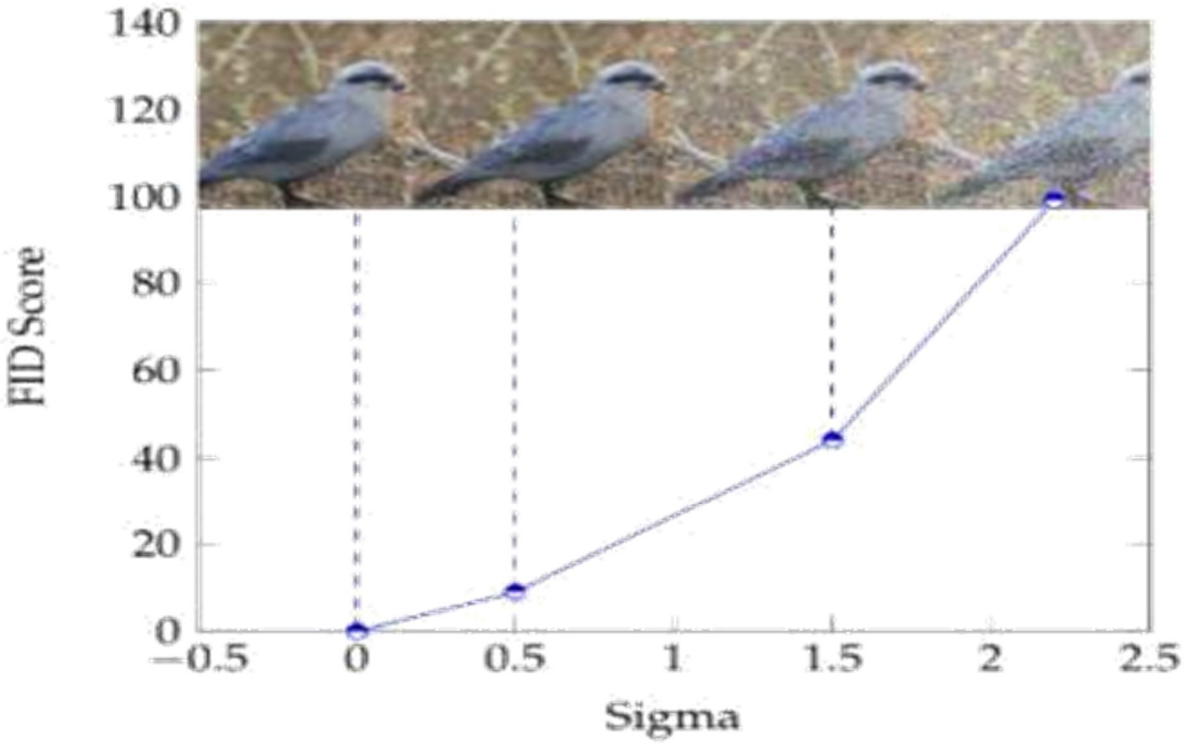
(d) GAN Discriminator

In the above figure (a,b) shows losses for G and D for CapGAN, respectively. For D, the loss decreases as the epochs increase. However, the loss of G starts increasing after the 60th epoch which indicates that D became too strong relative to the G. Beyond this point, G finds it almost impossible to fool D. When D loss decreases to a small value (i.e., 0.1 to 0.2) and G loss increases to a high value (i.e., 2 to 3), it means that model is trained, as G cannot be further improved. (c,d) are losses for G and D of GAN:To calculate the loss, all layers are kept as convolutional layers.In comparison, D loss for CapGAN is less than GAN.

However, these approaches have limitations, such as not being able to capture the hierarchical relationships among the entities of an object.

To address this limitation, we propose a new approach called CapGAN, which uses capsule networks in an adversarial process to better model the hierarchical relationships among the entities of an object. CapGAN consists of four main phases: input sentence, text encoding, image production, and image discrimination.





The input sentence is encoded into a vector representation using skip-thought vectors, which learn fixed-length representations of sentences in any natural language. The encoder network takes the sentence and generates a fixed-length representation using a recurrent neural network (RNN). The previous and next decoder networks take the embedding and try to generate the previous and next sentences, respectively.

Decoders are trained to minimize reconstruction error, which is back propagated to encoders for training. Noise is added before generating the fixed-length representations to make the embedding space more robust.

The generator network of CapGAN takes the caption vector of length 2400 obtained from the text encoding step and compresses it to acquire the text embedding of dimension 256. The text embedding is concatenated with noise, projected, and reshaped into a tensor of dimension $4 \times 4 \times 1024$. This tensor is passed through a series of deconvolutions for up sampling, and a tensor of dimension $64 \times 64 \times 3$ is obtained, which is the generated image from the given text.

The discriminator of CapGAN uses a capsule network along with CNN layers to retain more information by the vectors, thus capturing the relationship among different entities of an object in the input image. The discriminator resolves a binary classification problem of real or fake images using a sigmoid function and gives an output between 0 to 1.

The evaluation of CapGAN on the Caltech-UCSD Birds dataset and show that it outperforms the baseline models in terms of Inception score, Frechet Inception Distance, and Kernel Inception Distance.

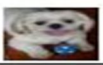
Overall, CapGAN provides a promising approach for automatic text-to-image synthesis by incorporating capsule networks in an adversarial process. The use of capsule networks helps to better model the hierarchical relationships among the entities of an object, resulting in more realistic and accurate image synthesis.

IV. RESULTS

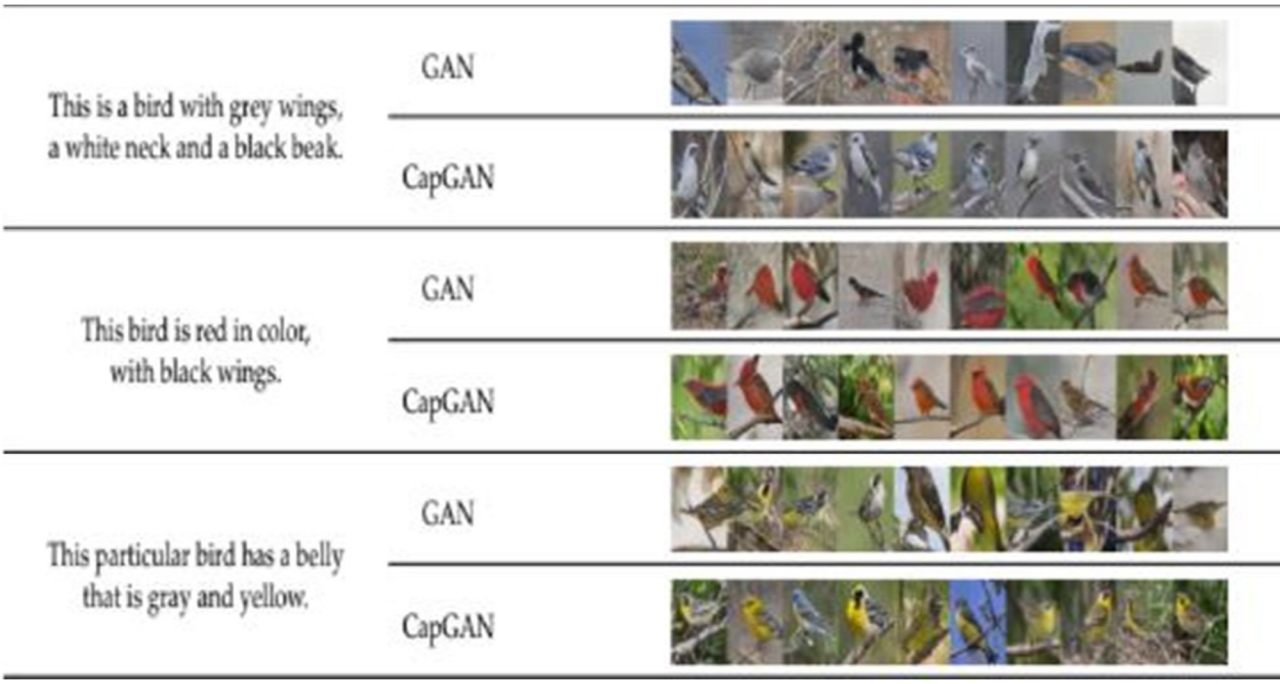
The CapGAN architecture is an approach to text-to-image synthesis that utilizes a capsule network in an adversarial process to generate images that better model the hierarchical relationships among the entities of an object. We conducted comprehensive experimentation using standardized datasets to evaluate the proposed model's performance, including the Oxford-102 dataset for flower images, Caltech-UCSD Birds 200 for bird images, and ImageNet for images of dogs.

The CapGAN model is trained using a fixed set of parameters and evaluated using the inception score (IS) and Fréchet inception distance (FID) metrics. The IS measures how different the score distribution of synthesized images is from the overall class balance, while the FID measures the similarity between generated and real-world images.

We found that CapGAN achieved the highest inception score and lowest Fréchet inception distance compared to other state-of-the-art models for text-to-image synthesis. This indicates that CapGAN's images are more recognized, meaningful, and have a greater diversity of information, while also being less distorted and closer to real-world images.

Text	Ground Truth	Generated Images Using CapGAN
This flower has a white petal with a yellow center.		
This flower has red petals with white center.		
This flower has a yellow petal with orange spots.		
This flower has pink petals with a pink center.		
This bird is yellow and black in color, with a long black beak.		
This particular bird has a belly that is gray and white.		
This is a brown and beige bird and brown on the crown		
White Shih-Tzu		

Additionally, we present visual results of the images generated using CapGAN and show that the model is able to learn the coherent structure details described in the input sentence. The losses for G and D calculated for GAN and CapGAN during training show that CapGAN's discriminator loss decreases to a small value, indicating that the model is trained and the generator cannot be further improved.



We also compare CapGAN to earlier state-of-the-art models for text-to-image synthesis, including GAN, StackGAN, StackGAN++, and TAC-GAN. CapGAN outperforms these models in terms of IS and FID, further highlighting its usefulness in generating high-quality images from text.

Through extensive quantitative evaluation, CapGAN demonstrates its efficiancy with an inception score (IS) of 4.05 ± 0.050, showcasing a remarkable 34% improvement over images synthesized using traditional GANs. Additionally, the Fréchet inception distance (FID) records at 44.38, reflecting an almost 9% enhancement compared to previous state-of-the-art models. These results underscore the effectiveness of CapGAN in generating images with complex scenes and highlight its advancement in addressing the challenges associated with text-to-image synthesis

V. DISCUSSION

The CapGAN architecture is an effective approach to text-to-image synthesis, particularly for complex scenes where objects are composed of multiple entities that are interlinked to form a whole part. The proposed model utilizes a capsule network at the discriminator level to grasp the orientational and relative spatial relationships between different elements of an object, resulting in visually more appealing images compared to conventional layers.

The CapGAN model also preserves the multimodality of the problem, meaning that there can be multiple correct answers for a single input sentence. The model generates numerous possible pixel configurations that can accurately depict the same description, ensuring diversity in the synthesized images.

Moreover, the integration of capsule networks at the discriminator level allows CapGAN to synthesize global coherent structures in complex scenes. The capsule networks extract the geometric information of an object in an image in the form of vectors and use it for inverse rendering, enabling the model to identify the spatial associations among an object's several entities in a scene.

Experimental results show that the CapGAN model outperforms other state-of-the-art models in terms of inception score and Fréchet inception distance, indicating that it generates more recognizable, meaningful, and diverse images that are closer to real-world images. The generated images by CapGAN are evidently closer to the given text and have more relative spatial and orientational association between objects and group of pixels, compared to images generated using conventional networks.

Furthermore, the CapGAN model captures the color(s), basic shape of each entity in the scene, as well as the spatial relationships between objects in complex scenes.

The images generated by CapGAN show a smooth transition between colors and better connections between different parts of the objects, compared to images generated by GAN. In conclusion, the CapGAN architecture is a promising approach to text-to-image synthesis, particularly for complex scenes. The integration of capsule networks at the discriminator level enables the model to grasp the orientational and relative spatial relationships between different elements of an object, resulting in visually more appealing images that are closer to the given text.

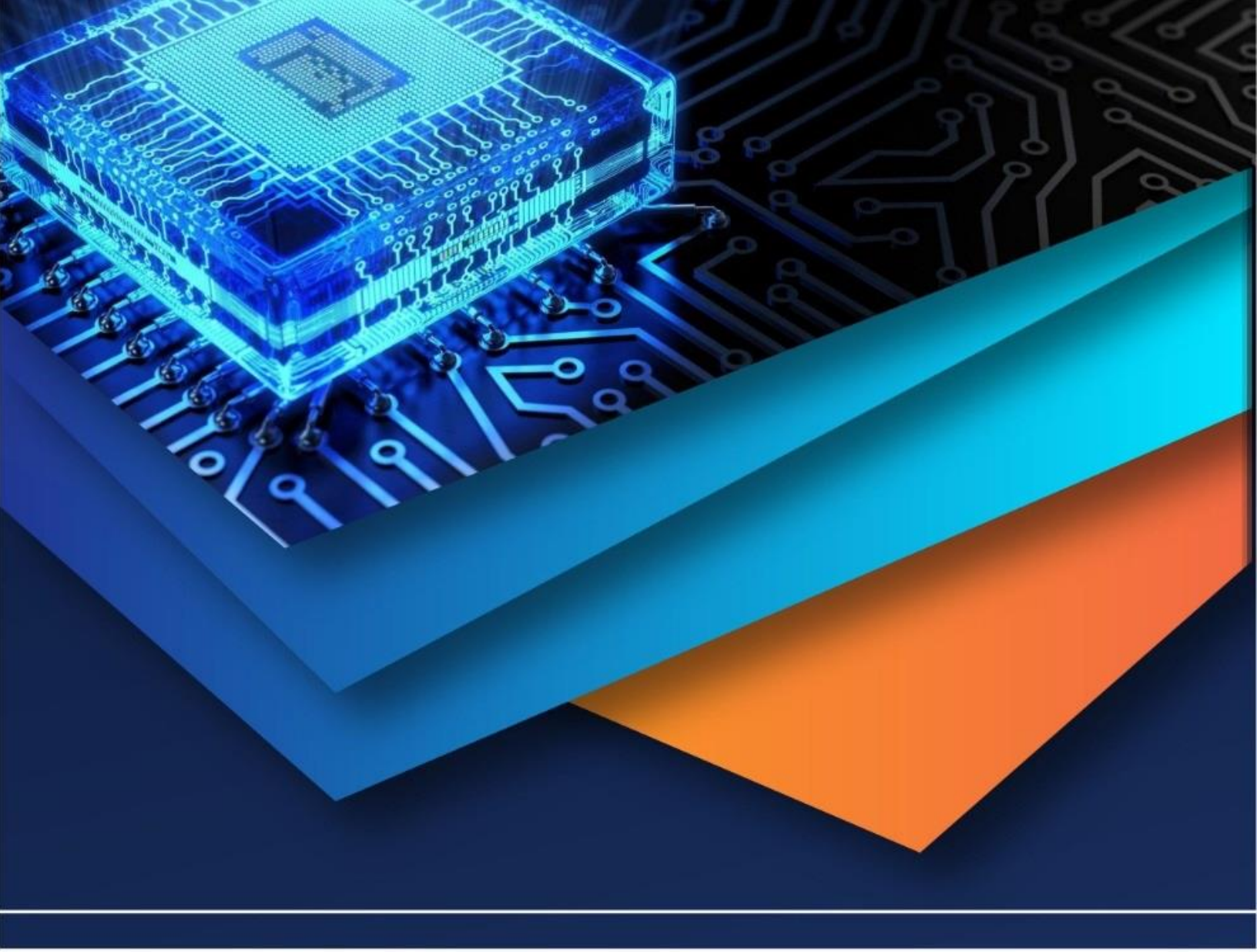
The CapGAN model also preserves the multimodality of the problem and synthesizes global coherent structures in complex scenes. Therefore, it outperforms other state-of-the-art models in terms of inception score and Fréchet inception distance. Further research can be done to improve the performance of the CapGAN model by incorporating attention mechanisms and exploring other architectures for the generator and discriminator networks.

VI. CONCLUSION

In this research, we proposed and evaluated a novel model called CapGAN for generating images from a given text statement. The model utilizes an adversarial process with two models, generator (G) and discriminator (D), trained simultaneously. The discriminator stage in CapGAN replaces convolutional layers with capsule layers, which incorporate orientation and relative spatial interactions between various objects. The experimental findings demonstrate the usefulness of the proposed model, particularly for generating images for complex scenarios. The model outperforms existing state-of-the-art models for the image synthesis problem. In future work, the model can be scaled up to generate higher resolution images by replacing the traditional deconvolutional neural network at the generator level with anti-capsule networks. Additionally, the results can be further improved by using multistage GAN architectures, where the output obtained in one phase is passed alternatively to the next phase. In conclusion, CapGAN is a promising approach for text-to-image synthesis, particularly for complex scenes, by incorporating capsule layers in the discriminator stage. The experimental results show that CapGAN outperforms existing state-of-the-art models for the image synthesis problem. In future work, the model can be further improved and scaled up to generate higher resolution images.

REFERANCES

[1] Xu, T.; Zhang, P.; Huang, Q.; Zhang, H.; Gan, Z.; Huang, X.; He, X. Attngan: Fine-grained text to image generation with attentional generative adversarial networks. arXiv preprint arXiv:1711.10485, 2017.
[2] Afshar, P.; Mohammadi, A.; Plataniotis, K.N. Brain Tumor Type Classification via Capsule Networks. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7– 10 October 2018.
[3] Lukic, V.; Brüggen, M.; Mingo, B.; Croston, J.H.; Kasieczka, G.; Best, P.N. Morphological classification of radio galaxies: Capsule networks versus convolutional neural networks. Mon. Not. R. Astron. Soc. 2019, 487, 1729– 1744.
[4] Hilton, C.; Parameswaran, S.; Dotter, M.; Ward, C.M.; Harguess, J. Classification of maritime vessels using capsule networks. In Geospatial Informatics IX; SPIE: Bellingham, WA, USA, 2019; Volume 10992, pp. 87– 93.
[5] Bass, C.; Dai, T.; Billot, B.; Arulkumaran, K.; Creswell, A.; Clopath, C.; De Paola, V.; Bharath, A.A. Image synthesis with a convolutional capsule generative adversarial network. In Proceedings of the International Conference on Medical Imaging with Deep Learning, London, UK, 8– 10 July 2019.
[6] Jaiswal, A.; AbdAlmageed, W.; Wu, Y.; Natarajan, P. Capsulegan: Generative adversarial capsule network. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8– 14 September 2018.
[7] Upadhyay, Y.; Schrater, P. Generative adversarial network architectures for image synthesis using capsule networks. arXiv preprint arXiv:1806.03796, 2018.
[8] Dash, A.; Gamboa, J.C.B.; Ahmed, S.; Liwicki, M.; Afzal, M.Z. TAC-GAN-text conditioned auxiliary classifier generative adversarial network. arXiv preprint arXiv:1703.06412, 2017.
[9] Zhang, Z.; Xie, Y.; Yang, L. Photographic text-to-image synthesis with a hierarchically-nested adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18– 23 June 2018.
[10] Chen, Q.; Koltun, V. Photographic Image Synthesis with Cascaded Refinement Networks. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22– 29 October 2017.
[11] Sangkloy, P.; Lu, J.; Fang, C.; Yu, F.; Hays, J. Scribbler: Controlling Deep Image Synthesis With Sketch and Color. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21– 26 July 2017.
[12] Nie, D.; Trullo, R.; Lian, J.; Petitjean, C.; Ruan, S.; Wang, Q.; Shen, D. Medical image synthesis with context-aware generative adversarial networks. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Quebec City, QC, Canada, 11– 13 September 2017.
[13] Dong, H.; Yu, S.; Wu, C.; Guo, Y. Semantic image synthesis via adversarial learning. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22– 29 October 2017.
[14] Wang, T.C.; Liu, M.Y.; Zhu, J.Y.; 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), Salt Lake City, UT, USA, 18– 22 June 2018.
[15] Liang, X.; Lee, L.; Dai, W.; Xing, E.P. Dual motion GAN for future-flow embedded video prediction. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22– 29 October 2017
[16] Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.P.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. In Proceedings of the CVPR 2017, Honolulu, HI, USA, 21– 26 July 2017a.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY