IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

6th National Conference on Science, Technology and Communication Skills – NCSTCS 2K23

Narula Institute of Technology, Agarpara, Kolkata, India

Vol. 10, Special Issue 3, September 2023

Image-to-Image Translation Using Cycle GAN

Tirthajyoti Nag[1], Jayasree Ghosh[2], Sonali Bhowmik[3] and Abhishek Bhowmik[4]

Student, Department of Computer Science and Engineering, JIS University, Kolkata, India[1,2]

Faculty, Department of Computer Science and Engineering, JIS University, Kolkata, India[3,4]

Abstract: Image-to-image translation tasks in computer vision and graphics often require to train models to convert input images to required output images. However, obtaining paired training data, where corresponding input-output image pairs are available, can be challenging or even impossible in many cases. We present a unique method that enables the translation of a source input image (X) to a target output image (Y) without requiring paired training data, while preserving the content representations. Our primary objective is to make the model learn a conversion function F: X→Y, that generates images from F(X) that closely resemble the target distribution Y. To accomplish this, we are introducing a loss function known as adversarial loss function which trains F to produce realistic target images. As the mapping from X to Y is inherently ambiguous, we also introduce an inverse mapping G: Y→X and enforce cyclic consistency by adding a loss term that ensures G(F(X)) ≈ X and F(G(Y)) ≈ Y. Our proposed method leverages Cycle-Consistent Generative Adversarial Networks (CycleGAN), which utilize unpaired datasets to automatically learn image pairings. We evaluate our approach on various translation tasks related to images where paired training data is unavailable, demonstrating its effectiveness and versatility.

Keywords: Image-to-image translation, Cycle GAN, Adversarial loss, Computer vision, Image pairings.

I. INTRODUCTION

The task of converting an input image into a desired output image known as Image-to-Image translation, is a fundamental task in computer vision and graphics. Traditional methods for image translation typically require a large dataset of paired images, where each input image is associated with a corresponding output image. However, obtaining such paired data can be costly and labor-intensive, hindering the scalability and applicability of these methods [1].

To address the limitations of paired data, recent advancements have been made in unpaired image-to-image translation techniques. Those approaches aim to teach the model the conversion between different domains without the need for explicit correspondences. One prominent method that has shown remarkable success in this area is Cycle-Consistent Generative Adversarial Networks (CycleGAN) [2].

CycleGAN introduces a novel framework for unpaired image translation by combining the power of generative adversarial networks (GANs) and the concept of cycle consistency. The key idea behind CycleGAN is to simultaneously train two mapping functions: one that translates images from the source input image to the target output image, and another that performs the reverse translation [3]. By enforcing a cycle consistency loss, that ensures the translation of an image from one source to another and then get back results in a close approximation of the source image, CycleGAN effectively learns the underlying mapping between the domains.

The CycleGAN methodology is advantageous due to its ability to leverage unpaired datasets, eliminating the need for costly manual annotation of paired images. This opens up a wide range of possibilities for image translation tasks where paired training data is unavailable or impractical to obtain. Furthermore, CycleGAN's ability to preserve content representations while transforming images provides a powerful tool for tasks such as style transfer, object transfiguration, and semantic manipulation.

In this paper, we present a comprehensive study on the application of CycleGAN for unpaired image-to-image translation. We demonstrate its effectiveness and versatility through extensive experiments on various tasks, including style transfer, domain adaptation, and image synthesis. Our results showcase the remarkable capabilities of CycleGAN in generating high-quality translations between diverse image domains.

The rest of this paper is organized as follows: Section 2 states the aim and objective in the field of image-to-image translation. Section 3 presents the methodology of CycleGAN, including the architecture, training procedure, and loss functions. Section 4 showcases the results obtained on different datasets and tasks. Section 5 tells about the future scope of the project. Finally, Section 6 states the conclusion.

Overall, this paper contributes to the advancement of image-to-image translation techniques, highlighting the strength of CycleGAN as a powerful tool for different computer vision and graphics applications, even if paired training data is not available.

II. AIM AND OBJECTIVE

The primary focus of this research endeavor is to delve into the realm of image-to-image translation, particularly when confronted with the challenge of working with limited analogous training data. The overarching goal is to pioneer an innovative and potent framework that harnesses the capabilities of Cycle-consistent Generative Adversarial Networks (GANs) in order to unravel the complexities of image translation in scenarios where conventional training data might be scarce.

At its core, the proposed framework aims to achieve a remarkable feat: to unravel the inherent intricacies of translating an input image into a desired output image, all the while preserving the underlying content representations. This intricate balance between transformation and content preservation is where the prowess of Cycle-consistent GANs comes into play. By harnessing the adversarial nature of GANs and incorporating the cyclical consistency principle, the framework endeavors to extract meaningful insights from the available data, facilitating the creation of a robust model that can perform image-to-image translations even in situations where closely related training data is lacking.

In essence, this research aspires to contribute significantly to the field of computer vision and image processing by presenting a pioneering approach to tackle the challenges of image translation with limited analogous training data. By amalgamating the power of deep learning, adversarial networks, and cyclical consistency, the envisaged framework strives to unlock new avenues of possibility for generating accurate and contextually relevant output images while retaining the essence of the original content.

III. METHODOLOGY

Our approach aims to automatically pair images from two unpaired datasets and generate images that maintain consistency with the input and output domains, thereby enabling the creation of new, realistic images that were not present in either of the original datasets.

A. Model Architecture

Our CycleGAN model employs a unique architecture consisting of two generators and two discriminators. The primary generator, denoted as G_AtoB, is responsible for translating images from source input A to target output B, while the secondary generator, denoted as G_BtoA, performs the reverse translation from target output B to source input A. The generator network architecture is designed to capture the intricate details and characteristics of the input images, enabling the generation of high-quality translations.

Both generators adopt a similar architecture, inspired by the Pix2Pix Patch GAN Discriminator. The discriminators play a crucial role in distinguishing between real and generated images, providing feedback for the training process. By utilizing separate generators and discriminators, our model facilitates bidirectional translation between the input and output domains, allowing for diverse image transformations [4].

To enhance the information content and spatial details within the generated images, we incorporate Instance Normalization in our model. This normalization technique ensures that each feature map retains comprehensive information about both the spatial data and the transforming information, promoting better preservation of the original content during the translation process.

The discriminator networks have the same architecture as the Pix2Pix Patch GAN Discriminator, which is composed of multiple convolutional layers with a PatchGAN design. This design allows the discriminators to analyse the realism of the images generated by the generator at the patch level, enabling fine-grained discrimination. The discriminators are trained to differentiate the real images from the target output images and the corresponding generated images, facilitating adversarial learning and promoting the generation of visually convincing translations.

By utilizing this architecture, our CycleGAN model is able to effectively capture the technique of conversion from the input domain to output domains, allowing for the generation of realistic and visually coherent images. The combination of the dual generators and discriminators, along with the utilization of Instance Normalization, enhances the overall performance and fidelity of the image-to-image translation process.

B. Adversarial Learning and Cycle Consistency Loss

To train our CycleGAN model, we employ adversarial learning in conjunction with a cycle consistency loss. This combination enables the generation of high-quality translations while ensuring consistency between the input and output domains.

1) Adversarial Learning:

In the adversarial learning process, our generators and discriminators work collaboratively to make the quality and realism of the image generated better. The discriminators have the responsibility for distinguishing between real input images from the target output image and the images that are generated produced by the corresponding generator. Simultaneously, the generators aim to trick the discriminators by generating images that are indifferentiable from the real ones. To achieve this, we employ a minimax game framework, where the generators goal is to minimize the adversarial loss, while the discriminators goal is to maximize it.

IARJSET

ISSN (Online) 2393-8021
ISSN (Print) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

6th National Conference on Science, Technology and Communication Skills – NCSTCS 2K23

Narula Institute of Technology, Agarpara, Kolkata, India

Vol. 10, Special Issue 3, September 2023

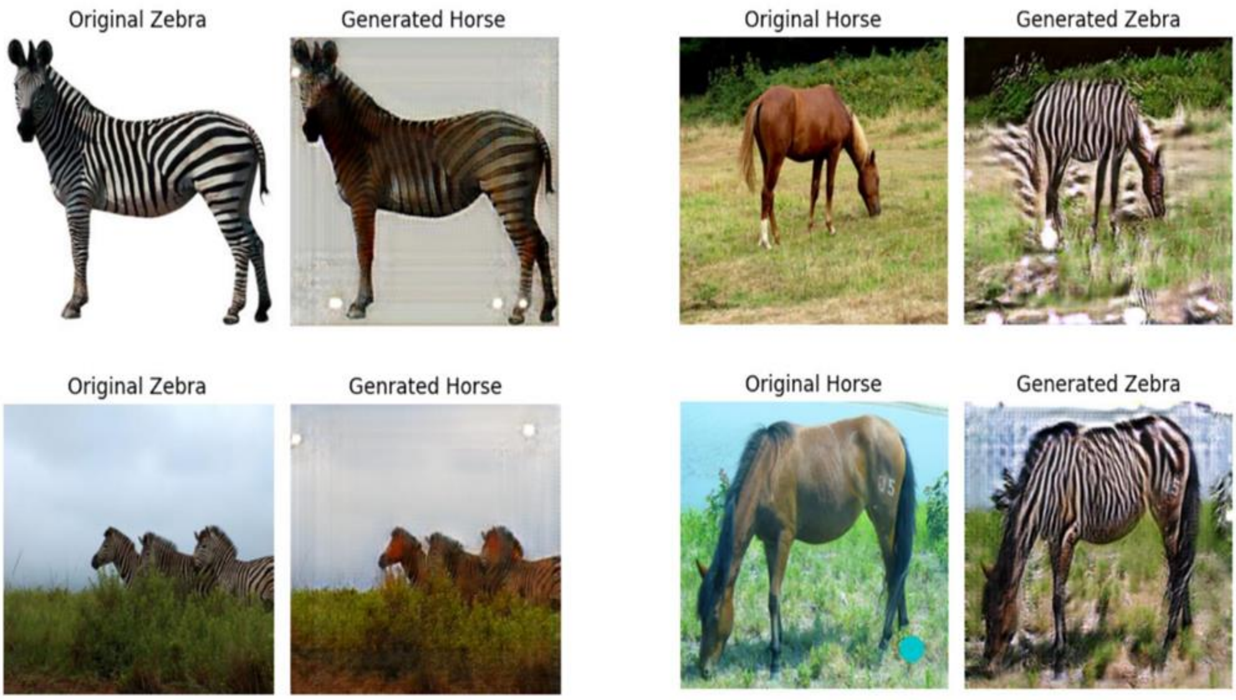
2) Cycle Consistency Loss

To enforce cycle consistency, a cycle consistency loss is introduced that acts as a regularization term during the training process. This loss ensures that the original input images are successfully reconstructed back from the translated images, preserving the content representations and maintaining consistency between the input and output domains. This loss is calculated by converting an image from the input domain A to the output domain B using the generator G_AtoB and then translating it back to domain A using the generator G_BtoA. The differentiating factors between the reconstructed image and the original image serves as a measure of the consistency between the two domains. By minimizing this loss, we encourage the generators to learn the mapping between the domains accurately, ensuring that the translated images retain the essential content from the original input [5].

The combination of adversarial learning and cycle consistency loss provides a powerful training mechanism for our CycleGAN model. The adversarial learning enables the generation of visually convincing translations, while the cycle consistency loss promotes the preservation of content representations and maintains consistency between the input and output domains. Together, these components contribute to the overall effectiveness and realism of the image-to-image translation process.

IV. RESULT

All experiments are done on Kaggle notebook, NVIDIA Tesla T4 GPU and in Python programming language. Few sample experimental results are shown in Fig. 1.



(a) Images of Original Zebra to Generated Horse (b) Images of Original Horse to Generated Zebra

Fig. 1 Sample Results

V. FUTURE SCOPE

The application of CycleGAN for image-to-image translation holds significant promise for further development and improvement. Although the current results showcase the potential of CycleGAN, there are several areas that warrant exploration to enhance its capabilities.

Firstly, expanding the training dataset to include a larger and more diverse collection of images can lead to improved translation quality and the ability to handle more complex transformations. Incorporating additional paired examples, if available, can further enhance the fidelity and accuracy of the generated images. Increasing the training duration beyond the current 25 epochs can allow the model to refine its mapping functions and capture finer details and nuances in the image translations. With more training, the generators can better grasp the intricate characteristics of the target domain, resulting in higher-quality and more realistic outputs. To address the issue of spatial information loss during transformation, innovative techniques can be explored. Preserving spatial details in the generated images, such as through advanced attention mechanisms or spatial consistency losses, can improve overall coherence and visual fidelity. Optimizing the selection and fine-tuning of regularization and loss functions is another avenue for improvement. Balancing adversarial loss, cycle consistency loss, and additional perceptual or content losses can lead to better alignment with the desired output distribution while maintaining content fidelity.

Integrating user interaction and control mechanisms can enhance the practicality and customization options of CycleGAN. Allowing users to guide the style or characteristics of the generated images through additional input or interactive interfaces can make the system more versatile and user-friendly.

By addressing these future directions, CycleGAN has the potential to advance image-to-image translation techniques, surpassing its current capabilities. With the utilization of larger datasets, longer training durations, improved loss functions, and user interaction, the quality, realism, and versatility of the generated images can be significantly enhanced, opening up new possibilities for practical applications.

VI. CONCLUSION

We have used CycleGAN in this paper, a powerful framework for image-to-image translation where the training data need not to be paired. Our technique utilizes a cycle-consistent adversarial learning strategy to train two generators and two discriminators, enabling automatic image pairing and translation between two unpaired datasets.

Through our experiments, we have demonstrated the potential of CycleGAN in generating realistic and high-quality translations. Despite training for a limited number of epochs, the results already showcase the capabilities of the model. By expanding the training duration and dataset size, CycleGAN has the potential to achieve even better performance.

However, there are still challenges to overcome. Spatial information loss during transformation remains a concern, and preserving structural details in the generated images requires further investigation. Additionally, the selection and optimization of regularization and loss functions can be explored to refine the overall performance of CycleGAN.

Looking ahead, there are several avenues for future research and development. Expanding the dataset, increasing training duration, and incorporating advanced techniques for spatial preservation. Additionally, exploring domain adaptation tasks, incorporating user interaction, and refining evaluation metrics can extend the applicability and usability of CycleGAN.

In conclusion, CycleGAN demonstrates significant potential in the field of image-to-image translation. With continued advancements and research, it has the capacity to surpass its current limitations and become a valuable tool for various applications, including style transfer, domain adaptation, and cross-domain synthesis. By addressing the remaining challenges and exploring future avenues, CycleGAN holds the promise of pushing the boundaries of image translation and opening up exciting possibilities for computer vision and graphics.

REFERENCES

[1]. Lin, J., Xia, Y., Qin, T., Chen, Z., & Liu, T. Y. (2018). Conditional image-to-image translation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 5524-5532).
[2]. Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1125-1134).
[3]. Chu, C., Zhmoginov, A., & Sandler, M. (2017). Cyclegan, a master of steganography. arXiv preprint arXiv:1712.02950.
[4]. Liu, M. Y., Breuel, T., & Kautz, J. (2017). Unsupervised image-to-image translation networks. Advances in neural information processing systems, 30.
[5]. Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision (pp. 2223-2232).