WEP Warwick Evans Publishing

Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 5
AIDML 2024

Exploring the Impact of Hyperparameters on the Generation Quality of CycleGAN

Zhaoxiang Tong *

School of Artificial Intelligence, Xidian University, Xi'an, Shaanxi, 710126, China

* Corresponding Author Email: 21009200870@stu.xidian.edu.cn

Abstract. Cycle-Consistent Generative Adversarial Networks (CycleGAN) has demonstrated remarkable proficiency in its ability to undertake image transference between differing domains without the need for paired examples, overcoming a major limitation of traditional image-to-image translation methods. These advantages make it a valuable addition to the toolbox of computer vision researchers and practitioners. However, there persists ample potential for further advancements. This article explores the effectiveness of refining the loss function and training parameters of CycleGAN, aiming to explore how these adjustments affect the outcomes. Specifically, this work plans to improve the loss weight allocation, balancing the translation from different domains. Additionally, the author also explores the number of training epochs and learning rate, halting the process at specific intervals to observe the impact of gradually decreasing the learning rate, as opposed to reducing it to zero. To assess the enhanced performance of CycleGAN, the author will employ both direct human visual perception and cycle consistency loss as evaluation metrics.

Keywords: CycleGAN; deep learning; style transfer.

1. Introduction

In the realm of computer vision and deep learning, there exists a method known as style transfer which enables the preservation of an image's core substance while altering its visual appearance, but its visual style (e.g., colors, strokes, textures, etc.) is changed to the style of another image [1,2]. In essence, style transfer is accomplished through either training a convolutional neural network or utilizing a pre-trained one, which has the ability to disentangle and reassemble the content and style components of images.

Mathematically and algorithmically, style transfer typically involves devising a loss function that comprises multiple elements, including Content Loss, Style Loss, and potentially other supplementary loss terms [3]. Among these, Total Variation Loss, for instance, serves to guarantee that the produced image not only preserves the salient features of the original content image but also emulates the aesthetic qualities of the intended style image [4].

Style transfer is an important branch of deep learning and computer vision research, and its core significance is reflected in the innovation and application of multiple dimensions. Especially in art creation and cross-media content generation, it overturns the traditional artistic expression method, allows users to transform the visual material of the real world into a variety of artistic styles, and greatly broadens the possibility boundary of art creation. It has brought revolutionary tools and solutions to many fields, including entertainment, design, medical image processing, and remote sensing image analysis.

With the rapid development of technology, style transfer technology has achieved a breakthrough from the original single and rough artistic style simulation to today's complex and delicate artistic effect reproduction on the basis of ensuring real-time performance and improving computational efficiency. In this process, researchers continue to deeply explore and optimize the neural network architecture, aiming to retain the content structure and deep semantic information of the original image more accurately, while ensuring the accuracy and artistic quality of artistic style transfer.

In recent years, style transfer technology has been widely used in practical applications, and has been successfully integrated into many frontier fields such as Augmented Reality (AR) or Virtual Reality (VR), video game development, and video post-production. It shows strong scene adaptability and flexible artistic style transfer ability, which further confirms the great practical value of this technology in practice and its broad prospects in future applications [5].

Style transfer methods cover a variety of deep learning-based techniques, this includes but is not limited to neural Style transfer (such as traditional methods based on content and style features extracted by VGG network for matching and Fast Style proposed by Johnson). Cycle-Consistent Generative Adversarial Network (CycleGAN) achieves it without paired data, which utilizes cycle consistency loss to enable free translation of images between two different domains [6]. Moreover, alternative techniques for style transfer exist, including the Adaptive Instance Normalization (AdaIN) method [7] and the Whitening and Coloring Transform (WCT) approach [8], which realize style transfer through innovative feature re-mapping mechanisms. With the evolution of technology, new style transfer technologies used an attention mechanism in conjunction with a generative adversarial network continue to emerge, which continue to promote the development and application expansion of the field of style transfer.

CycleGAN has shown unique advantages in the field of image translation [6]. Its core feature is that it can achieve high-quality image conversion between two different domains without paired training data. Unlike traditional methods that rely on paired images (e.g., an image in domain A versus its counterpart in domain B), the CycleGAN is capable of accomplishing the style transfer task between different domains by only utilizing unpaired datasets of source and target domains. By constructing two generators and the corresponding discriminator, and introducing the cycle consistency loss to ensure that the structural and semantic information in the transformation process is preserved, it achieves excellent performance in the unsupervised learning environment, and greatly reduces the difficulty of data collection and preprocessing when applying such techniques.

This work plans to adjust some parameters of CycleGAN and try to improve the performance of CycleGAN algorithm by observing some metrics about CycleGAN.

2. Method

2.1. Dataset

The Horse2Zebra dataset in CycleGAN is a typical example dataset used to demonstrate the capabilities of the CycleGAN model [9]. This dataset contains images from two different domains: one is a collection of images of horses and the other is a collection of images of zebras. There are 1067 images of horses and 1334 images of zebras in the training set.

During the training phase of CycleGAN, the neural network autonomously learns to identify and extract the feature representations from the input images.The CycleGAN model usually contains several convolutional layers, each of which performs feature extraction on the input image, and the model gradually learns a representation that can effectively distinguish the horse and zebra features.

2.2. Model

CycleGAN is a deep learning-based unsupervised image-to-image translation model, which is unique in that it can complete image style transfer from two different domains without paired training data. The core principle of the model was to use the Generative Adversarial Network (GAN) framework and the constraint of cycle consistency to ensure high-quality and structurally fidelity bidirectional transformation among the target domain and the source domain [10]. Its principal strengths involve the capacity to manage unpaired datasets, execute style transfer while conserving essential image characteristics, and exhibit robust generalization capabilities, making it extensively utilized across various domains like artistic style transfer, image completion, and cross-domain data augmentation.

This work will make a series of innovative changes to the loss function and training parameters of CycleGAN and observe the effects of these changes.

This study adjusts the loss weight parameters bidirectionally, encompassing the refinement from domain A to domain B and the reverse process from domain B back to domain A. In CycleGAN, the loss weights of these two directions directly affect the balance and accuracy of the model's mapping between the two domains. Therefore, by carefully adjusting these weights, it is expected to further improve the model's ability to achieve more accurate cross-domain conversion while retaining the original content of the source image.

The influence of different training epochs and learning rates on the model performance is also explored. Increasing the number of epochs can help improve model performance by giving the model more opportunities to fit the training data, but too many epochs can lead to overfitting, where the model is too fit to the training set and does not generalize well to unseen data. In CycleGAN, due to the complex learning objective such as cycle consistency loss, an appropriate number of epochs can help find a better image transformation space. Learning rate decay is a common regularization technique used to control learning effort late in training, prevent overfitting, and allow the model to fine-tune the weights later in the process. Is an important determinant of the weight update step size. A large initial learning rate may accelerate convergence, but may also cause the model to fall into poor local minima or training instability. A smaller initial learning rate may require more epochs to converge, but may result in better generalization and a more stable learning process. In this experiment, by appropriately increasing or decreasing the number of training epochs and adopting a dynamic learning rate strategy, it facilitates the model's search for an optimal local minimum, thus enhancing the generated images' quality and authenticity.

This work also tries to stop in the process of reducing the learning rate, using latest_net_G_A and latest_net_G_B to compare the results of the test dataset and the corresponding loss function to find the difference between them

3. Result

3.1. Training Details and Evaluation

In this experiment, the author implemented CycleGAN using PyTorch deep learning framework, and the hardware environment used in the experiment is equipped with NVIDIA GeForce RTX 4060 Ti GPU-16GB.

For the evaluation of the experimental results, a two-dimensional evaluation system is used:

The first is the subjective evaluation method: this is a direct evaluation method based on human visual perception. Through manual observation and comparison, the resemblance between the synthetic target image produced by CycleGAN and the actual source image is examined. Specific evaluation criteria include but are not limited to color fidelity, that is, whether the hue distribution in the synthesized image approximates that of the original image; The fidelity of texture details focuses on whether the model can accurately replicate the tiny features in the original image. And shape preservation ability, to investigate whether the main structure and overall form are reasonably maintained in the process of style transfer to avoid deformation or distortion.

The second is Cycle Consistency Loss: as the core evaluation metric of CycleGAN, this loss function reflects the internal self-constraint mechanism of the model in training and transformation tasks. The fundamental idea involves mapping the source domain image through a model to the target domain, subsequently reversing the transformation to remap the converted target domain image back into the source domain. After completing this "cyclical" process, the disparity between the retrieved image and the original source image is assessed. If the content within the initial input image can be effectively reproduced following two successive transformations, that is, the cycle consistency loss value is small, it indicates that the model effectively captures and retains the key information of the original image without paired supervision, and achieves high-quality image transformation effects without relying on pairwise training data.

3.2. Quantitative Results and Visualization

This work successfully implemented the above parameters tuning with the Horse2Zebra training set. Primarily, this study demonstrates the effects brought about by varying cycle-consistency loss weights. The parameters lambda_A and lambda_B individually regulate the respective cycle consistency loss intensities when transforming from domain A to domain B and then back to domain A (and its reciprocal direction). These two parameters determine how much weight the model attaches to the constraint of cycle consistency during training. In this work, the control variable method is used, set two sets of data, the first set of lambda_A=5.0, lambda_B=5.0, the second set of lambda_A=10.0, lambda_B=10.0, and the rest of the parameters are unchanged, while training 200 epochs, the results are as follows:

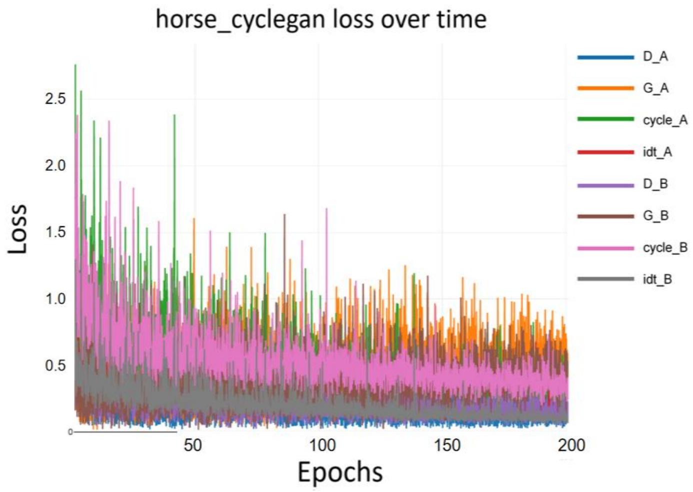

Fig. 1 Loss curvesw with lambda A=5 and lambda B=5 (Figure Credits: Original).

Fig. 2 Result visualization with lambda A=5 and lambda B=5 (Figure Credits: Original).

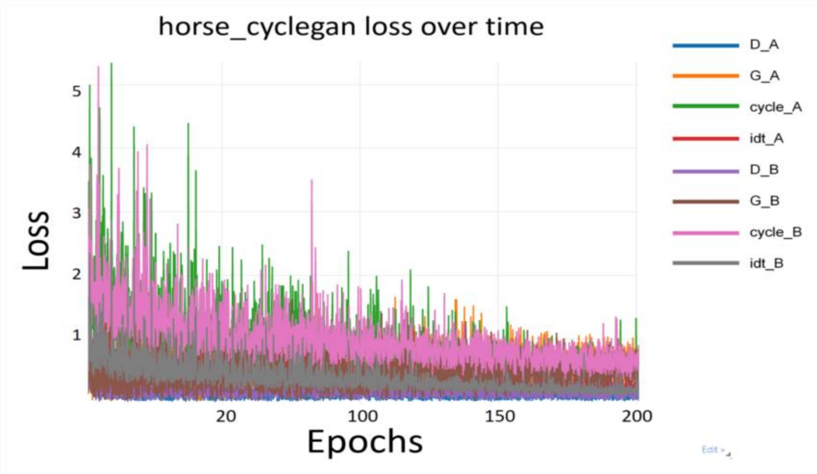

Fig. 3 Loss curves with lambda A=10 and lambda B=10 (Figure Credits: Original).

Fig. 4 Result visualization with lambda A=10 and lambda B=10 (Figure Credits: Original).

Fig. 1 and Fig. 2 are obtained after adjusting the parameters as lambda_A=5 lambda_B=5, and Fig. 3 and Fig. 4 are obtained after adjusting the parameters as lambda_A=10 lambda_B=10. They all got their final results after 200 epochs of training

Moreover, this study employs the controlled variable approach by altering specific training parameters such as the overall number of epochs with an initial CycleGAN learning rate, the epochs at which the learning rate linearly decreases to zero, and the starting learning rate for the Adam optimizer, all aimed at enhancing the quality and authenticity of the produced images. Below are the corresponding experimental outcomes.

Fig 5 shows the total number of epochs trained at initial learning rate is 100, and the number of epochs where the learning rate linearly decays to zero is 100. Fig. 5, from top to bottom, respectively shows the results with learning rate of 0.001, 0.0005, and 0.0002. Fig. 6 shows the results with a learning rate of 0.0005, 50 initial learning rate training epochs and 50 learning rates linearly decays to zero.

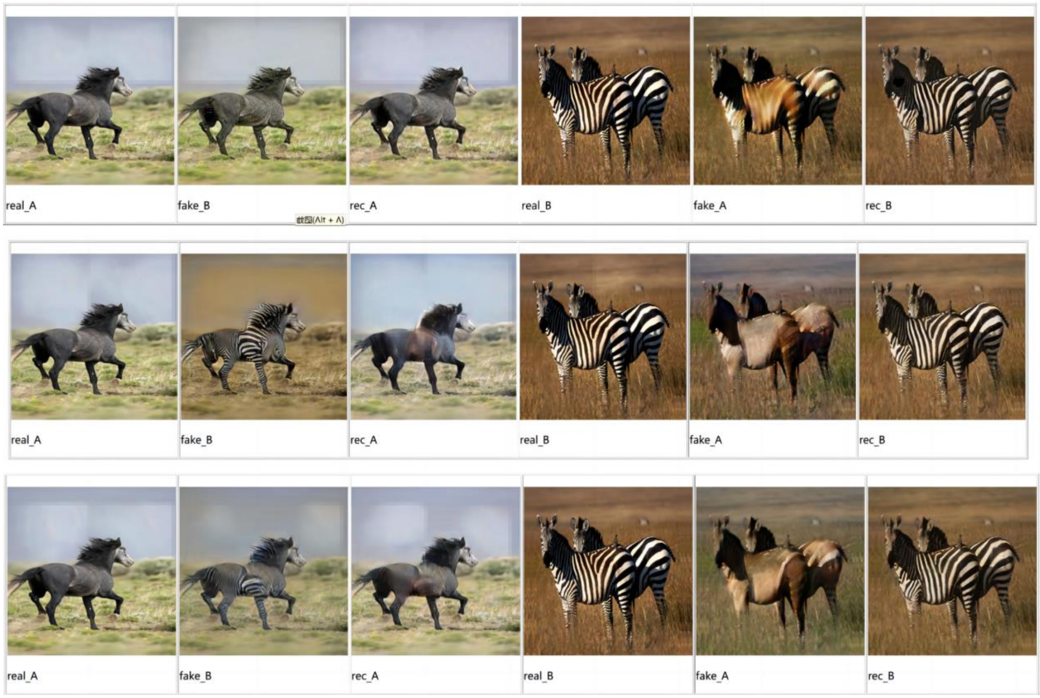

Fig. 5 Visualization of results from different learning rates. From top to bottom are respectively 0.001, 0.0005 and 0.0002 (Figure Credits: Original).

Fig. 6 Visualization of results using learning rate 0.0005, where first 50 epochs with initial learning rate and latter 50 epochs with learning rate linearly decays to zero. (Figure Credits: Original).

When the learning rate is reduced to 0.000400, the training is stopped, and the results and corresponding loss curves of latest_net_G_A and latest_net_G_B tested on this dataset are obtained and compared with the learning rate reduced to 0, corresponding results are shown in Fig. 7, Fig. 8, and Fig. 9.

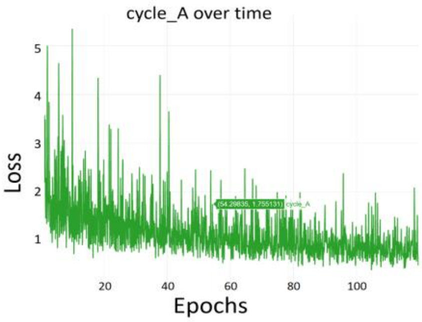

Fig. 7 Loss curve with learning rate 0.0004 (Figure Credits: Original).

Fig. 8 Visualization results of testing with learning rate 0.0004 (Figure Credits: Original).

Fig. 9 Visualization results of testing when learning rate drops to 0 (Figure Credits: Original).

Fig. 7 is the cyclic consistency loss with respect to A->B->A when the learning rate drops to 0.0004, Fig. 8 is the test result, and Fig.9 is the test result when the learning rate eventually drops to 0.

4. Discussion

In the first set of experiments, the direct evaluation of human visual perception based on lambda_A=10 lambda_B=10 was better than that based on lambda_A=5 lambda_B=5. When looking at the cyclic consistency loss, the cycle consistency loss of the model with lambda_A=10 lambda_B=10 at the beginning is about twice that of the model with lambda_A=5 lambda_B=5

It shows that the model with lambda_A=10 and lambda_B=10 can better retain the information of the original A-domain image during the process of performing the cyclic mapping, that is, the image converted from the B domain back to the A domain is more similar to the original A-domain image.

In the second set of experiments, it could be observed that when the learning rate is 0.001, the gradient update is too aggressive, causing the model to miss effective solutions, and the results obtained are not satisfactory. When the learning rate is 0.0005, the obtained model can convert between horses and zebras well, and when the learning rate is 0.0002, the obtained model is the best.

In the third set of experiments, it could be observed that the model obtained with a learning rate decayed to 0.000400 is inferior to the one obtained with a decayed learning rate of 0. However, upon examining the loss function curve, it becomes apparent that this model still retains some capacity for learning throughout its training cycle.

5. Conclusion

Form the aforementioned results, it could be concluded that by increasing the weight, the model will pay more attention to the source image during training. Conversely, this constraint is weakened, and the model may focus more on direct style transfer effects at the expense of a certain degree of cyclic fidelity.

A higher initial learning rate can speed up convergence. However, it may also lead to suboptimal local minima or training instability. In contrast, a lower initial learning rate may require more convergence cycles, but provides enhanced generalization capability and a more stable learning process.

Alternatively, increasing the number of epochs can provide more opportunities for the model to fit the training data and thus improve its performance.

The model does not achieve complete convergence when the learning rate drop is non-zero. However, it can still effectively fine-tune the parameters in subsequent stages, thereby facilitating convergence towards a more refined solution and enabling adaptation to dataset changes or prevention of premature convergence to local optima.

CycleGAN has had a profound impact in the field of image processing and computer vision, and its potential application value and industrial prospects have expanded over time. Although CycleGAN has a certain history as a relatively mature technology, there is still a huge potential and space to be tapped. Through continuous exploration and improvement, CycleGAN can not only keep pace with The Times and adapt to the rapidly developing technological environment and social needs, but also is expected to give rise to a series of new industries and solutions with high added value.

References

[1] Jing, Yongcheng, Yezhou Yang, Zunlei Feng, Jingwen Ye, Yizhou Yu, and Mingli Song. Neural style transfer: A review. IEEE transactions on visualization and computer graphics, 2019, 26(11): 3365-3385.

[2] Singh, Akhil, Vaibhav Jaiswal, Gaurav Joshi, Adith Sanjeeve, Shilpa Gite, and Ketan Kotecha. Neural style transfer: A critical review. IEEE Access, 2021, 9: 131583-131613.

[3] Kotovenko, Dmytro, Artsiom Sanakoyeu, Sabine Lang, and Bjorn Ommer. Content and style disentanglement for artistic style transfer. In Proceedings of the IEEE/CVF international conference on computer vision, 2019: 4422-4431.

[4] Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge. A neural algorithm of artistic style. ArXiv Preprint, 2015: 1508.06576.

[5] Sandamini, Asangika, Chamodi Jayathilaka, Thisara Pannala, Kasun Karunanayaka, Prabhash Kumarasinghe, and Dushani Perera. An Augmented Reality-based Fashion Design Interface with Artistic Contents Generated Using Deep Generative Models. In 2022 22nd International Conference on Advances in ICT for Emerging Regions, 2022: 104-109.

[6] Zhu, Jun-Yan, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision, 2017: 2223-2232.

[7] Huang, Xun, and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In Proceedings of the IEEE international conference on computer vision, 2017: 1501-1510. 2017.

[8] Chiu, Tai-Yin. Understanding generalized whitening and coloring transform for universal style transfer. In Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 4452-4460.

[9] Horse2zebra Dataset, 2017, URL: https://www.kaggle.com/datasets/balraj98/horse2zebra-dataset. Last Accessed:2024/03/12

[10] Creswell, Antonia, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A. Bharath. Generative adversarial networks: An overview. IEEE signal processing magazine, 2018 35(1): 53-65.