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

Impact of Hyperparameters on the Quality of Image Translation Using CycleGAN

Yichen Hu *

School of Computing and Data Science, Xiamen University Malaysia, Selangor, Malaysia

* Corresponding Author Email: CST2109155@xmu.edu.my.com

Abstract. The study delves into the intricate relationship between hyperparameters and image translation quality in Cycle-Consistent Generative Adversarial Network (CycleGAN), which is one of the most widespread tools for unsupervised image-to-image translation. The hyperparameters investigated include learning rate, identity loss, and cycle consistency loss, which are aimed at enhancing the model's quality of producing realistic and good images. Furthermore, by employing horse2zebra dataset author was able to execute various experiments in order to understand what influence hyperparameters in CycleGAN have to its performance in style transfer problems. The analysis shows that changing the learning rate and identity loss weight have a drastic effect on whether the translation done yields detailed or correct images, tallied with the low-learning rate producing more explicit and accurate image translation. Loss weight transform manipulation of the identity is a pivotal element in the restoration of the original and the transformed features from the images. The study provides more insights into the understanding of how CycleGAN behaves in different training environments, and it stresses the significance of adequate parameters fine-tuning for the optimal image translation.

Keywords: CycleGAN; image translation; hyperparameter.

1. Introduction

The digital world has emerged in the last decade and there are several new technologies have brought changes in several processes including art and design. In this regard, Style transfer is one of the methods that has been expanding in recent years. Deep learning and computer vision are working on it. Style transfer combines two or more images to create a new hybrid with the better of both [1]. The technology is utilized in art and design as well as in everyday life to improve visual media and feed machine-learning models. Convolutional Neural Networks were crucial to this field's advancement. Convolutional Neural Networks (CNNs) enable more sophisticated modifications. Cycle-Consistent Generative Adversarial Network (CycleGAN) can learn and generate complicated style changes without paired examples, making it an excellent tool for understanding pictures in this dynamic, shifting environment [2,3]. This unsupervised method lets users exchange new trends without matching picture pairs.

CycleGAN's approach and style transfer needed improvement. The original goal was to create images that resembled famous works. Though difficult, it kept me going. Over the years, professionals have studied many styles and ways. A few rudimentary systems can just change colors and textures. Others use neural networks to interpret and adjust complicated visual features. Generational Adversarial Networks (GANs) and their descendants have enabled larger and more complicated advancements [4,5]. CycleGAN allows two-way unsupervised image translation that retains many source images while achieving the desired style.

This unique study examines how hyperparameters, CycleGAN model efficiency, and output quality interact complexly. The novel design and network theory are well-documented, but no research has examined how different training sets affect the model's capacity to produce realistic, high-quality images. Several factors such as learning rate, identity loss, adversarial loss, and cycle consistency loss can affect the final output. This study reveals how vital it is to adjust CycleGAN style transfer characteristics, which are generally overlooked. By studying this domain, academics and consumers can improve CycleGAN in many situations. It also fills literature gaps.

This study examines how changing the learning rate, identity loss, and cycle consistency loss weight affects CycleGAN style transfer projects. As per expectation, this study will illuminate how the model operates in diverse training scenarios and improve style transfer approaches. This paper advances the field by emphasizing hyperparameter selection and the delicate balance between model design and training approach. This creates more versatile and powerful artistic and practical image-altering tools. Researchers expanded the knowledge of CycleGAN and established the stage for future research to increase neural network performance in various fields.

2. Method

2.1. Dataset

In this study, the used dataset is horse2zebra. This dataset has been widely used for evaluate GANs models such as cycleGAN. The images of both domains are divided into a training subset and a testing subset. The training dataset contains 1067 horse images and 1334 zebra images, and the test datasets have 120 horse images and 140 zebra images downloaded from ImageNet [6]. All the images are resize to 256 x 256 pixels dimensions.

2.2. CycleGAN

For a long time, image-to-image translation required paired training data, but CycleGAN eliminates this requirement. Studies introduced GANs, which have two producers and two discriminators [4]. Generators translate images between domains. The discriminators compare the target domain's real images against the generators' to determine if the translated images are real. This competition forces generators to generate increasingly realistic graphics, expanding the styles that can be conveyed using unsupervised learning.

Cycle consistency allows CycleGAN to rebuild the original image after translation to the target domain and return. This cycle consistency loss is crucial for preserving visual information while translating. The target area style and relevant data can be saved together. CycleGAN is unique since it ensures consistency without paired situations. This makes it ideal for switching artistic styles, improving photographs, and scientific imaging without matched image pairs.

CycleGAN has many benefits. Allowing realistic style transfer without observation opens up new creative and beneficial options. It can manage unrelated datasets and is adaptable, making it useful in various fields [7]. Despite its strengths, CycleGAN has issues. When translating anything difficult, such as preserving the meaning but employing a different style, faults or artifacts might lower picture quality. Training CycleGAN models using high-resolution images or big datasets is computationally intensive and time-consuming.

Generator-discriminator links are important to CycleGAN architecture. Image translation without a partner was cleverly solved by this architecture. Two of these pieces make up CycleGAN. Different sets translate images differently. The study requires this dual structure for the model to transfer styles between areas without matched image pairs.

2.2.1. Generators

CycleGAN generators are represented by G_{AB} and G_{BA}. They truly translate images. G_{AB} does domain A to B, while G_{BA} does the inverse [8]. The target domain's picture distribution is copied by these deep neural networks. They then alter the assigned photographs to appear randomly selected from that distribution. Image style is recorded and shown by slowly encoding, transforming, and decoding with convolutional layers.

2.2.2. Discriminators

If the model is adversarial, discriminators D_A and D_B judge with generators. D_A evaluates pictures using domain A criteria, while D_B employs domain B criteria. Discriminators employ convolutional neural networks to distinguish authentic photographs from their areas from fakes. User feedback on how realistic generator visuals are helps improve translation accuracy and interest.

2.2.3. Adversarial Loss

The adversarial loss function for the entire process is expressed as follows:

$$L_{GAN}(G_{AB}, D_B, A, B) = E_{b\sim p_{data}(b)}[logD_B(b)] + E_{a\sim p_{data}(a)}[\log(1 - D_B(G_{AB}(a)))] \quad (1)$$

$$L_{GAN}(G_{BA}, D_A, B, A) = E_{a\sim p_{data}(a)}[logD_A(a)] + E_{b\sim p_{data}(b)}[\log(1 - D_A(G_{BA}(b)))] \quad (2)$$

The concept involves the distributions of the target domain B denoted as $b\sim p_{data}(b)$ and the source domain a denoted as $a\sim p_{data}(a)$. $D_B(b)$ represents the score assigned by the discriminator D_B to sample b, while $D_B(G_{AB}(a))$ represents the score given by discriminator D_B to the fake image GAB(a) produced by the generator GAB. The interaction between the generator G_{BA} and the discriminator D_A follows a similar principle.

2.2.4. Cycle Consistency Loss

This concept is special to CycleGAN. It is easy to switch images between domains (A to B and back), and much of the original material is maintained. The cycle consistency technique prevents translation distortion of source images [9]. Forcing generators to learn reversible mappings that match the intended style does this. The cycle consistency loss, denoted as L_{cyc}, is defined as:

$$L_{cyc}(G_{AB}, G_{BA}) = \lambda_A \cdot E_{a\sim p_{data}(a)}[\| G_{BA}(G_{AB}(a)) - a \|_1] \quad (3)$$

$$L_{cyc}(G_{BA}, G_{AB}) = \lambda_B \cdot E_{b\sim p_{data}(b)}[\| G_{AB}(G_{BA}(b)) - b \|_1] \quad (4)$$

Where λ_A and λ_B are the given weight to the cycle consistency loss for the forward and backward, respectively.

2.2.5. Identity Loss

The concept of identity loss is employed to mitigate the effects of extensive migration and to prevent an excessive blending of colors between the input and output.

$$\lambda_{id} \cdot E_{b\sim p_{data}(b)}[\| G_{AB}(b) - b \|_1] \quad (5)$$

These components allow the CycleGAN architecture to create new images with a new style but the same content, achieving the optimal balance between originality and realism. The balance of CycleGAN makes it good at single image-to-image translation, and useful in commerce, science, the arts, and more.

3. Experiments are Results

3.1. Experimental Setup

This study used the same base model as in [10]. The discriminator architecture was based on PatchGAN, employing local patches of dimensions 140×140.6 The discriminator comprises six downsampling convolutional layers with normalization and Leakey ReLU operations, culminating in a 30x30 result matrix. To enhance the robustness of adversarial training, a buffer is employed to retain 50 previously generated images.

For the implementation, experiments were conducted on a robust hardware and software setup. The author utilized a computer equipped with an 11th Gen Intel(R) Core (TM) i9-11900K CPU and an NVIDIA Geforce GTX 3090 GPU, running on a Windows 10 operating system. The development environment was set up with Python 3.8, PyTorch 1.8, supported by CUDA 11.0 and CuDNN 11.0. Additional libraries like Tensorboard, Numpy, Scipy, Opencv-python, Seaborn, and Matplotlib were used for various computational and visualization purposes.

Regarding the training settings, some of the parameters suggested in [7] was modified. All models were trained for 150 epochs with a batch size of 4. Experiments are conducted with 6 different hyperparameter settings on the cycleGAN model (Table 1).

Table 1. Experimental hyperparameter configuration

	λ_A	λ_B	λ_{id}	Learning rate
1	10	15	0.5	0.0002
2	10	10	0.5	0.001
3	10	5	2	0.0002
4	10	5	2	0.001
5	10	20	1	0.0003
6	10	20	3	0.0003

3.2. Comparative Results

In terms of image quality, the initial group of parameters within the six sets analyzed demonstrates superior results compared to the others. The visualization results are shown in Fig. 1. Both the real_A-fake_B1-rec_A1 and real_B-fake_A1-rec_B1 image sets exhibit exceptional clarity, capturing all intricate details and closely resembling the input colors. Conversely, the second group exhibits the lowest image quality, with the fake_B2 transformed image displaying less distinct zebra features and the rec_B2 reconstructed image lacking specific details, especially in the zebra's eye region. The difference between the first and second groups can be attributed to the varying values of λ_B and the learning rate. In Experiment 1, λ_B is set at a high level while the learning rate is low. The image quality of the fourth group is also suboptimal, as it shares the same learning rate as the second group, which is the highest among the experimental sets. It is plausible that the learning rate has a more significant impact on image generation than λ_B, as higher learning rates tend to produce rougher image quality.

The visual outcomes of experiments three and four exhibit notable distinctions, with the learning rate being the variable under examination. Experiment 4 demonstrates a significantly higher learning rate compared to experiment three. The findings indicate that the transformed and reconstructed images in experiment three are markedly subpar, characterized by blurriness, indistinct outlines, evident distortion, and increased feature loss. Conversely, the group in Experiment 3, with a lower learning rate, notably captures more features, boasts high-definition quality, and presents clear background features. This observation suggests that a reduced learning rate leads to slower but more accurate learning, which facilitates improved outcomes with a stronger focus on acquiring details. Conversely, a heightened learning rate accelerates the learning process and convergence but results in a rough learning trajectory, leading to blurry images and feature loss. Consequently, for individuals seeking training outcomes with intricate details, it is recommended to start with a lower learning rate and adjust it in response to training advancements.

In experiments 5 and 6, notable variations were observed in the outcomes of the output image where only the hyperparameter In experiments 5 and 6, notable variations were observed in the outcomes of the output image where only the hyperparameter λ_{id} is different. The experiments outcomes also emphasize the critical role of λ_{id} that assists in balancing the aspects which will be kept intact attributing to the original domain and the ones that will be transformed in integration through the CycleGAN framework. Experiment 5 showed an enhanced intensity of the reconstructed image(rec_A5) shooting the horse facial features, reducing blurriness and elimination of "zebra appearance" residue. Given these observations with Experiment 5 that a low setting for the λ_{id} parameter makes the model emphasize adaptation in the new domain perhaps at the expense of preserving the identity of the original, Experiment 6 showed the opposite direction - a translated image with a higher λid value (fake_B6) more stayed faithful to the original attributes when a backward re-mapping was done and exemplified additionally a more complete zebra spots appearance. However, The results of Experiment 6 displayed a coloring that was somewhat different from the original image. This could occur as an identity loss oblivious those changes mean, changing for instance the brightness and contrast that could happen during the domain translation. Moreover, the model with higher λ_{id} degree mainly focuses on the keeping of the initial image features. Thereby, the model is potentially restricted to adaptation on new domains. In conclusion, to adapt to various goals and considerations, precise tuning of the λ_{id} degree is essential. Setting the afterburners at a value greater than the lambda coefficient is feasible when you are planning to achieve more flexibility in a new region and also keep up with imaging that would introduce novel features. On the contrary, opting for a higher percentage of Luma decay is preferable in situations where the author aim to preserve the key attributes of the original images. After thorough tuning of different hyperparameters, such as λ_{id}, one can achieve the optimal model configuration that best suits the specific task requirements. The optimized configuration, in turn, helps improve performance and enhance image translation tasks.

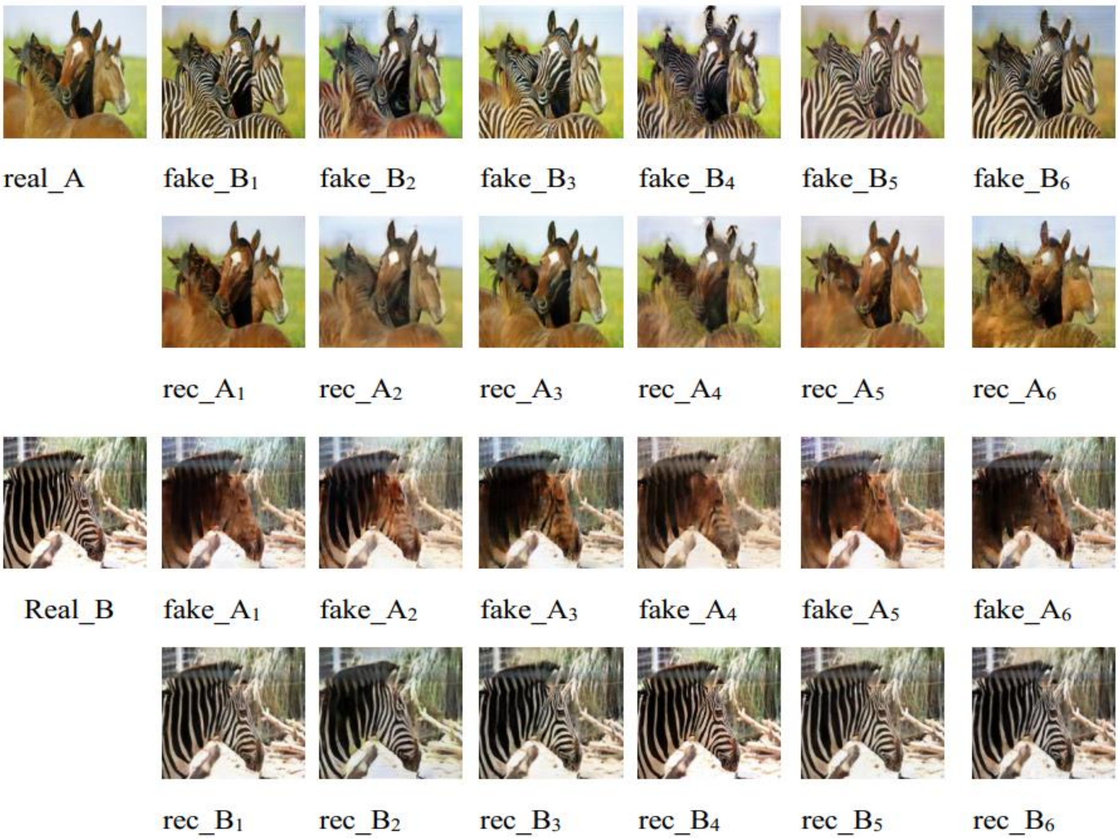

Fig 1. Result of CycleGAN translations under different hyperparameter settings (Figure Credits: Original)

4. Conclusion

The study demonstrates the impact of adjusting the learning rate and identity loss weight on enhancing the effectiveness of unsupervised image-to-image translation. The tweaking of these hyperparameters which means a lot in the process of translation result in a whole lot of difference. By means of such constant adjustments of the hyperparameters and visual tests the study found some glaring differences between one photo and another. The author found in particular that the amount of identity loss weighting in cycleGAN has a significant impact on the image features and the final quality. Moreover, the investigation found that this selection leads to faster, either better or worse image quality convergence, which depends on the learning rate and the accuracy of the model in result generation. Several empirical evidences show that progressively reduced learning rate normally improves the image quality, although it can prolong the learning process. It is needed to have the ability to capture the image details. Tuning of hyperparameters is a fundamental pillar of the image generation process as it is found that the optimal values differ from the task to the task as well as from the dataset to the dataset. In general, the experiment was quite limited in scope in terms of the hyperparameters it considered. The others were overlooked while the best hyperparameter configuration is being emphasized, which may result in a not a truly optimal setting. Besides that, the researches on the other hyperparameters which might have an effect on the generation of images are also necessary and the empirical analysis can also be adopted as a method to analyze the influences in a comprehensive way.

References

[1] Fu, Xuhui. Digital Image Art Style Transfer Algorithm Based on CycleGAN. Computational Intelligence and Neuroscience 2022: 1-10.

[2] Sharma, Neha, Reecha Sharma, and Neeru Jindal. Comparative analysis of CycleGAN and AttentionGAN on face aging application. Sādhanā, 2022, 47(1): 33.

[3] Li, Zewen, Fan Liu, Wenjie Yang, Shouheng Peng, and Jun Zhou. A survey of convolutional neural networks: analysis, applications, and prospects. IEEE transactions on neural networks and learning systems, 2021, 33(12): 6999-7019.

[4] Aggarwal, Alankrita, Mamta Mittal, and Gopi Battineni. Generative adversarial network: An overview of theory and applications. International Journal of Information Management Data Insights, 2021, 1(1): 100004.

[5] 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.

[6] Horse2zebra Dataset. URL: https://www.kaggle.com/datasets/balraj98/horse2zebra-dataset. Last Accessed: 2024/03/21.

[7] Luleci, Furkan, F. Necati Catbas, and Onur Avci. CycleGAN for undamaged-to-damaged domain translation for structural health monitoring and damage detection. Mechanical Systems and Signal Processing, 2023, 197: 110370.

[8] Hindarto, Djarot, and Endah Tri Esti Handayani. Revolution in Image Data Collection: CycleGAN as a Dataset Generator. Sinkron: jurnal dan penelitian teknik informatika, 2024, 9(1): 444-454.

[9] Park, Yae Won, Seo Jeong Shin, Jihwan Eom, Heirim Lee, Seng Chan You, Sung Soo Ahn, Soo Mee Lim, Rae Woong Park, and Seung-Koo Lee. Cycle-consistent adversarial networks improves generalizability of radiomics model in grading meningiomas on external validation. Scientific reports, 2022, 12(1): 7042.

[10] 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.