Comparison of Transfer Style Using a CycleGAN Model with Data Augmentation

Gerardo Lugo-Torres, José E. Valdez-Rodríguez,
Diego A. Peralta-Rodríguez, Hiram Calvo

Instituto Politécnico Nacional,
Centro de Investigación en Computación,
Mexico

{glugot2022, jvaldezr2018, hcalvo}@cic.ipn.mx
dperaltar1100@alumno.ipn.mx

Abstract. Image-to-image translation (I2I) is a specialized technique aimed at converting images from one domain to another while retaining their intrinsic content. This process involves learning the relationship between an input and its corresponding output image through a dataset of aligned pairs. Our study utilizes the CycleGAN model to pioneer a method for transforming images from the domain of Monet's paintings to a domain of varied photographs without the need for paired training examples. We address challenges such as mode collapse and overfitting, which can affect the integrity and quality of the translated images. Our investigation focuses on enhancing the CycleGAN model's performance and stability through data augmentation strategies, such as flipping, mirroring, and contrast enhancement. We propose that judicious dataset selection for training can yield superior outcomes with less data compared to indiscriminate large-volume training. By online scraping Monet's artwork and curating a diverse, representative image subset, we fine-tuned our model. This targeted approach propelled our results to 2nd place in the Kaggle challenge "I am something of a Painter Myself" as of August 3rd, 2023, demonstrating the efficacy of our enhanced training protocol.

Keywords. Generative adversarial network, image-to-image translation, data augmentation, cycle consistency.

1 Introduction

The image-to-image translation transforms an input image from one visual domain to another while preserving its semantic content. In other words, it involves changing the appearance or style of an image while retaining its underlying structure or content. The image-to-image translation aims to learn a mapping function that can convert images from a source domain to a target domain.

The source and target domains can represent different visual characteristics, such as style, color, texture, or even the presence or absence of particular objects.

Traditional methods, such as patch-based algorithms or filter-based approaches, have been effective but are often tailored for specific tasks and lack the ability to generalize across different domains.

Image-to-image translation has been used in a broad range of real-world applications across multiple industries and disciplines. Some relevant examples:

- **Modality Translation:** Consist of translating between different imaging modalities (e.g., from MRI to CT scans) can be useful for medical diagnostics when only one type of imaging is available [23, 20].

- **Simulated Training:** Translating synthetic or simulated data for improved machine learning training [13, 15].

- **Satellite to Map Translation:** Converting satellite images into more interpretable map views can aid in various types of terrain analysis and planning [9, 8].

Table 1. Dataset description

Dataset	Size	Description
Monet300	300	Dataset from Kaggle competition consisting of landscapes.
Monet900	900	Dataset hand picked from Monet1969 with a variety of landscapes, portraits and wide variety of colors.
Monet1172	1172	Dataset generated by flipping, rotating and contrast enhancement of Monet 300.
Monet1337	1337	Dataset generated by randomly selecting i images from Monet 1969.
Monet1969	1969	Dataset obtained by web scrapping[1]. Includes an almost complete gallery of Claude Monet 2500 known artworks.
Photos	7038	Dataset from Kaggle competition consisting of wide variety of photos used to generate the transfer style of our trained model from the Monet images.

– **Design Visualization:** Converting 2D blueprints into 3D images for better visualization and understanding of architectural designs.

These examples only scratch the surface; the possibilities are continually expanding as the technology matures. Image-to-image translation models open up even more possibilities for innovation and application.

In order to complete this task, various techniques can be employed, including generative adversarial networks (GANs) [10], variational autoencoders (VAEs) [12], conditional GANs (cGANs) [22], and other deep learning architectures.

These models have shown remarkable advancements in various computer vision and image processing applications, including image synthesis, segmentation [7], and style transfer [26]. Additionally, these models are trained on large datasets containing pairs of input and target images, learning the mapping between the domains through optimization.

However, a fundamental challenge in image-to-image translation is the unavailability of paired training data in many real world applications.

For instance, in medical imaging, obtaining perfectly aligned images from different modalities (e.g., MRI to CT scans [14]) is often impractical or even impossible. This lack of paired data has been a bottleneck for the effective deployment of GANs in diverse applications.

Cycle-Consistent Adversarial Networks (CycleGANs) offer a groundbreaking solution to this problem by learning to translate images from one domain to another in the absence of paired training examples.

The core innovation lies in introducing a cycle-consistency loss, which ensures that an image translated from one domain to another can be reverted back to the original image, thus preserving the inherent structure and content. Despite its transformative potential, the CycleGAN model is not without its limitations.

There are issues related to mode collapse [6], overfitting to specific styles [24], and the computational intensity of the training process. In this context, data augmentation techniques have been identified as a potential avenue to enrich the training process, thereby enhancing the robustness and performance of the model.

Yet, the interplay between CycleGAN architecture and various data augmentation strategies remains an underexplored area of research [5, 16, 21]. In this work, we investigate using different data augmentation approaches in the performance of a CycleGAN model for Monetesque style transfer in images.

We aim to give some insight about the impact of data augmentation strategies and the CycleGAN performance by conducting an investigation into the use of data augmentation techniques in conjunction with CycleGAN for image-to-image translation tasks.
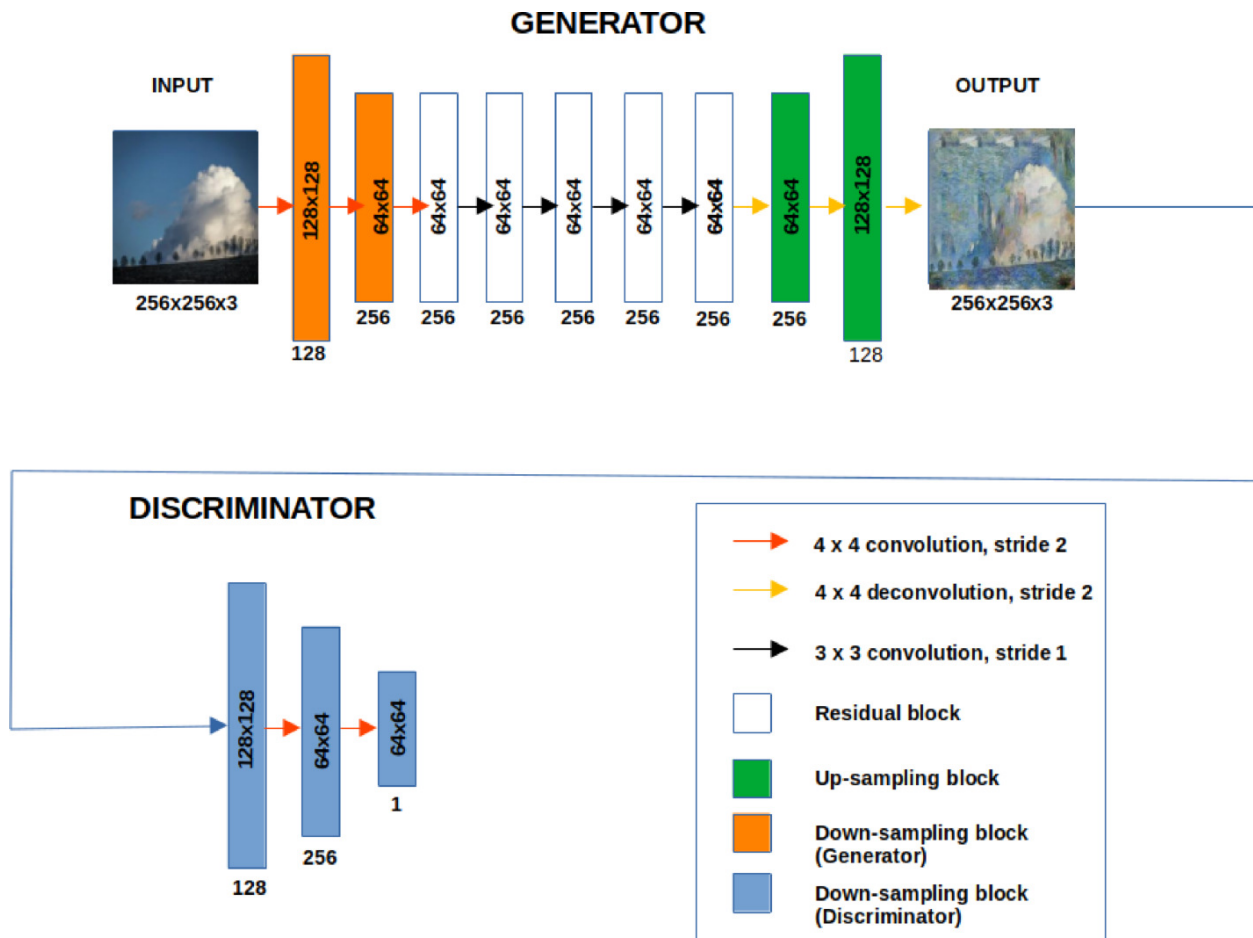
---

[1]www.claudemonetgallery.org/

Fig. 1. Architecture of Cycle GAN

While several prior works have focused on the CycleGAN model in style transfer [16, 2], there are very few [4] focusing on the importance of the relationship between the quality of the data input and the performance of the model metrics.

We focus in the MiFID[1] score since it was the one used in the Kaggle competition for the assessment of the quality of the model.

This work is divided as follows: Section 2 describes the state of the art and related works. Section 3 describes the proposed methodology, Section 4 describes the experiments, Section 5 and 6 summarize our discussion and draw our conclusions.

---

[1]Memorization-informed Frechet Inception Distance

2 Related Work

Most of the work in literature deals with the transfer style of the different artist using supervised methods. The most important work using CycleGAN was done by Zhu et al. [26] being a pioneering work.

They proposed the CycleGAN model and exemplified its advantages over supervised GAN models through an application of image-to-image translation all in the absence of one-to-one mapping between source and target pairs.

In another line of research, Yi et al. (2017) [25] proposed DualGAN, a model similar to CycleGAN, with a different approach to

Fig. 2. Sample of dataset : A) Dataset mostly landscapes, B) Dataset handpicked with more varied paintings

the implementation of the generator and discriminator networks. This work did not explicitly made use of data augmentation but it established important foundations for unpaired image-to-image translation.

More recently, Lee et al. (2021) [11] show the performance of data augmentation in the Pix2Pix model, but with paired training data. They explored several data augmentation techniques, including flipping, rotation, and scaling.

Their study revealed significant improvements in model generalization capabilities, prompting us to explore similar strategies in the context of CycleGAN.

While data augmentation has been extensively studied in deep learning, its implementation in the context of GANs and specifically CycleGAN is relatively under-explored. Ratner et al. (2017) [18] used data augmentation in GANs to generate more diverse images but did not explore unpaired image-to-image translation tasks.

Their approach, however, does offer interesting insights into how data augmentation can improve the quality and diversity of generated images.

Also, Almaihairi et al. (2018) [1] proposed the Augmented CycleGAN model to deal with some significant limitations such as the problem of mode collapse presented in the CycleGAN model.

They do this by extending the original framework to support multi-domain image translation for translation between more than two domains, enabling the generation of images across a broader range of visual styles or attributes.

Our work aims to explore the impact of data augmentation in the performance of the CycleGAN for unpaired image-to-image translation tasks.

We build upon previous studies and our own knowledge by incorporating data augmentation techniques into the CycleGAN model and made use of MiFID metric to evaluate the model performance.

3 Methodology

For this section we intend to present the main features of the CycleGAN model through the description of the architecture used in this work and the metric applied for the assessment of the model performance.

3.1 Generative Adversarial Networks

A generative adversarial network (GAN) is a type of deep learning model consisting of two neural networks: a generator network and a discriminator network. The GAN framework was introduced by

Table 2. Summary of experiment results

Dataset	Epochs	MiFID
Monet300	100	48.2689
	150	43.9139
	175	44.9882
Monet900	25	38.7693
	50	34.8291
	75	40.4339
Monet1172	50	47.4199
	75	42.1799
	100	47.7829
Monet1337	75	39.0477
	125	38.7693
	150	40.1230
Monet1969	25	47.8782
	50	45.6861
	75	46.05

Goodfellow et al. in 2014 [10]. The main objective of a Generative Adversarial Network (GAN) is to produce synthetic data that exhibits a high degree of realism, particularly in the context of generating images, which closely resemble samples from a specified target dataset.

The generator network accepts either random noise or a latent input as its input and endeavors to produce samples that closely resemble the distribution of the target data.

Conversely, the discriminator network undergoes training to differentiate between authentic samples provided from the target dataset and synthetic samples produced by the generator.

The training process of a Generative Adversarial Network (GAN) entails a competitive interplay between the generator and discriminator neural networks. The primary objective of the generator is to generate samples that exhibit a higher degree of realism in order to deceive the discriminator.

Conversely, the discriminator's primary goal is to accurately distinguish between real and counterfeit samples. The learning process is driven by the antagonistic connection between the two networks.

Given a discriminator D and a generator G, this two networks play two-player minimax game with value function $V(G, D)$:

$$\min_G \max_D V(G, D) = \mathbb{E}_{\boldsymbol{x} \sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] \\ + \mathbb{E}_{\boldsymbol{z} \sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))], \quad (1)$$

where p_{data} represents the distribution of generator over the input data x and $p_{\boldsymbol{z}}$ a prior input noise variables. GANs have inspired numerous variations and extensions, such as conditional GANs (cGANs), Wasserstein GANs (WGANs), progressive GANs, which further improve the stability and quality of generated samples [17], and CycleGANs.

3.2 CycleGAN

A CycleGAN (Cycle-Consistent Adversarial Network) is a type of GAN specifically designed for unsupervised I2I translation. It was introduced as a new way to learn mappings between two different image domains without needing paired training data.

The key idea behind CycleGAN is to leverage the concept of cycle consistency. In image translation tasks, the goal is to learn a mapping between images from a source domain and images from a target domain without explicitly paired examples.

The CycleGAN architecture is different from other GANs in a way that it contains 2 mapping functions (G and F) that act as generators and their corresponding Discriminators (D_x and D_y): The generator mapping functions are as follows:

$$G : X \to Y, F : Y \to X, \quad (2)$$

where X is the input image distribution and Y is the desired output distribution. And the cost function used is the sum of adversarial loss and cyclic consistent loss:

$$L(G, F, D_x, D_y) = L_{\text{advers}}(G, D_y, X, Y) + \\ L_{\text{advers}}(F, D_x, Y, X) + \lambda L_{\text{cycl}}(G, F, X, Y). \quad (3)$$

With an objective function with the form of:

$$\min_{G,F} \max_{D_x, D_y} L(G, F, D_x, D_y). \quad (4)$$

The training process of a CycleGAN involves two main components:

1. **Adversarial Loss:** The training of the generators and discriminators is conducted through the utilization of adversarial learning. The primary objective of the generators is to produce images that deceive the discriminators by causing them to classify the generated images as authentic.

   Conversely, the discriminators strive to accurately differentiate between real images and those that have been generated. The utilization of an adversarial loss function contributes to the enhancement of both the quality and realism of the generated images:

$$\text{Loss}_{\text{adv}}(G, D_y, X, Y) = \frac{1}{m}\Sigma(1 - D_x(G(x_i)))^2,$$
$$\text{Loss}_{\text{adv}}(F, D_x, Y, X) = \frac{1}{m}\Sigma(1 - D_y(F(y_i)))^2. \quad (5)$$

2. **Cycle-Consistency Loss:** This principle asserts that an image that undergoes translation from one domain to another and subsequently back should have a high degree of similarity to the original input image.

   The loss function employed in this process guarantees the maintenance of consistency in the mapping between the images in both directions, so aiding in the preservation of the original image content:

$$\text{Loss}_{\text{cyc}}(G, F, X, Y) = \frac{1}{m}[(F(G(x_i)) - x_i) + (G(F(y_i)) - y_i)]. \quad (6)$$

The incorporation of the cycle-consistency loss incentivizes the generators to acquire knowledge about cycle-consistent mappings, hence facilitating the generators' ability to comprehend and represent the common information shared across the two domains.

This limitation serves to mitigate the potential distortion or loss of significant content that may occur during the process of translation. A fundamental weakness of the CycleGAN model is that it learns deterministic mappings.

In CycleGAN and other similar models [17, 2], the conditionals between domains correspond to delta functions: $\hat{p}(a \mid b) = \delta(G_{BA}(b))$ and $\hat{p}(b \mid a) = \delta(G_{AB}(a))$, and cycle consistency forces the learned mappings to be inverses of each other.

Table 3. Leaderboard of Kaggle event[1]

| Place	Team	MiFID
1	HUST AIA PRCD	34.48525
2	**Gerardo Lug**	**34.82910**
3	CLIPTraVeLGAN	35.01656
4	chenccckkk	35.07934
5	MLCV	35.31007
6	GudrunGertold	36.96598
7	Nandita Bhattacharya	37.06163
8	Coffee L	37.29003
9	Alena Shevtsova	37.48513
10	Datendullis	37.68987
11	Issam Ben Moussa	37.71797
12	Andrey Nesterov	38.26549
13	rabbie	38.64153
14	Eishkaran Singh	39.08037
15	Yuanfei Xu	39.08037

When confronted with intricate inter-domain connections, CycleGAN tends to acquire an artificial one-to-one correspondence instead of accurately reflecting the genuine, organized conditional distribution.

The presence of deterministic mappings poses a challenge in achieving optimized cycle consistency, particularly when the domains exhibit significant differences in complexity. In such scenarios, the mapping from one domain to another typically results in a one-to-many relationship.

3.3 Architecture of the Proposed CycleGAN

The CycleGAN generator is composed of three distinct components, namely the Encoder, Transformer, and Decoder. The UNET architecture will be employed for the generator. In order to construct the generator, we establish our downsample and upsample techniques.

The process of downsampling involves reducing the two-dimensional dimensions, specifically the width and height of an image, by a factor known as the stride. The stride refers to the measurement of the distance covered by each step taken by the filter.
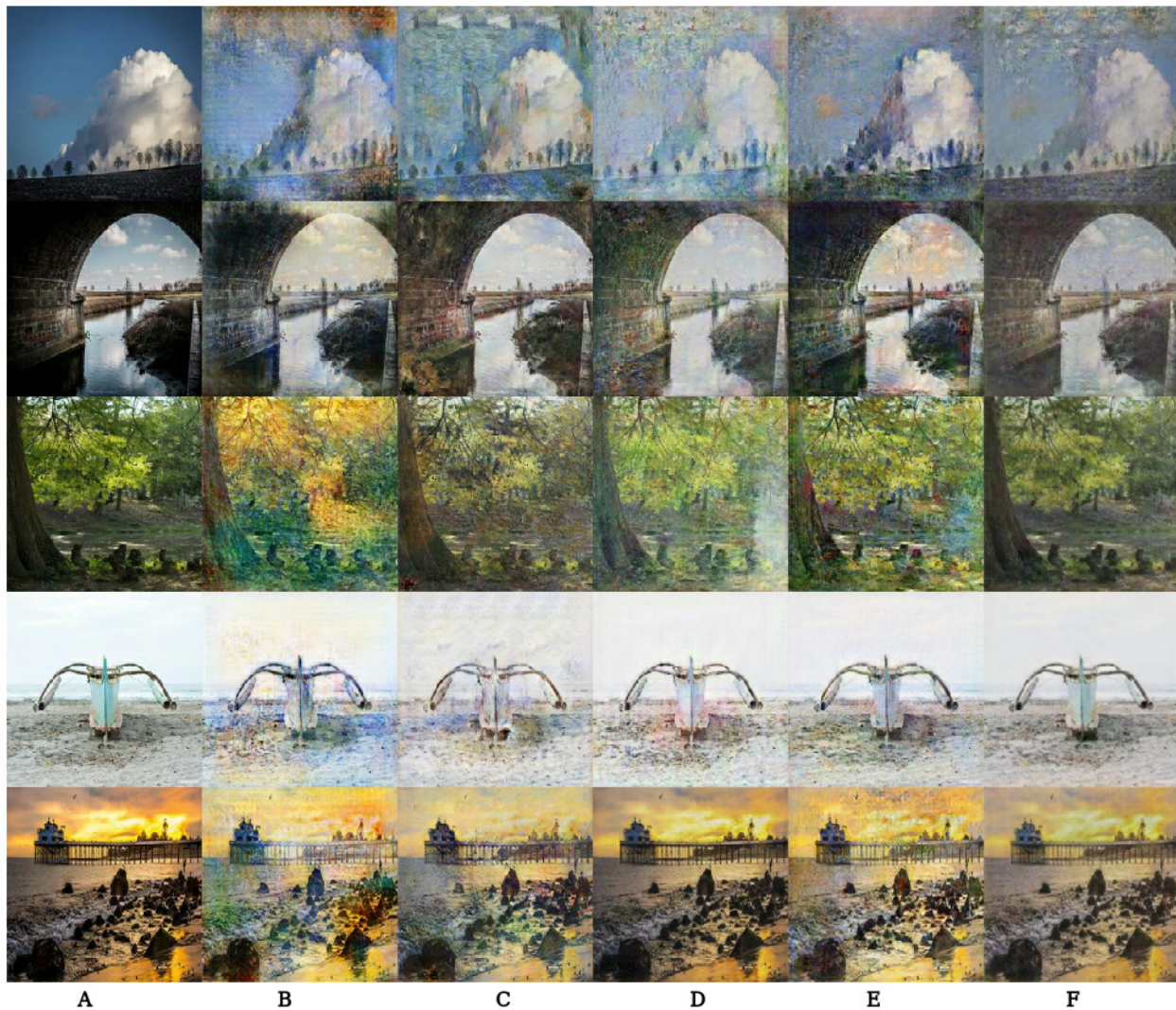
Fig. 3. Results obtained from our CycleGAN model: A) Original Image, B)Monet300 with 150 epochs, C) Monet900 with 50 epochs, D) Monet1172 with 75 epochs, E) Monet1337 with 100 epochs, F)Monet1969 with 50 epochs

Given a stride value of 2, the filter is applied to alternate pixels, resulting in a reduction of both the width and height dimensions by a factor of 2. In this study, instance normalization was employed as an alternative to batch normalizing.

The process of upsampling involves increasing the dimensions of an image, which is in contrast to downsampling where the dimensions are reduced. The Conv2DTranspose layer performs the inverse operation of the Conv2D layer.

The initial step of the generator involves downsampling the input image, followed by upsampling while simultaneously establishing lengthy skip connections.

Skip connections are employed to mitigate the issue of vanishing gradient by integrating the output of a layer with numerous layers through concatenation, rather than solely connecting it to a single layer.
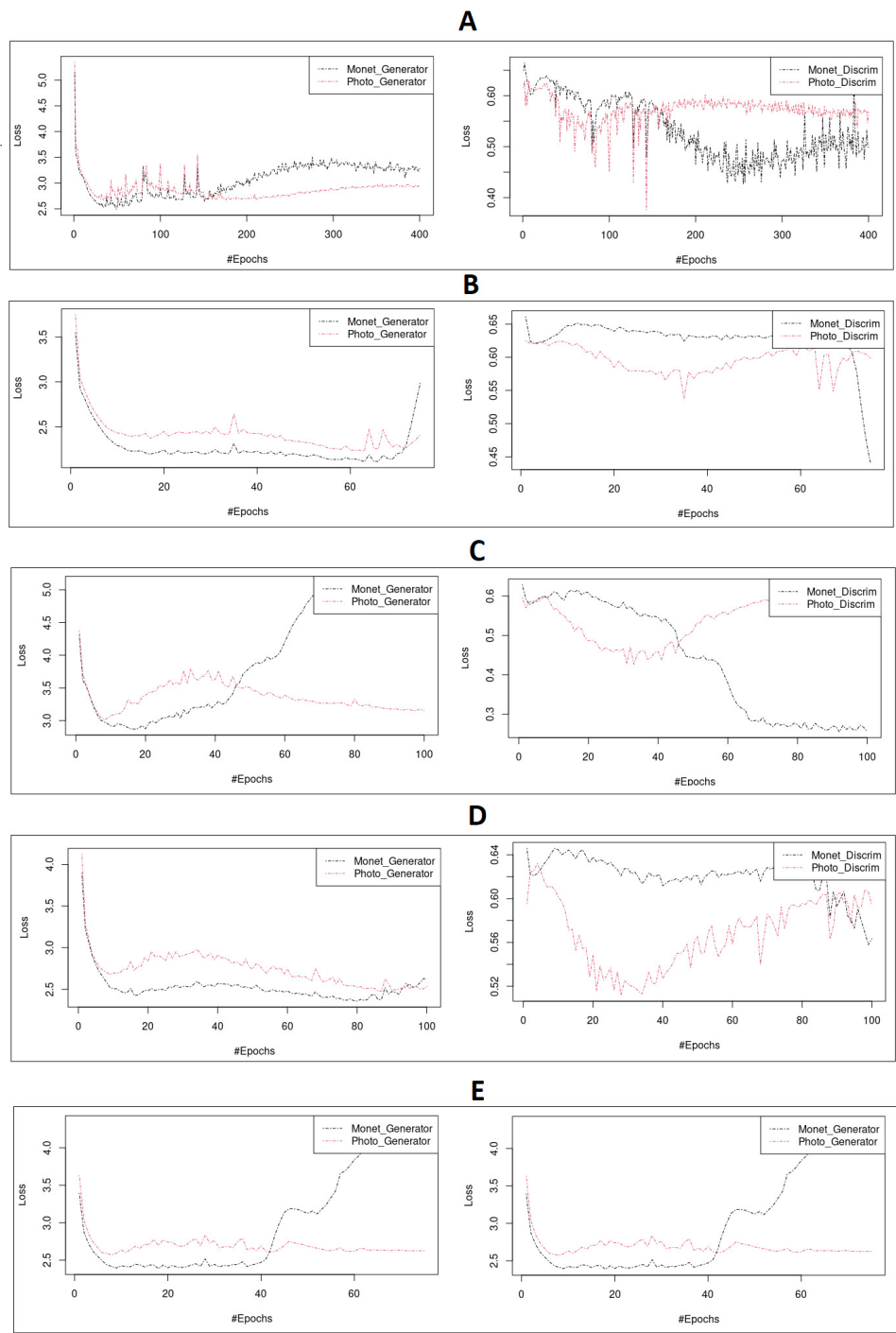
Fig. 4. Loss function (Top: Generators, Bottom: Discriminators): A) Monet300, B) Monet900, C) Monet1172, D) Monet1337, E) Monet1969

In this process, the output of the downsample layer is symmetrically concatenated with the output of the upsample layer. The architecture of the discriminator employs the PatchGAN discriminator.

The distinction between a PatchGAN and a conventional GAN discriminator lies in their respective mapping functions. In the case of a standard GAN, the mapping is performed from a 256×256 picture to a singular scalar output, which serves as an indicator of authenticity ("real" or "fake").

On the other hand, the PatchGAN operates by mapping from a 256×256 image to a different output, which encompasses many patches inside the image to an N $\times$ N (here 64×64) array of outputs X, where each X_{ij} represents whether the patch i, j in the image is real or fake—first, a 4×4 convolution-InstanceNorm-LeakyReLU layer with 128, 256 and 512 filters and stride of size 2.

InstanceNorm on the first layer of 64 filters is not applied. After the last layer, we apply convolution operation to produce a 1×1 output. The general design of the implemented architecture for this work is shown in Fig. 1.

3.4 Datasets and Pre-Processing

All the images for the model were of size 256 x 256 pixels and processed into TFRecord files with a batch size of 25. Additionally, the images were scaled to a [-1, 1] scale.

Because we were building a generative model, we do not need the labels, so we will only return the image from the TFRecord. The description of the used datasets are shown in Table 1 and a sample of the dataset contents are shown in Fig. 2.

3.5 Performance Metric

Memorization-informed Frechet Inception Distance (MiFID): Bai et al. (2021) [3] conduct the first generative model competition. The researchers made adaptations to the Frechet Inception Distance (FID) metric in order to impose penalties on models that generate images that closely resemble the training dataset:

$$\mathrm{MiFID}\left(S_g, S_t\right) = m_\tau\left(S_g, S_t\right) \cdot \mathrm{FID}\left(S_g, S_t\right), \quad (7)$$

where S_g is the generated set and S_t is the original training set. m_τ is the memorization penalty which is based on thresholding the memorization distance s of generated and true distribution defined as:

$$s\left(S_g, S_t\right) = \frac{1}{|S_g|} \sum_{x_g \in S_g} \min_{x_t \in S_t}\left(1 - \frac{|\langle x_g, x_t\rangle|}{|x_g| \cdot |x_t|}\right), \quad (8)$$

$$m_\tau(S_g, S_t) = \begin{cases} \dfrac{1}{s(S_g, S_t) + \epsilon} & (\epsilon \ll 1), \text{ if } s < \tau, \\ 1 & \text{otherwise.} \end{cases} \quad (9)$$

Lower memorization distance is associated with more severe training sample memorization.

4 Experiments and Results

We train our CycleGAN model with the different datasets shown in Table 1 and evaluate their performance using the MiFID metric.

Training of the model was stopped when performance of the MiFID score lowered, giving us an indication of mode collapse and overfitting affecting the performance of the model [26].

The best results of the various performed experiments are shown in Table 2 and compared with Kaggle competition leaderboard shown in Table 3[1]. The results from Table 2 synthesize the most relevant experiments where we show the progression of the model's performance, reporting the best result and intermediate results between it.

For example, for Monet300 the best result obtained was with 150 training epochs with a MiFID metric of 43.9139 and the values of epochs 100 and 175 with a MiFID score of and epochs 48.2689 and 44.9882 respectively show that the performance of the model will not provide us with more training improvement due to overfitting and mode collapse.

This can be seen more clearly in Figure 4 where the loss functions of the generators and discriminators of our CycleGAN model are shown with the different datasets over several training epochs.

---

[1]103 teams Kaggle competition leaderboard "I'm Something of a Painter Myself.Use GANs to create art - will you be the next Monet?". Date: August 3rd,2023.

It can be observed that despite being the same model architecture, its performance is significantly affected by the training dataset, where the best result was obtained with the Monet900 dataset with a MiFID score of 34.82910.

This is quite a good performance positioning us in the 2nd place in the Kaggle[1] competition when we submit our results. An adequate and representative selection of the transfer domain, in this case, Monet images, can generate better results despite having fewer images for training.

In the same way, we notice that the use of data augmentation strategies in Monet1197 and Monet1969 showed a similar performance or below the Monet300 dataset.

The performance of a model may depend not only on its architecture but also on the quality of the training dataset. It is essential to carefully consider and prepare the dataset to generate a representative sample that effectively represents the desired distribution.

This insight has been highlighted by recent studies [16]. The dataset Monet1969 comprised the most extensive collection of Monet paintings from the evaluated datasets with the worst performance.

This is counterintuitive since one would expect that, as in the CNN models, having a more extensive dataset would imply a better model performance [19]. However, a handpicked sample performs better for the same model CycleGAN architecture in the case of our experiments.

The behavior of the loss function of the generators and discriminators, Monet to Photo and Photo to Monet, is shown in Fig. 4.

It is interesting to notice that although the loss functions are not very useful in evaluating the performance of the CycleGAN models, they can provide the necessary information, such as if there is overfitting in the generators and poor training of the discriminants.

In the same way, since the convergence in the CycleGAN models does not usually occur as in the CNN models, we have that the best performance of the model usually occurs at the intersection of the loss functions of the two discriminators and of the two generators [26].

Fig. 3 shows some comparative results of the best outputs with the experiments of each dataset. Visually it can be seen that overall all results are pretty good in high-contrast images but have less convincing transformations when faced with complex cross-domain relationships such as medium and low-contrast images.

5 Discussion

In our work, we explored the impact of different data augmentation strategies on the effectiveness of the CycleGAN model.

As CycleGAN relies on unpaired image-to-image translation, it presents unique challenges in handling data diversity and robustness. Data augmentation has been acknowledged as an effective technique in traditional supervised learning for increasing the robustness of models by creating more diverse training samples.

Our experiments revealed that implementing these data augmentation strategies significantly improved the model's robustness against overfitting.

We observed an enhanced ability of the model to generalize across diverse transformations and variations in the source and target images. We compared the performance of the CycleGAN model with four different data augmentation approaches and showed that not all augmentation strategies were equally beneficial.

For instance, the data augmentation strategy of flipping, mirroring, cropping, and contrast enhancement in Monet1172 did not contribute significantly to model performance. In contrast, the augmentation approach in Monet900, by carefully selecting the training data, obtained a better performance with a MiFID performance of 34.8291 that managed to position itself in second place in the Kaggle competition [1].

Limitations of the CycleGAN model were observed, such as the overfitting problem and mode collapse of the generator, which limits our ability to obtain better results since the CycleGAN model learns deterministic mappings.

---

[1] 103 teams Kaggle competition leaderboard "I'm Something of a Painter Myself.Use GANs to create art - will you be the next Monet?". Date: August 3rd,2023

6 Conclusion and Future Work

In this work, we performed a comparison with different data augmentation strategies in the performance of a CycleGAN model for generating Monetesque-style images. We managed to show that the importance of the dataset is not related to its size but to the quality of data representation concerning the distribution of interest to be generated, in this case, Monetesque images.

For future works, to address these issues and improve results, the use of multimodal and multi-domain models such as Augmented CycleGAN, Mode Seeking Generative Adversarial Networks (MSGAN), and Domain-supervised GAN (DosGAN) have been proposed.

Also, reinforcement learning has been considered to make the CycleGAN model more robust to mode collapse and correct the overfitting of the generators.

Dataset and Code

The code used in this paper and the dataset is available[2], with also some other transfer style experimnets run on the artworks of Cezanne, Van Gogh, Diego Rivera, Ukiyo style and Hokusai.

Acknowledgments

References

1. **Almahairi, A., Rajeshwar, S., Sordoni, A., Bachman, P., Courville, A. (2018).** Augmented CycleGAN: Learning many-to-many mappings from unpaired data. International conference on machine learning, pp. 195–204. DOI: 10.48550/ARXIV.1802.10151.

2. **Antoniou, A., Storkey, A., Edwards, H. (2017).** Data augmentation generative adversarial networks. DOI: 10.48550/ARXIV.1711.04340.

3. **Bai, C. Y., Lin, H. T., Raffel, C., Kan, W. C. W. (2021).** On training sample memorization: Lessons from benchmarking generative modeling with a large-scale competition. Proceedings of the 27th Association for Computing Machinery Special Interest Group on Knowledge Discovery and Data Mining Conference on Knowledge Discovery and Data Mining, pp. 2534–2542.

4. **Bao, F., Neumann, M., Vu, N. T. (2019).** CycleGAN-based emotion style transfer as data augmentation for speech emotion recognition. Proceedings of Interspeech, pp. 2828–2832. DOI: 10.21437/Interspeech.2019-2293.

5. **Branikas, E., Murray, P., West, G. (2023).** A novel data augmentation method for improved visual crack detection using generative adversarial networks. IEEE Access, Vol. 11, pp. 22051–22059. DOI: 10.1109/access.2023.3251988.

6. **Ding, Z., Jiang, S., Zhao, J. (2022).** Take a close look at mode collapse and vanishing gradient in GAN. IEEE 2nd International Conference on Electronic Technology, Communication and Information, pp. 597–602. DOI: 10.1109/icetci55101.2022.9832406.

7. **Eslami, S., Williams, C. (2012).** A generative model for parts-based object segmentation. Advances in Neural Information Processing Systems, Vol. 25.

8. **Ganguli, S., Garzon, P., Glaser, N. (2019).** GeoGAN: A conditional GAN with reconstruction and style loss to generate standard layer of maps from satellite images. DOI: 10.48550/ARXIV.1902.05611.

9. **Gao, P., Tian, T., Li, L., Ma, J., Tian, J. (2021).** DE-CycleGAN: An object enhancement network for weak vehicle detection in satellite images. IEEE Journal of Selected Topics

---

²github.com/Lugo1025/Monet-CGAN

in Applied Earth Observations and Remote Sensing, Vol. 14, pp. 3403–3414.

10. **Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y. (2020).** Generative adversarial networks. Communications of the ACM, Vol. 63, No. 11, pp. 139–144. DOI: 10.48550/ARXIV.1406.2661.

11. **Lee, H., Lee, H., Hong, H., Bae, H., Lim, J. S., Kim, J. (2021).** Classification of focal liver lesions in ct images using convolutional neural networks with lesion information augmented patches and synthetic data augmentation. Medical Physics, Vol. 48, No. 9, pp. 5029–5046. DOI: 10.1002/mp.15118.

12. **Liang, D., Krishnan, R. G., Hoffman, M. D., Jebara, T. (2018).** Variational autoencoders for collaborative filtering. Proceedings of the 2018 world wide web conference, pp. 689–698. DOI: 10.48550/ARXIV.1802.05814.

13. **Liu, L., Pan, Z., Qiu, X., Peng, L. (2018).** SAR target classification with CycleGAN transferred simulated samples. IEEE International Geoscience and Remote Sensing Symposium, pp. 4411–4414. DOI: 10.1109/igarss.2018.8517866.

14. **Liu, Y., Chen, A., Shi, H., Huang, S., Zheng, W., Liu, Z., Zhang, Q., Yang, X. (2021).** CT synthesis from MRI using multi-cycle GAN for head-and-neck radiation therapy. Computerized Medical Imaging and Graphics, Vol. 91, pp. 101953. DOI: 10.1016/j.compmedimag.2021.101953.

15. **Martin, B., Edwards, K., Jeffrey, I., Gilmore, C. (2023).** Experimental microwave imaging system calibration via cycle-GAN. IEEE Transactions on Antennas and Propagation, Vol. 71, No. 9, pp. 7491–7503. DOI: 10.1109/tap.2023.3296915.

16. **Ngoc-Trung, T., Viet-Hung, T., Ngoc-Bao, N., Trung-Kien, N., Ngai-Man, C. (2021).** On data augmentation for GAN training. IEEE Transactions on Image Processing, Vol. 30, pp. 1882–1897. DOI: 10.1109/tip.2021.3049346.

17. **Pang, Y., Lin, J., Qin, T., Chen, Z. (2021).** Image-to-image translation: Methods and applications. IEEE Transactions on Multimedia, Vol. 24, pp. 3859–3881.

18. **Ratner, A. J., Ehrenberg, H., Hussain, Z., Dunnmon, J., Ré, C. (2017).** Learning to compose domain-specific transformations for data augmentation. Advances in Neural Information Processing Systems, Vol. 30. DOI: 10.48550/ARXIV.1709.01643.

19. **Shorten, C., Khoshgoftaar, T. M. (2019).** A survey on image data augmentation for deep learning. Journal of Big Data, Vol. 6, No. 1, pp. 1–48. DOI: 10.1186/s40537-019-0197-0.

20. **Teng, L., Fu, Z., Yao, Y. (2020).** Interactive translation in echocardiography training system with enhanced cycle-GAN. IEEE Access, Vol. 8, pp. 106147–106156. DOI: 10.1109/access.2020.3000666.

21. **Tran, N. T., Tran, V. H., Nguyen, N. B., Nguyen, T. K., Cheung, N. M. (2020).** Towards good practices for data augmentation in GAN training. arXiv:2006.05338.

22. **Wang, T. C., Liu, M. Y., Zhu, J. Y., Tao, A., Kautz, J., Catanzaro, B. (2018).** High-resolution image synthesis and semantic manipulation with conditional GANs. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8798–8807. DOI: 10.1109/cvpr.2018.00917.

23. **Welander, P., Karlsson, S., Eklund, A. (2018).** Generative adversarial networks for image-to-image translation on multi-contrast MR images-a comparison of cyclegan and unit. DOI: 10.48550/ARXIV.1806.07777.

24. **Yazici, Y., Foo, C. S., Winkler, S., Yap, K. H., Chandrasekhar, V. (2020).** Empirical analysis of overfitting and mode drop in gan training. IEEE International Conference on Image Processing, pp. 1651–1655.

25. Yi, Z., Zhang, H., Tan, P., Gong, M. (2017). DualGAN: Unsupervised dual learning for image-to-image translation. Proceedings of the IEEE International Conference on Computer Vision, pp. 2849–2857. DOI: 10.4 8550/ARXIV.1704.02510.

26. Zhu, J. Y., Park, T., Isola, P., Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. Proceedings of the IEEE international conference on computer vision, pp. 2223–2232. DOI: 10.1109/iccv.2017.244.