# Ablation Study of CycleGAN with Earth Mover's Distance

Cheng Duan
300060887
*School of EECS*
*University of Ottawa*
Ottawa, Canada
cduan092@uottawa.ca

Boming Shi
300072593
*School of EECS*
*University of Ottawa*
Ottawa, Canada
bshi099@uottawa.ca

***Abstract*—Generative adversarial network(GAN) has achieved great progress in many applications. Also the combination of "top-down" discriminator and "bottom-up" generator has provided a new thought to generative model. In the paper "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks", Zhu et al. presented cycleGAN to implement unpaired image translation. Although the results shown in the paper are impressing, there are some drawbacks in cycleGAN, such as mode collapse and difficulty in transfiguration. In our paper, our goal is to explore the performance of cycle consistency loss and improve cycleGAN with Wasserstein GAN.**

## I. Introduction

### A. Background and Motivation

Transfer learning is a typical machine learning approach which aims to use the knowledge learned from a domain A to another domain B. One of major applications of applying the idea of transfer learning is image translation. In fact, the problem of image-to-image transferring has largely drawn researcher's attention for many years [11] [6] [4].

GANs (Generative adversarial networks) is one of the most popular deep learning tools nowadays since it equips computer with the ability to learn, modify and create pictures or videos. Based on cGAN, Pix2Pix [4] is part of the reason why GANs become popular. While in Zhu et al. [11] research which is inspired by Pix2Pix, image-to-image translation makes another step forward by using CycleGAN which takes unpaired images as input datasets.

There are two major reasons so that unpaired image-to-image translation is more practical. Firstly, it is costly to obtain and collect enough paired images for network training. Without enough number of training instances, the trained network is not reliable. This also restricts the possibilities of applying image translation mechanism into real-world application. Apart from that, unbalanced number of images for input domain and target domain is another vital issue. One of the major task in image translation is to learn the differences between input domain and target in terms of data structures and distributions so as to construct correct mapping functions. However, while there are sufficient number of input domain images and insufficient number of target domain images, the model and algorithm cannot comprehensively learn the required knowledge.

### B. Objectives

Zhu's paper shows many remarkable results of applying the model in different applications. However, in their experiments, we noticed several fail images were produced appearing due to mode collapse. By checking through their network design, they did apply two methods including adopting Shrivastava et als strategy in network and replacing negative log likelihood by a least square loss.

Our contribution is to supply the ablation studying did by authors. Based on observations above, we intend to using a WGAN which was presented to be able to prevent mode collapse to replace original GAN network. More specifically, we firstly replaced original loss functions with WGAN's loss function. In addition, the training process was modified to adapt a clip value in order to restrict parameter range.

## II. Related Work

### A. GAN Network

Generative Adversarial Networks (GANs) [2] has been widely explored in image generation and have shown very good results. On image-to-image translation, several state-of-art approaches have been investigated and proposed [9] [7] [5] [3]. The goal of a standard GAN is to solve the following value function:

$$\min_G \max_D V(D,G) = E_{x\sim P_{data}(x)} log(D(x)) + E_{z\sim P_z(z)} log(1 - D(G(z)))$$

### B. cycleGAN

One of the impressive method "CycleGAN" was developed by Zhu et al. [11]. The method can be applied on several different applications including object transfiguration (figure 1), season transferring (figure 2) and photo generation from paintings (figure 3).

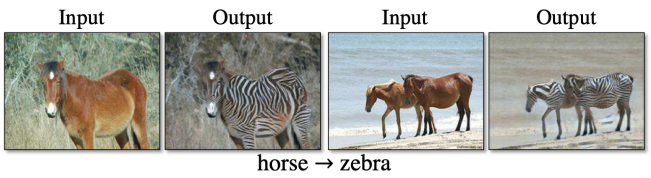


Fig. 1: Object transfiguration

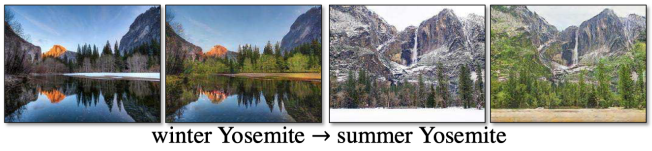


Fig. 2: Season transferring

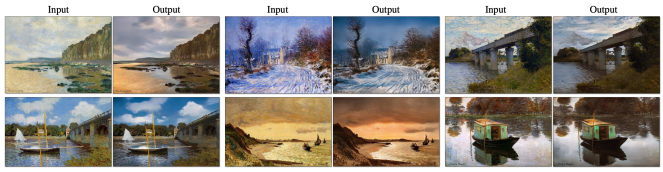


Fig. 3: Photo generation from paintings

The goal is to learn the image mapping from an input domain to the target domain where paired training data are not available [11]. More specifically, as shown in 10 they proposed cycleGAN which could learn a mapping function $G : X- > Y$ such that the distribution of images from G(X) is indistinguishable from the distribution [6]. In addition, they coupled the function G with an inverse mapping function $F : Y- > X$. Also, they introduce a cycle consistency loss to push $F(G(X)) = X$ and vice versa [6].

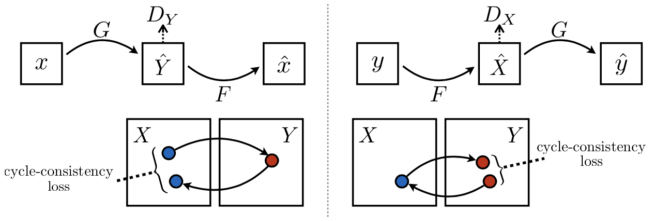


Fig. 4: Image at left hand side shows the training process of mapping from domain x to $\hat{Y}$ using mapping function G and mapping from y to $\hat{x}$ using mapping function F. Besides, specifically define cycle-consistency loss. The image at right hand side shows similar process.

Qualitative results are presented in [11] on several tasks where paired training data does not exist, including collection style transfer, object transguration, season transfer, photo enhancement, etc. Quantitative comparisons against several prior methods demonstrate the superiority of our approach. The idea of cycleGAN was inspired by Pix2Pix. With the benet of image-to-image translation, they made another step forward by taking the idea of cycle consistency into their loss calculation.

For the G net, the author suggests a network with high performance on neural style transfer and super-resolution. For the D net, a 70 * 70 Patch-GANs with the high performance was mentioned for classifying whether 70 * 70 overlapping image patches are real or fake.

In order to achieve the objectives, the model firstly tries to optimize the adversarial loss of two mapping functions:

$$\mathcal{L}_{GAN}(G, D_Y, X, Y) = E_{y \sim P_{data}(y)}[logD_Y(y)] + E_{x \sim P_{data}(x)}[log(1 - D_Y(G(x)))]$$
$$\mathcal{L}_{GAN}(F, D_X, X, Y) = E_{x \sim P_{data}(x)}[logD_X(x)] + E_{y \sim P_{data}(y)}[log(1 - D_X(F(y)))]$$

Besides, the cycle consistency loss includes both forward and backward cycle consistency loss:

$$\mathcal{L}_{cyc}(G, F) = E_{x \sim P_{data}(x)}[\|F(G(x)) - x\|_1] + E_{y \sim P_{data}(y)}[\|G(F(y)) - y\|_1]$$

are optimised using L1. Overall, the full loss function to be optimised is:

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{GAN}(G, D_Y, X, Y) = \mathcal{L}_{GAN}(F, D_X, Y, X) = \lambda \mathcal{L}_{cyc}(G, F)$$

Besides, in Harry et al. implementation version of cycleGAN, they improved the original algorithm in terms of convergence speed by adding skip connection between input and output in the G net.

In addition, in [4], Kim et al. transited from supervised nature to the unsupervised nature of the problem. They used samples from 2 dierent domains and discovered the relations between them. It showed a great advantage of detecting multiple representations of the same image which can be used to increase the accuracy of downstream applications.

### C. Wasserstein GAN

WGAN (Wasserstein GAN) was introduced by Martin et al. [1] in 2017. In the paper the author addressed several problems of GANs and one of them is the instability of GAN's training process which also named as mode collapse. Martin and his co-authors managed to propose a new distance measurement - EMD (Earth Mover's Distance) which has the ability to avoid vanishing gradient (figure 5). Thus the model collapse problem can be prevented. The goal of WGAN is to solve:

$$W(P_{data}, P_G) = \max_{D \in 1-Lipschitz} (E_{x \sim P_{data}} D(x) - E_{x \sim P_G} D(x))$$

The purpose of $1 - Lipschitz$ function is to smooth the discriminator. (figure 5) shows that discriminator of GAN varies severely, when discriminator of WGAN varies smoothly, so GAN tends to generate similar images which are very close to the training samples while WGAN is able to generate enhance the variety of images.

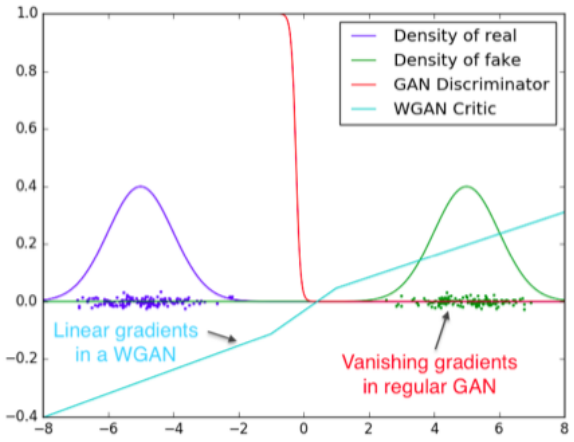


Fig. 5: Gradient signals of regular GAN and WGAN in distinguishing two Gaussian

The author also proposed a simple example to state why EMD (the distance measurement used by WGAN) is better than JSD (the distance measurement used by regular GANs). The settings [1] was set as there is a $Z$ which is uniformly distributed over $(0, 1)$. Let $p_d$ be the distribution of $(0, Z) \in R^2$. Let $G_\theta(Z) := (\theta, z) \in R^2$ and $p_g$ be the distribution of $G(Z)$. Then the results show that:

$$EMD(p_d, p_g) = |\theta|$$

$$JSD(p_d, p_g) = \begin{cases} log2 & if \theta \neq 0 \\ 0 & otherwise \end{cases}$$

It is obvious that the JSD is not differentiable with parameter $\theta$, while EMD can. Such phenomenon ensures that a GAN training with EMD minimisation is trainable throughout entire process even reach the end of the training. In other words, the vanishing problem will not occur and so that mode collapse problem will not happen.

## III. Method

### A. Datasets

In the paper, the authors used a imageNet to train and test their model. However, it is not practical for this project to experiment with same settings. Therefore, other smaller alternatives will be used for the training and testing in this project.

For the training and testing in early stage, MNIST will be used. In later stage, a standard horse2zebra [10] dataset is available for the experiments of this project. It's a 117.9M set containing training and testing pictures (figures 6, 7).

Fig. 6: Horse Images

Fig. 7: Zebra Images

The horse images were set as input domain which including 120 images. The zebra images were set as target domain which including 140 images. All of them are RGB images with dimension of 256 by 256.

### B. Network Architecture

The network we used is the ones from original cycleGAN paper [8]. For the G net, the author adopts a network with high performance on neural style transfer and super-resolution. For the D net, a $70 \times 70$ Patch-GANs with high performance was mentioned in the paper for classifying whether $70 \times 70$ overlapping image patches are real or fake [11].

### C. Ablation Study

The authors performed a ablation study so as to prove that by adding cycle consistency to their loss function, the performance of image translation increased. The study was designed by generating 5 sets of images while one of them using the model trained with full loss function. The rest 4 set were generated using models with different part of the full loss function.

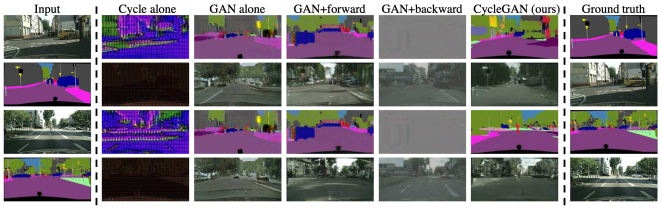


Fig. 8: Ablation study performed by original authors

A drawback of this ablation study is that the images generated with part of loss function are actually useless as the problem of mode collapse happened. It is not highly persuasive to conclude cycle consistency is vital in this model by comparing these images. Therefore, we modified original cycleGAN model by replacing GAN with WGAN and reproducing ablation study so as to learn by what extent the cycle consistency helped during training.

### D. Modification of cycleGAN

In order to adopt WGAN into the model, several modifications on loss function and training process were made based on original cycleGAN. To take advantage of Google Colab's free GPU, we transferred original Python implementation into a jupyter notebook and made related modifications within the notebook.

*1) Loss Function:* The original loss function used for training is the regular loss function for regular GAN networks which has been shown in section II-B. The loss function is composed of two sets of adversarial loss and two cycle consistency losses including one forward cycle consistency loss and one backward cycle consistency loss.

During the modification, both of the modifications will remain same. The adversarial losses will be modified. More specifically, we intend to modify the loss function for discriminators. Originally, the goal of discriminator is to maximising following scores of sample from $P_{data}$ and minimising scores of data from $P_G$ which can be formulated as:

$$V(G,D) = E_{X\sim P_{data}(X)} log(D(x)) + E_{X\sim P_G(X)} log(1-D(x))$$

When G is fixed, it is to:

$$D^* = arg\max_D V(G,D)$$

The above function can be represented with KL divergence as:

$$\max_D V(G,D) = -log4 + 2JSD(P_{data}\|P_G)$$

With this function, we may replace the $JSD$ with  where $\beta$ is a positive value. It is to ensure that a function f is $\beta$-Lipschitz.

While in terms of implementation, we replaced original loss function of discriminator

```
> d_loss_a_real =
    tf.losses.mean_squared_error(a_logit,
    tf.ones_like(a_logit))
> d_loss_b2a_sample =
    tf.losses.mean_squared_error(b2a_sample_logit,
    tf.zeros_like(b2a_sample_logit))
> d_loss_a = d_loss_a_real + d_loss_b2a_sample
>
> d_loss_b_real =
    tf.losses.mean_squared_error(b_logit,
    tf.ones_like(b_logit))
> d_loss_a2b_sample =
    tf.losses.mean_squared_error(a2b_sample_logit,
    tf.zeros_like(a2b_sample_logit))
> d_loss_b = d_loss_b_real + d_loss_a2b_sample
```

with new loss functions

```
> wd_a = tf.reduce_mean(a_logit) -
    tf.reduce_mean(a2b_logit)
> d_loss_a = -wd_a
>
> wd_b = tf.reduce_mean(b_logit) -
    tf.reduce_mean(b2a_logit)
> d_loss_b = -wd_b
```

*2) Training Process:* The training process requires to be modified as well. The $\beta$ which mentioned in section III-D1 is also used here as the value of clip. The value is used to clip the value of each weight parameter of $f$ in range $[-\beta, +\beta]$ after every updating of $f$.

The modifications were implemented with the features of Tensorflow as follows:

```
> with
    tf.control_dependencies([d_a_train_op]):
>    d_a_train_op = tf.group(*(tf.assign(var,
    tf.clip_by_value(var, -clip, clip)) for
    var in d_a_var))
```

## IV. Results and Discussion

### A. Experiment Design and Results

In order to to explore whether and how cycle consistency and WGAN contributes to the results, we conducted several different experiments with different loss functions: WGAN loss without cycle consistency loss, WGAN loss with forward cycle consistency loss, WGAN loss with backward cycle consistency loss, WGAN with full loss, WGAN with full loss. Sample images are showed as following, the left columns are input images, medium columns are conversion and reconstruction:

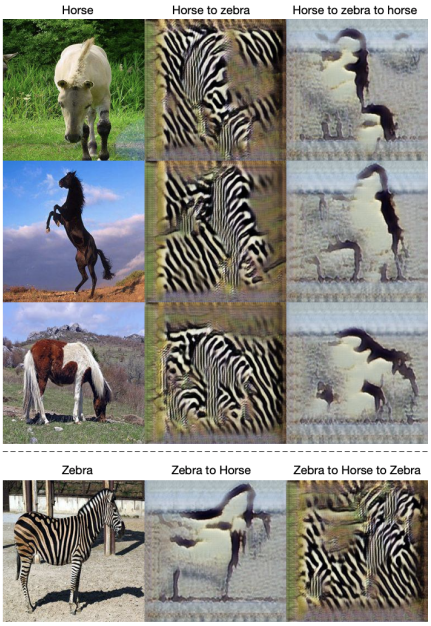


Fig. 9: WGAN alone

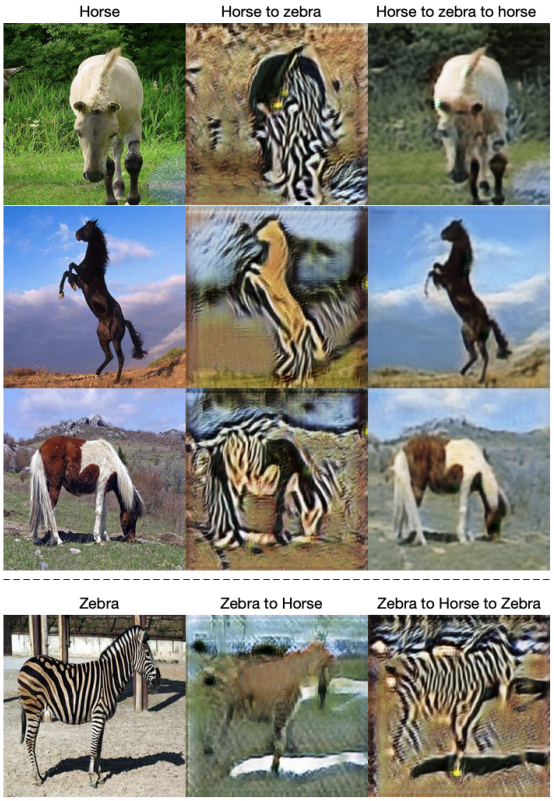


Fig. 10: WGAN + forward cycle consistency

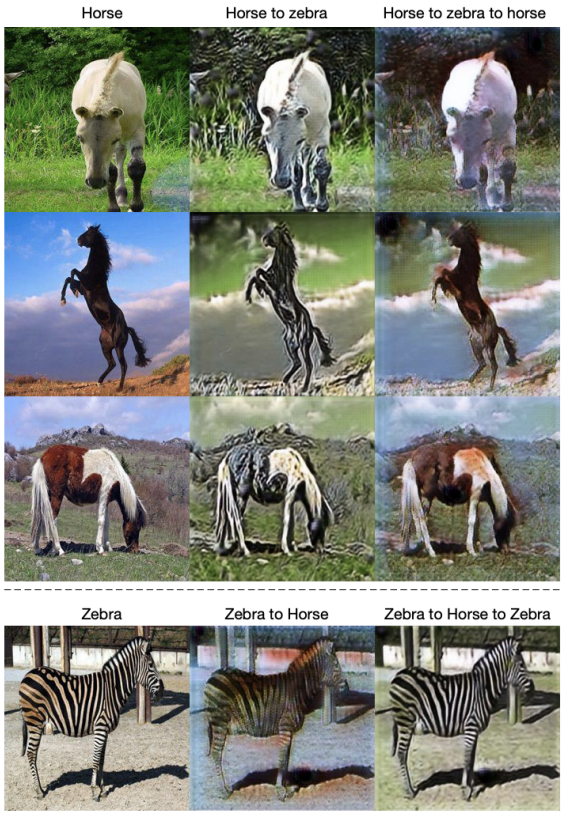


Fig. 11: WGAN + backward cycle consistency

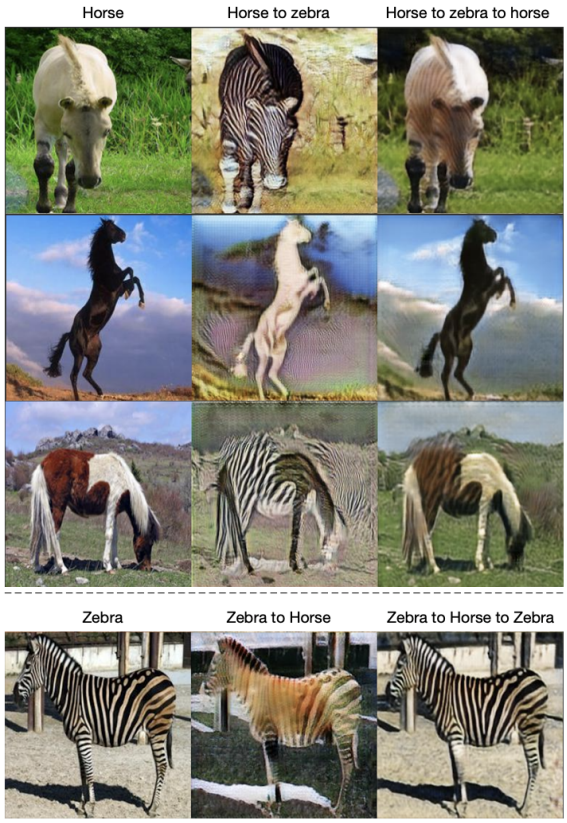


Fig. 12: Origianl cycleGAN


Fig. 13: WGAN + cycleConsistency

### B. Conversions and Reconstructions

From the results, it can be concluded that WGAN without cycle consistency has a severe mode collapse and has disability in generating paired images. With forward loss, the forward reconstruction will be improved while backward loss improve the performance of backward reconstruction. As is expected, WGAN with cycle consistency can convert between horse and zebra. With full cycle consistency loss, the reconstruction image will be almost same as the original image. The substitution of WGAN loss with WGAN loss does not improve the conversion result. However, the result shows that WGAN enhance the variety of generating and reduce mode collapse, which can be applied in other translation tasks such as transfigure.

## V. Conclusion

The modifications including new loss functions and training process on original cycleGAN network is successful. Comparing with original ablation study, the WGAN does solve the mode collapse problems to some extent. While using partial loss function, the network could reconstruct recognisable images instead of similar useless images due to mode collapse. Besides, it is pretty obvious that when all models were trained with same number of epochs and same datasets, original cycleGAN still have the best performance. Overall, WGAN solves mode collapse problem to some extent and supplied original ablation study which meets our objectives. However, one drawback of this experiment is the lack of datasets and limited computing resources. It is worthy to learn how far could the training of WGAN reach.

## Acknowledgment

We would like to thank Prof. Lang for his valuable guidance and suggestions towards our project.

## References

[1] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan. *arXiv preprint arXiv:1701.07875*, 2017.

[2] Jean Pouget-Abadie Mehdi Mirza Bing Xu David Warde-Farley Sherjil Ozair Aaron Courville Goodfellow, Ian and Yoshua Bengio. Generative adversarial nets. In *In Advances in neural information processing systems*, pages 2672–2680, 2014.

[3] Jun-Yan Zhu Tinghui Zhou Isola, Phillip and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *arXiv preprint*, 2017.

[4] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1125–1134, 2017.

[5] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Auto-encoding variational bayes. In *arXiv preprint*, 2013.

[6] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. Unsupervised image-to-image translation networks. In *Advances in Neural Information Processing Systems*, pages 700–708, 2017.

[7] Ming-Yu Liu and Oncel Tuzel. Coupled generative adversarial networks. In *Advances in neural information processing systems*, pages 469–477, 2016.

[8] LynnHo. LynnHo/CycleGAN-Tensorflow-PyTorch. https://github.com/LynnHo/CycleGAN-Tensorflow-PyTorch, May 2018.

[9] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. In *arXiv preprint arXiv:1411.1784*, 2014.

[10] Taesung Park. Cyclegan datasets. https://people.eecs.berkeley.edu/~taesung_park/CycleGAN/datasets/. Accessed:2017.

[11] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Computer Vision (ICCV), 2017 IEEE International Conference on*, pages 2242–2251. IEEE, 2017.