GROUP2 SECLEGAN: IMPROVEMENT OF THE CYCLEGAN WITH SEGMENTATION

Vince Chen[⋆], *Jin Wu*[⋆], *and Jiayi Luo* [⋆]

University of California San Diego, La Jolla, CA 92093-0238

ABSTRACT

CycleGAN model has been one of the most widely used machine learning model in the ML/DL field. However, it has certain limitations and would fail some of the cases. Therefore, in this paper, we would like to fix the original limitations that CycleGAN has. Thus we propose an upgrade based on the original CycleGAN model by adding one more segmentation layer right before the CycleGAN layer, which we name our method as SecleGAN model. Our model could perfectly solve the problem that original CycleGAN mistreated the part of background as the target as well and transform the background into other domains. In the end, our SecleGAN model is able to only transform the real target yet remains the background information as much as possible. Therefore, we consider our model as an improvement of the original CycleGAN.

Index Terms— GAN, CycleGAN, Semantic Segmentation

1. INTRODUCTION

Nowadays, CycleGAN[1] has gained wide attention and frequently been applied into many different fields. For example, it is known that CycleGAN algorithm has been used in diagnosing and recognizing lesions and tumors thanks to it's excellent performance augmentation. Moreover, it's also applied into image reconstruction because of its ability to perform image-to-image transformation. Specifically, Jack Clark used CycleGAN to convert ancient Babylon and London maps into modern Google Map Satellite views, which helps people better understand the layout of these ancient remains. Even Yann LeCun, the Godfather of the deep learning, said that ”GANs is the most interesting idea in the past 10 years in machine learning”. Therefore, our team truly believes that diving deep into the field of GANs is a very meaningful. This is the main reason that we initialize our project topic.

1.1. Problem

Although CycleGAN is a powerful algorithm and can be broadly generalized into different field, it still has some limitations that cannot be overlooked. For example, sometimes the generator of the CycleGAN collapses, and it would fail some of the cases. Specifically, in the below image, the CycleGAN means to transform the horse into zebra. However, the person who's riding the horse together with some background contents are also transformed into the zebra, which in turn results in this failure.

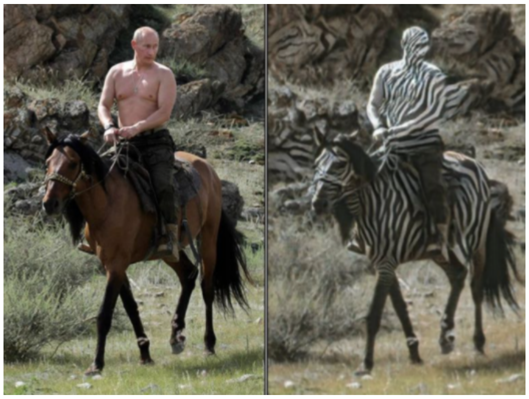

Fig. 1. The famous failure of CycleGAN

The above failure does not meet our expectation from CycleGAN. Therefore, we would like to fix this issue.

1.2. Proposed solution

Our team proposes a solution to solve the above problem, in which we name as the SecleGAN algorithm. SecleGAN is able to transform the target into other domain yet preserves the background information as much as possible.

The SecleGAN model consists of two main part: the CycleGAN part and semantic segmentation part. For the CycleGAN part, we remain the original CycleGAN model. For the semantic segmentation part, we used the U-net model. The reason that we chose U-net over other semantic image segmentation model is that U-net is fast to train and use. In the future, we aim to test SecleGAN model with other state-of-art semantic image segmentation models and compare the results.

Specifically, the process of going through the SecleGAN model is that, an input image with the transformed object will first be processed by the semantic image segmentation U-net,

[⋆] Ourselves

and the object instance will be pulled out of the background contents by the U-net unit. After that, the object instance alone will go through the traditional CycleGAN unit to be transformed into the target instance. Finally, we would then use a mask to put the transformed instance of the object back into the original background content. In this way, we will only transform what is needed and make sure that the CycleGAN algorithm does not contaminate the background and cause information loss.

2. DATASET

The dataset that we are using is Caltech-UCSD Birds-200-2011 dataset[2], which is an extended version of the original CUB-200 dataset. The Caltech-UCSD Birds dataset contains 11,788 images of birds from 200 different categories. Therefore, there are 200 different kinds of bird categories with different features, which is ideal for our purpose. Moreover, as an image dataset, the number of data in this dataset is sufficient and very useful in terms of training our model.

During the training stage, all of the 11,788 images will be used to train the SecleGAN model. As usual, we splitted the 80 percents as the training data and 20 percents as the testing data. And for our SecleGAN transformation stage, we used two classes from this dataset, which are Pine Grosbeak Yellow bellied Flycatcher. The reason for choosing these two categories is that they are highly contrasted in body colors, and is optimal to see the transformation results.

3. MODEL

3.1. Structure

The overall structure of our SecleGAN model is shown below.

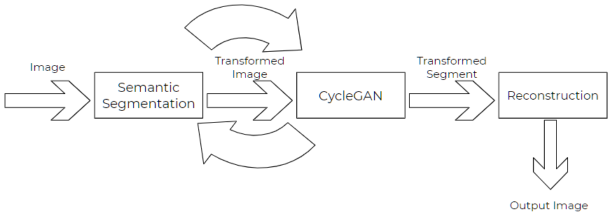

Fig. 2. SecleGAN architecture

To reiterate here, the input image would go through the semantic segmentation layer (U-net) and then the object without the background will be processed by the CycleGAN to get the transformed segment. Fianlly, the segment is put back to the original background using mask to reconstruct the output image.

We believe that similar result can be also achieved by swapping the order of the semantic segmentation and CycleGAN model, where the image will be first transformed by the CycleGAN. Often, part of the background after the direct transformation is also mis-transformed and cause information loss. We then use the U-net to pull the transformed instance out and use a mask to put it back to the original background to avoid information loss. However, during our implementation, we figured out that the accuracy of the U-net has decreased because we are feeding the transformed image into the U-net but not the original one. Often times, the transformed image has some background transformed as well, which will,in a sense, ’trick’ the U-net and thus result in less excellent result.

3.2. CycleGAN

CycleGAN[1] is an extension of GAN[3] architecture.

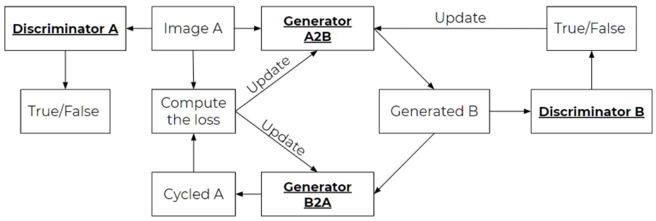

Fig. 3. CycleGAN network architecture in forward cycle

The CycleGAN architecture contains two separate GANs running in cycles. The two cycles are denoted as forward cycle and backward cycle. In forward cycle, the input image from domain A will be used to train the discriminator A, then it’s transformed into domain B, the target domain, by the generator A2B, which the result we denote as transformed B image. Unlike the tradition GAN model which constantly update the generator using weights, the transformed B image will then be transformed back to domain A by another generator B2A. And the result that comes out of generator B2A is denoted as reconstructed A image. Lastly, the CycleGAN would compare the reconstructed A image with the original A image to compute the loss and update the model. In backward cycle, the exact same logic is applied for the image in domain B.

3.3. U-net

As explained in previous sections, our model uses U-net[4] as the semantic image segmentation model. Regular Convolutional Neural Network cannot achieve the output of the semantic segmentation since the output should be a reconstruced image rather than a single label, whereas the output of the regular CNN will be the classification labels or softmax outputs. It is notable that not only the convolutional layer is involved in U-net, but also the up-sample layer which is similar to the process of reconstructing new images.

After a considerable amount of research on the choice of semantic segmentation model, we think that U-net is the most

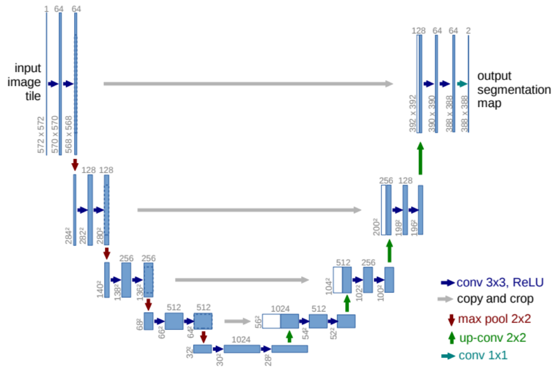

Fig. 4. U-net architecture

adequate neural network architecture for semantic segmentation. The detailed comparisons will be included in the next section.

4. RELATED WORK

First, our team compared with different semantic image segmentation model, and chose the U-net architecture as our segmentation model. Initially, we read some literature about Pix2Pix technique[5], Pix2Pix is a poplar method in image to image transformation. However, we figured out that this technique requires parallel data. Specifically, if we want to train our model to transform a landscape picture from winter to summer, we will need to have the images of the same landscape in both summer and winter. Moreover, if we want to train another model to transform a horse to zebra, we won't be able to obtain this parallel data since it's impossible to get the zebra pattern to show on a horse in real life. Despite the harsh requirement, Pix2Pix is still a powerful generative model, but it's not ideal for our SecleGAN model.

Fully Convolutional Networks[6] is another technique for semantic segmentation. It main idea is that the final dense layer of the network can be replaced by a 1x1 convolutional layer. However, the accuracy of the transformation for FCN is lower if we have a large size of input image. Therefore, to address this issue, we decided to use the U-net, which it shared the very similar structure as FCN but had improvement when dealing with large size input.

Some investigations about previous work on the improvement of the CycleGAN is also conducted. There had been some research aiming to improve many flaws that CycleGAN had. As described previously, CycleGAN required paired data to train the model and then perform transformation. Almahairi et.al developed a new method for CycleGAN to learn from unpaired data [7]. This is a notable improvement towards the limitation of the CycleGAN. However, our main focus is to fix the potential failure cases of the transformation, and using unpaired data will not necessarily aid our work.

Another work is to improve the loss function of the CycleGAN[8]. Jarda used L1 loss for the CycleGAN loss function, and added Adam optimization function as well for better results. As a result, Jarda was able to improve the transformation accuracy of the CycleGAN when testing using the same dataset. However, as he mentioned as well, his model could not solve our person to zebra problem.

After some research, we began to implement our SecleGAN model, and we believe that our model performs better yet remains as simple as it is.

5. EXPERIMENT

5.1. Models and Parameters

The parameters for Cycle GAN training involves λ, which is the weights factor that balance between L_{GAN} and L_{cyc} as defined previously. However, in order to achieve better performance and explore, we have introduced an identity loss and weighted the three losses, instead of two. The three losses are

Notation	Meaning
L_{GAN}	GAN loss(L_2)
$L_{identity}$	Identity loss(L_1)
L_{cyc}	Cycle loss(L_1)

Table 1. Three kinds of losses been weighted

After trails and evaluation with cross-validation, the weightsused in this experiment is $[1, 5, 10]$, respectively. So the overall loss is defined as:

$$L = L_{GAN} + 5 * L_{identity} + 10 * L_{cyc} \tag{1}$$

Identity loss is computed by compare the original image from A and transformed images from B to A, using $G_{B2A}(A)$.

In addition to weight factors, the learning rate is also important. We found out it is very easy to miss the local minimum/maximum if a larger step is taken. The loss function will bouncing around, instead of ”monotonically” decreasing. After trails with

$$\begin{aligned} \eta \in &(0.0001, 0.0002, 0.0005, 0.001, \\ &0.002, 0.005, 0.01, 0.02, 0.05, 0.1) \end{aligned} \tag{2}$$

we set the learning rate to $\eta = 0.0002$ eventually. Note that the learning rate is small because the loss is very sensitive to learning rate.

5.2. Loss plots and Final Results

Below are the loss plots of CycleGAN at 450 epochs. Note that there will be two generators and two discriminators for a single CycleGAN model. Moreover, we plots the losses for each discriminators with original data and reconstructed data to better demonstrate our model.

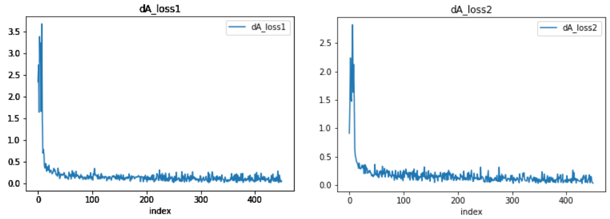

Fig. 5. Discriminator A with OG data and reconstructed data

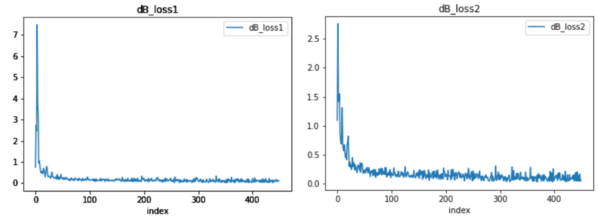

Fig. 6. Discriminator B with OG data and reconstructed data

Besides, we also have the loss plots for our two Generators.

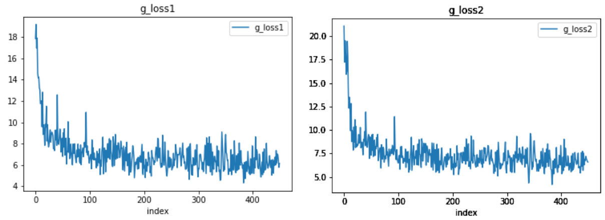

Fig. 7. Loss of two generators

After training with CycleGAN model, we also trained the U-net model, since the U-net model's accuracy maintained at high level after roughly 15 epochs, we trained it for 20 epochs.

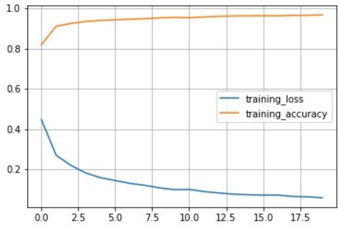

Fig. 8. Accuracy and Loss curve for U-net

Below shows our final transformation results using SecleGAN, and we can see that our model has a notable improvement compared with original CycleGAN model, where our model preserves much more background information.

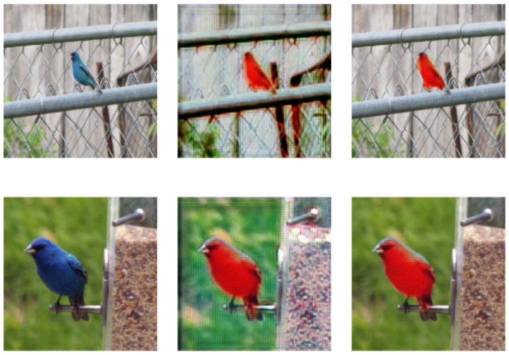

Fig. 9. Original Result vs CycleGAN vs SecleGAN

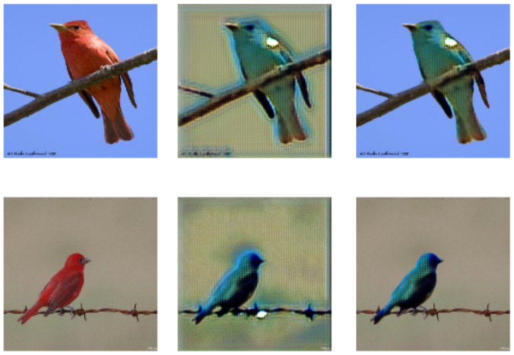

Fig. 10. Original Result vs CycleGAN vs SecleGAN

Figure 9 shows the transformation results from class A to class B, and figure 10 shows the one from class B to class A. Needlessly to say, SecleGAN outperforms CycleGAN by very much.

6. CONCLUSION

In the end, we managed to improve the original CycleGAN by using the U-net to preserves the background information as much as possible and use the mask to prevent the CycleGAN from contaminating the background as well. CycleGAN itself is a powerful tool for the image-to-image transformation, and the biggest weakness is that it sometimes transform the background contents that should not be transformed. By using our SecleGAN model, this problem could be resolved if our dataset comes with the mask. In the future, we will also try to upgrade our SecleGAN and hopefully it could fit all different kinds of datasets and produce much better results compared to CycleGAN.

7. REFERENCES

[1] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *CoRR*, abs/1703.10593, 2017.

[2] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The Caltech-UCSD Birds-200-2011 Dataset. Technical report, 2011.

[3] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.

[4] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015.

[5] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. *CoRR*, abs/1611.07004, 2016.

[6] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation, 2015.

[7] Amjad Almahairi, Sai Rajeswar, Alessandro Sordoni, Philip Bachman, and Aaron Courville. Augmented cyclegan: Learning many-to-many mappings from unpaired data, 2018.

[8] Aamir Jarda. Improving the efficiency of the loss function in cycle-consistent adversarial networks.

8. CONTRIBUTION

All of the team members are self-motivated and working hard for this project. Jin Wu is the one who came up with this entire SecleGAN idea, which we think it's novel. He also worked on the CycleGAN part, where he read the literature, designed the CycleGAN network and train the network for over 450 epochs. Vince Chen worked on the U-net part, where he also read the literature, built the U-net architecture from the literature and trained the U-net architecture. Moreover, he also helped working on using the mask to pick out the instance. Jiayi Luo selected the dataset and work on the data preprocessing, and he also work on using the mask to pick out the instances of from a picture. Finally, he connected the CycleGAN, U-net together and make sure that the entire flow is working to produce ideal output. For the report part, we all grouped together to write the report.

9. REPLY TO REVIEW

Answer to group 25:

- I believe we could only qualitatively measure the performance, so does the loss since for now we do not think there's a proper metric for measuring the loss for our cases.
- Answered above.
- We used two classes for training CycleGAN but used the entire dataset to train the U-net.
- We've done the literature review and also tried using other semantic segmentation models, and so far U-net is the best semantic segmentation model.
- This is an interesting topic, we will try this one later.
- We've tried our best. However, if we are using mask, the boundary is often obvious.

Answer to group 32:

- Yes, the approach worked well for other varieties of classes as well, we have tried and it worked well. The reason that we chose these two classes is that the colors of these types of birds are highly in contrast, which helps visualization.
- It's explained in our paper in detail. Thanks for the suggestion.
- So far, we could only visualize the results but we can compare the results that come from CycleGAN and SecleGAN and they are different qualitatively.
- Yes, but we believe that SecleGAN model will also work for the Horse2Zebra dataset. The reason that we use bird dataset instead of Horse2Zebra dataset is that it comes with masks and it's easy for operation. However, in the future, we will continue to test our model in different benchmark dataset and compare the results with CycleGAN.

Answer to group 33:

- No, the model needs to work with successful segmentation.
- Yes, it works for any algorithms that is capable of doing image-to-image transformation.
- No, there is no other work that combines the segmentation with the GAN model, and that's why we believe our idea is novel.
- it's explained in our paper in deatil.
- Answered above at other groups' answers.