Shape Transformation with CycleGAN Using an Automobile as an Example

Akira Nakajima[a] and Hiroyuki Kobayashi[b]
Robotics & Design Engineering Osaka Institute of Technology University, Japan

Keywords: CycleGAN, Shape Transformation, Image Processing.

Abstract: AI technology has developed remarkably in recent years, and AI-based image generation tools have spread rapidly. CycleGAN is one of the image generation AIs and specializes in image style transformation, and has the problem of being able to change colors and patterns but not shapes. The reason may be that the model considers the background as a part of the conversion target, which can be solved by removing the background. In this study, the number of backgrounds is limited to a certain number, and CycleGAN is used for shape transformation.The evaluation is done by comparing the result of this experiment with the image transformation when the input is an image with the background removed.Comparison of the proposed and conventional methods showed comparable results.

1 INTRODUCTION

In recent years, deep learning has developed remarkably, and GAN(Goodfellow et al., 2014) is a technology that has been attracting a lot of attention.The method pix2pix(Isola et al., 2017) proposed by Isola et al. performs a style transformation such that a realistic object is generated from handwritten edges by obtaining transformation rules for each image pair and a unique loss function between each domain (a collection of images with the same features). In addition, CycleGAN(Zhu and Li, 2017), a method proposed by Zhu et al. performs Image-to-Image in the framework of unsupervised learning by removing pair constraints on training data from pix2pix. This allows learning the correspondence between domains and performing image style transformation as long as two domains with common features are available.

CycleGAN is good at transforming image styles such as color and pattern, etc. CycleGAN can capture the same shape as a feature and change its color or pattern, but it is not good at transformations involving shape changes. We limited the transformation target to cars and considered whether CycleGAN could perform shape transformation. It is known from the paper(Wu et al., 2019) that the reason why shape transformation is difficult is that the background of the image is recognized as a part of the object, and feature extraction cannot be performed well. This study performs CycleGAN shape transformation without removing the background by limiting the number of backgrounds.

2 METHOD

2.1 Image Processing

In this study, CycleGAN-based shape transformation was performed on a car. A box-shaped old car was transformed into a curved current car. The old and new cars were cropped from the original images and merged with 10 different landscape images. An example of the dataset is shown in Figures 1 and 2.Since some of the old cars had side mirrors attached to the tip of the hood, the side mirrors were removed to unify the features of the old cars.

Figure 1: Example data set of a new car.

[a] https://orcid.org/0009-0002-1142-9470
[b] https://orcid.org/0000-0002-4110-3570

Nakajima, A. and Kobayashi, H.
Shape Transformation with CycleGAN Using an Automobile as an Example.
DOI: 10.5220/0012233100003543
In *Proceedings of the 20th International Conference on Informatics in Control, Automation and Robotics (ICINCO 2023) - Volume 1*, pages 736-739
ISBN: 978-989-758-670-5; ISSN: 2184-2809

Figure 2: Example of data set for an old car.

2.2 CycleGAN

The model structures of CycleGAN generators and discriminators are shown in Figures 3 and 4. The generator consists of an encoder, a transformer, and a decoder.The encoder reduces the image size by convolution. The transformer transforms the domain. The decoder restores the image size to its original size by inverting convolution. The discriminator reduces the image size by convolution and extracts a wide range of features across the image by average pooling.

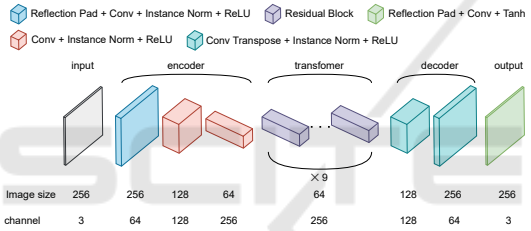

Figure 3: Generator structure.

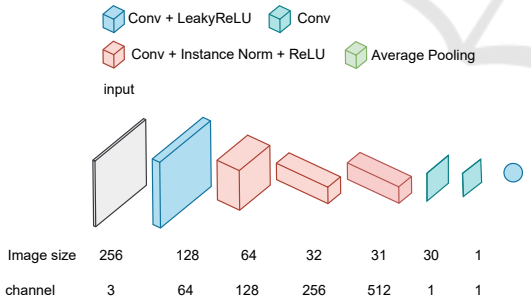

Figure 4: Discriminator structure.

The learning of CycleGAN is shown in Figures 5. Two sets of images, A and B, are needed to train the model. The model consists of four networks, two each of generators and discriminators. The generator performs the image transformation between the two domains. The discriminator identifies whether the input belonging to A or B is a real one in the data set or a fake one generated by the generator, and returns a value close to 1 if it is a real one, and close to 0 if it is a fake one. The parameters of the generator and discriminator are updated based on the result.

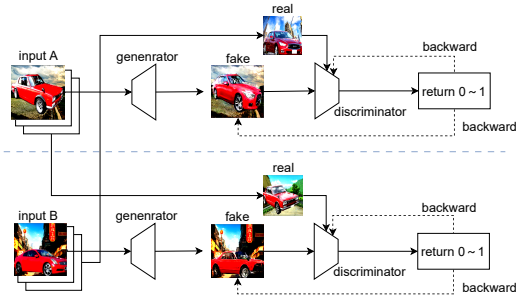

Figure 5: CycleGAN Learning Process.

First, consider the loss function that converts from domain A to domain B. The discriminator identifies whether the input image is a real image in the data set or a fake image generated by the generator. The generator also learns to generate images that fool the discriminator. These two types of learning are represented by the adversarial loss shown in (1). G is the generator that converts A to B, Dy is the discriminator that distinguishes real B from fake B, and X is the data set of domain A, Y is the data set of domain B. Conversely, a similar loss function can be defined for domain A from domain B.

$$L_{gan}(G,D_Y,X,Y) = E_y[\log D_Y(y)] + E_x[\log(1-D_Y(G(x))] \quad (1)$$

In addition to the adversarial loss, CycleGAN introduces the cycle consistency loss shown in (2).

$$L_{cyc}(G,F) = E_x[||F(G(x))-x||_1] + E_y[||G(F(y))-y||_1] \quad (2)$$

The loss function of CycleGAN is expressed as a weighted addition of the adversarial loss and the cycle consistency loss.

$$L(G,F,D_X,D_Y) = L_{GAN}(G,D_Y,X,Y) + L_{GAN}(F,D_X,Y,X) \quad (3)$$

3 RESULTS

CycleGAN was run on a dataset of 2,700 images of old and new cars with the backgrounds replaced. Figures 6 and 7 show the image transformations for the white background. Figures 8 and 9 show the image transformations of the experimental results with 200 epochs. Figure 10 shows the output results of a model trained on a dataset of only one type of background with 200 epochs. In Figures 6, 7, 8, 9 and 10, left is the input and right is the output.

Figure 6: Conventional method(old to new).

Figure 7: Conventional method(new to old).

Figure 8: Proposed method(old to new).

Figure 9: Proposed method(new to old).

Figure 10: One type of background dataset.

The image transformation with background of the proposed method is comparable to the conventional image transformation without background.The output image on the right in Figure 9 shows that the lower right background area is not complemented. This is the issue that the background must be complemented when the shape changes.Figure 8 shows that when only one type of background was trained, the shape transformation was able to preserve the background.

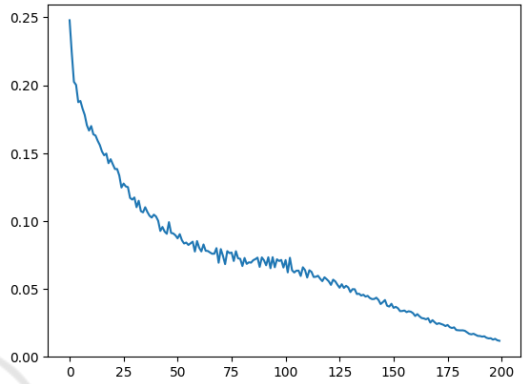

Figure 11: Discriminator loss.

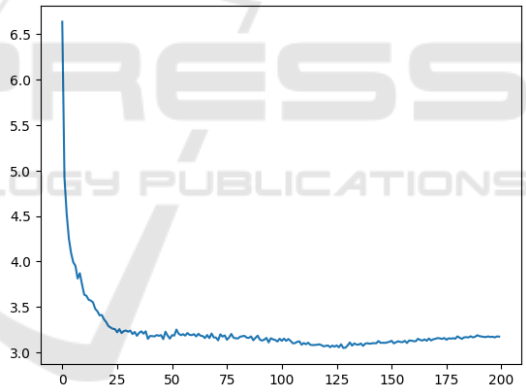

Figure 12: Generator loss.

Discriminator losses dropped steadily. Generator losses dropped to around epoch 125 but have been rising since then.In GAN, the training of the discriminator is followed by the training of the generator, and the learning balance is important to improve the accuracy of the generated images. In this case, the discriminator has overwhelmed the generator, so the performance of the discriminator needs to be reduced.

4 DISCUSSIONS

The accuracy of the resulting images was comparable between the model trained on the dataset of one background and the model trained on the dataset of

ten backgrounds from Figures 8 and 10. This is because the models generate images by extracting rules common to the datasets at the convolution layer and mimicking them. In one type of case, information about the car is lost in order to accurately represent the background. In the case of ten types, the car features were successfully extracted without being recognized as important features by the model because the background is random.The issue of background completion pointed out in section 4 is solved by transforming the network structure of the discriminator. As shown in Figure 13, the discriminator has two sizes of filters, large and small. The large filter captures global features and the small filter captures local features.The small filter completes detailed background areas, and the large filter captures the car outline. We also consider increasing the number of filters in the convolutional layer to improve the accuracy of the generator.As shown in Figure 14, the generator extracts features with two encoders and connects all the two feature maps together.

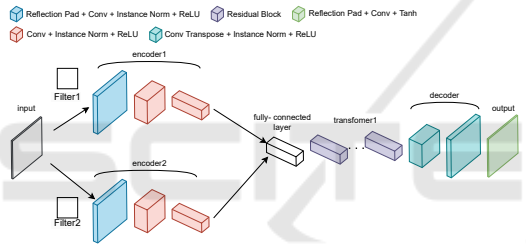

Figure 13: New Generator.

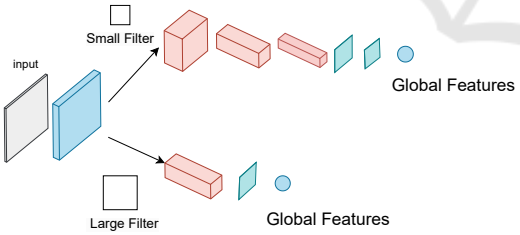

Figure 14: New Discriminator.

5 CONCLUSIONS

In this study, we transformed the shape of a car using CycleGAN while preserving the image background; since it is difficult to simply read the accuracy of the generated image from the loss function of GAN, we will not evaluate the results numerically. In the future, we will compare the results after and before the model change for visual evaluation.

REFERENCES

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., and Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, 27.

Isola, P., Zhu, J.-Y., Zhou, T., and Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Wu, W., Cao, K., Li, C., Qian, C., and Loy, C. C. (2019). Transgaga: Geometry-aware unsupervised image-to-image translation. In *CVPR*.

Zhu, C. and Li, G. (2017). A three-pathway psychobiological framework of salient object detection using stereoscopic technology. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV) Workshops*.