Improved Unsupervised Style Transfer - Mask Cycle-GAN

Jianlin Du, Zehao Guan, Yi Zhou, Wendi Cui

Language Technologies Institute, School of Computer Science, Carnegie Mellon University

Introduction

Cycle-GAN has achieved great results in unsupervised style transfer. However, it is sometimes not precise. For example, in figure 1, when the original Cycle-GAN transfers a horse to zebra, it applies stripes to not only the horse body, but also the human riding the horse. Besides, the background of the image is also patternized with stripes and rendered a lower saturation than the input.

Figure 1: Failure of Cycle-GAN

To remedy this problem, we propose the **Mask Cycle-GAN** which consists of a segmentation network and a Cycle-GAN. Our model achieves a satisfying performance by rendering precise images with patterns applied only on the target object. Meanwhile, the background becomes less noisy and more realistic now.

Related Work

- **Cycle-GAN**

Cycle-GAN adds a second generator which converts a generated image back. There is a constraint as it needs to keep the difference between input and converted back image down.

Cycle-GAN fails to identify objects within output images. For example, when trying to convert an image of horse to an image of zebra, the saturation of the output background will be reduced because of the black and white stripes on zebras. Our project plans to reduce this side effect so that only the object that needs to be transferred with the style will be rendered.

- **Mask R-CNN**

They present a conceptually simple, flexible and general framework for object instance segmentation. The applied approach efficiently detects objects in image while simultaneously generating a high-quality segmentation mask for each instance.

Methods

- **Pipeline**

There are two important components in the pipeline as the figure 2. One is the **Mask R-CNN** for generating mask of the target object; The other is a **Cycle-GAN** which takes images with four channels (RGB+mask).

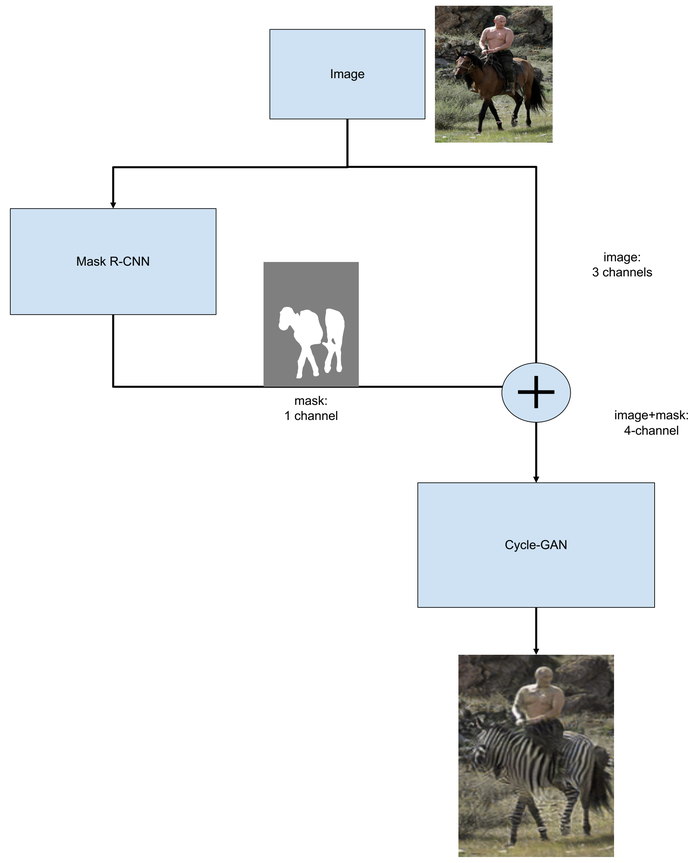

Figure 2: Pipeline

- **Mask R-CNN fine tuning**

We started with a pre-trained Mask R-CNN on COCO dataset (which includes horse and zebra classes). In order to further improve the ability of the network to produce masks, we fine-tune the pre-trained model to only recognize people, horses, and zebras and ignore other classes.

Then, we preprocessed the dataset of horses and zebras provided in the Cycle-GAN repository with the fine-tuned Mask R-CNN to generate the masks.

- **Cycle-GAN training**

Besides RGB, we added the forth channel (mask) to the original model which takes input from masks produced from Mask R-CNN. To ensure that they are on the same scale as the other three channels, the pixel value of masked area is set to 255 and that of the background is set to 127.

Then, we trained Cycle-GAN on the images with masks for 200 epochs, after which we feed the test picture (Putin riding the horse) and successfully transferred the horse in the picture to zebra as shown in the Figure 4.

Results

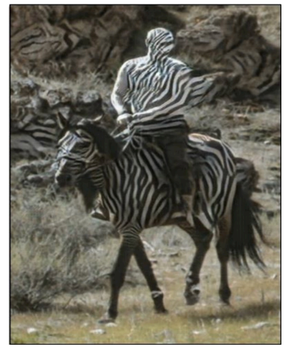

Figure 3: Cycle-GAN

Figure 4: Mask Cycle-GAN

Datasets

- Mask R-CNN training dataset

	Train	Test
horse	1500	200
zebra	1500	200
person	1500	200

- Mask Cycle-GAN training dataset

	Train	Test
horse	1068	141
zebra	1335	141

Conclusion

Our model successfully solve original Cycle-GAN's problem of sometimes misunderstanding what is the target object and what is background, and proved that generator of Cycle-GAN can successfully perceive the meaning of masks.

A limitation of our approach is that although the mask is good most of the time, Mask R-CNN may make mistakes which may affect the following result of Mask-Cycle-GAN. However, we believe there will be better segmentation techniques to produce more accurate mask in the future.

One of the possible future improvements is to use a smooth/fuzzy mask, instead of setting pixel values to either 255 or 127.

References

[1] Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2223-2232).

[2] He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In Proceedings of the IEEE international conference on computer vision (pp. 2961-2969).