11-785 Project Midterm Report: Mask Cycle-GAN

Jianlin Du, Zehao Guan, Yi Zhou, Wendi Cui

Github link: https://github.com/Yi-Zhou/dl_proj

1. Introduction

Cycle-GAN has achieved great results in style transfer. However, it is sometimes not precise enough. For example, as figure 1 demonstrates, if we want to transfer a horse to zebra, it applies the stripes to not only the horse body, but also the background, even the human riding the horse. Besides, the background of the image is also patternized with the strip and rendering a lower saturation than original one. We suppose Cycle-GAN could render better result if we provide it with a mask of the object that we want to apply style transfer on. It can be applied to resolve the ambiguity in the image for Cycle-GAN to generate more precise transformation. Our team plans to combine the two neural networks, by feeding in the segmentation of the input image (from Mask R-CNN), in addition to the input image, into Cycle-GAN to achieve better accuracy.

Figure 1: Failure of Cycle-GAN

2. Baseline

In the pipeline, there are two important components as illustrated in figure 2: one is Mask R-CNN (or other segmentation networks) for generating mask for the object we want to apply the style-transfer on; The other is Cycle-GAN, that takes both the image and the mask as inputs.

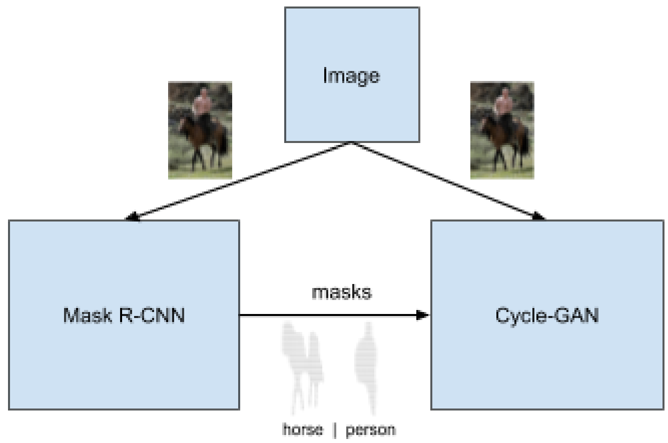

Figure 2: Model Structure

We are starting with a Mask R-CNN for style-transfer between horse and zebra, so the Mask R-CNN pre-trained on COCO (which includes both horse and zebra) dataset can well handle the task of generating mask. Here is the example output of the Mask R-CNN.

Figure 3: Mask R-CNN Output Mask

We can see that the horse and rider are generally well separated, and Cycle-GAN will have a much better idea on what to style transfer after getting the mask.

We have also modified the Cycle-GAN to let it take in 4 channels(R, G, B, mask) instead of the original 3 channels. Since it is still being trained right now, unfortunately, we cannot show the result for Cycle-GAN now.

3. Future Work

For the future phases, we can figure out a way to incorporate instance-level segmentation to Cycle-GAN. In the situation where multiple horses are presented in the image, we can take advantage of the output of Mask R-CNN where individuals of the same classes are separately masked to only transfer style on specific ones. Another thing we notice is that there are overlapping between masks of horses and human. Thus when we use those generated masks directly to apply patterns, there will be discrepancies. We are thinking of getting rid of the overlapping region by analyzing the overlapped parties to get a cleaner cutoff.

4. Related Work

1) GAN

A generative model that is trained via an adversarial process, in which we simultaneously train two models: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G.

2) Cycle-GAN

Contrary to GAN with only one generator from input to generated image, Cycle-GAN adds a second generator which converts a generated image back. So there is a constraint as it needs to keep the difference between input and converted back image down.

Limitation: 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.

3) 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.

5. Reference

[1] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672-2680).

[2] 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).

[3] 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).

[4] Einstein, A., B. Podolsky, and N. Rosen, 1935, "Can quantum-mechanical description of physical reality be considered complete?", Phys. Rev. 47, 777-780.

[5] Gatys, L. A., Bethge, M., Hertzmann, A., & Shechtman, E. (2016). Preserving color in neural artistic style transfer. *arXiv preprint arXiv:1606.05897*.

[6] Wang, H., Liang, X., Zhang, H., Yeung, D. Y., & Xing, E. P. (2017). Zm-net: Real-time zero-shot image manipulation network. *arXiv preprint arXiv:1703.07255*.

[7] Jing, Y., Liu, Y., Yang, Y., Feng, Z., Yu, Y., Tao, D., & Song, M. (2018). Stroke controllable fast style transfer with adaptive receptive fields. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 238-254).

[8] Jing, Y., Yang, Y., Feng, Z., Ye, J., Yu, Y., & Song, M. (2017). Neural style transfer: A review. *arXiv preprint arXiv:1705.04058*.

[9] Gatys, L. A., Ecker, A. S., & Bethge, M. (2015). A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576*.