Cartoon-to-real: An Approach to Translate Cartoon to Realistic Images using GAN

K M Arefeen Sultan[*], Labiba Kanij Rupty[†], Nahidul Islam Pranto[‡],
Sayed Khan Shuvo[§], Mohammad Imrul Jubair[¶]
Department of Computer Science and Engineering,
Ahsanullah University of Science and Technology Dhaka, Bangladesh
{[*]*krsultan069*, [†]*labknr98*, [‡]*nahidul19967*, [§]*sayedhossainkhan36*}*@gmail.com*, [¶]*mohammadimrul.jubair@ucalgary.ca*

Abstract—**We propose a method to translate cartoon images to real world images using Generative Aderserial Network (GAN). Existing GAN-based image-to-image translation methods which are trained on paired datasets are impractical as the data is difficult to accumulate. Therefore, in this paper we exploit the *Cycle-Consistent Adversarial Networks (Cycle-GAN)* method for images translation which needs an unpaired dataset. By applying CycleGAN we show that our model is able to generate meaningful real world images from cartoon images. However, we implement another state of the art technique – *Deep Analogy* – to compare the performance of our approach.**

1. Introduction

What if we could see real images of one of the most famous cartoon movies – Spirited Away (2001)? How would it feel to see the protagonist, chihiro's real life version? Isn't this most of the caroon-lovers have dreamt of while watching cartoon movies?

In this paper, we present a method – *Cartoon-to-Real* – to materialize the above desire by performing cartoon to real world image translation. However, it is extremely time consuming and tedious to create a sufficient paired dataset, hence we develop an unpaired one. We extracted cartoon images from different cartoon movies and real images from internet, i.e. *flickr*, where a cartoon and a realistic image in a pair have no correlations among themselves. Using our Cartoon-to-Real, we achieve significant result in translating the cartoon images to realistic ones.

2. Background and Related Works

Recently, Generative Adversarial Network (GANs) [1] have achieved astounding results in image synthesis such as – text-to-image translation [2], image inpainting [3], super-resolution [4] etc. Moreover, GAN is widely used in image-to-image translation, for example – CycleGAN [5] – that uses unpaired training data. It trains two sets of GAN to map *class R → class C* and *C → R* respectively. Recently, CartoonGAN [6] is proposed to translate real word images

to cartoon images which converges faster than CycleGAN [5] and performs satisfactorily (see Figure. 1).
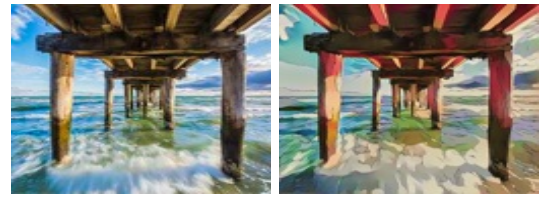
(a) Input image   (b) Output image

Figure 1: Results of CartoonGAN [6] approach. Here, a real world image (a) is translated a cartoon image (b).

3. Proposed Methodology

The main target of our *cartoon-to-real* is to perform the reverse of CartoonGAN (***Cartoon→Real***) and we exploit the CycleGAN [5] technique for this purpose. The model contains two mapping functions $G : C \to R$ and $F : R \to C$ where C denotes cartoon and R denotes the real domain. There are 2 discriminators (D_R, D_C) and 2 generators (G, F) for the translation process. While performing $C \to G(C)$, D_R tries to enforce the translation $G(C)$ to domain R, and the vice-versa for D_C and F. For the regularization, we implement two *cycle consistency losses* [5] where authors proposed that the learned mapping functions should be cycle-consistent to avoid direct mapping distribution. The loss is written as -

$$\mathcal{L}_{cyc}(G, F) = \mathbb{E}_{c \sim p_{data}(c)}[||F(G(c)) - c||_1] \\ + \mathbb{E}_{r \sim p_{data}(r)}[||G(F(r)) - r||_1] \quad (1)$$

Hence, the full objective is -

$$\mathcal{L}(G, F, D_C, D_R) = \mathcal{L}_{GAN}(G, D_R, C, R) \\ + \mathcal{L}_{GAN}(F, D_C, R, C) + \lambda \mathcal{L}_{cyc}(G, F). \quad (2)$$

where λ is the weight or relative importance of the two objectives. Therefore, our aim can be described as -

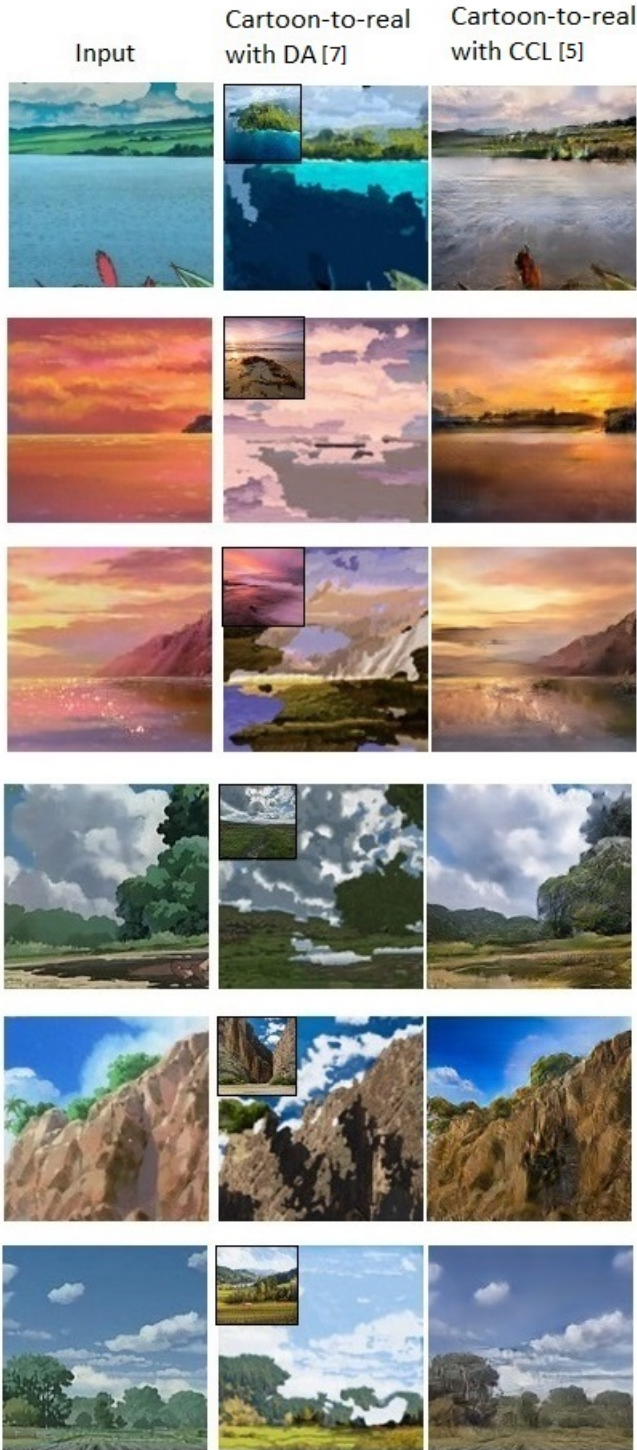$$G^*, F^* = \arg \min_{G,F} \max_{D_c,D_r} \mathcal{L}(G, F, D_C, D_R) \quad (3)$$

Figure 2: Comparison between cartoon to real translation using Deep Analogy (DA) and using cycle consistency loss (CCL). *Left column* shows the input, *middle column* shows the output of using Deep Analogy (with the style image on top corner), and the *right column* presents the results of using cycle consistency loss. It is visible that the cartoon-to-real with cycle consistency loss [5] shows the better translation.

4. Experiment Results

We developed two unpaired datasets to train our network. For the cartoon domain, we collected almost 3.1K images scrapped from various movies, e.g. *Pokemon*, *My Neighbour Totoro* and *Kiki's Delivery*. We used *flickr* dataset for the real images' domain. Images are resized to 128×128 resolution. For implementation we used PyTorch and for hardware we used `Nvidia GTX 1060`. We compared our paper with the outputs from another state of the art work called – *Deep Analogy* [7]. Our result along with the outputs from Deep Analogy are presented in Figure 2. It exhibits that outputs with cycle consistency loss produce more realistic images than that with Deep Analogy.

5. Conclusion

In this paper, we performed image translation from cartoons to real world images. We used cycle consistency loss to generate images so that the generate image will not be directly mapped into any distribution of target domain. Our research is yet on progress. However, we observed that our results are not completely satisfactory and our upcoming target is to minimize the limitation. In future, we want to investigate on preserving the content of input image from cartoon domain for better translation.

References

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

[2] S. E. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, "Generative adversarial text to image synthesis," *CoRR*, vol. abs/1605.05396, 2016. [Online]. Available: http://arxiv.org/abs/1605.05396

[3] R. A. Yeh, C. Chen, T. Lim, M. Hasegawa-Johnson, and M. N. Do, "Semantic image inpainting with perceptual and contextual losses," *CoRR*, vol. abs/1607.07539, 2016. [Online]. Available: http://arxiv.org/abs/1607.07539

[4] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," *CoRR*, vol. abs/1609.04802, 2016. [Online]. Available: http://arxiv.org/abs/1609.04802

[5] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *arXiv preprint*, 2017.

[6] Y. Chen, Y.-K. Lai, and Y.-J. Liu, "Cartoongan: Generative adversarial networks for photo cartoonization," 2018.

[7] J. Liao, Y. Yao, L. Yuan, G. Hua, and S. B. Kang, "Visual attribute transfer through deep image analogy," *arXiv preprint arXiv:1705.01088*, 2017.