Discriminative Region Proposal AdversarialNetworks for High-Quality Image-to-Image Translation

Wenjie Niu

July 19,2018

After several days of reading the paper of Discriminative Region Proposal Adversarial Networks for High-Quality Image-to-Image Translation, I have access to the framework, structure and basic professional knowledge such as Convolutional Neural Network(CNN) and Generative Adversarial Network(GAN). However, there are some difficulties for me like the deprivation of formula in the understanding of GAN.

Abstract

In this paper, authors present Discrimitive Rigion Proposal Adversarial Network(DRPAN) for high-quality image-to-image translation. In order to reduce the artifacts and blur problems during the process of translation, they explore the patch discriminator. It produces the score map and find the most artificial part by a sliding window. Then put the discriminitive region mask the coresponding real image to obtain a fake-mask real image. A reviser based on GANs is used to distinguish the real to fake-mask real to help improving the translation quality.

1. Introducion

The procedure of image-to-image translation task is descomposed into three iterated steps. First is to generate a fake image with some local artifacts by generator. Second is to propose the most fake region from the genrated image using DRPnet as shown in figure. 1. Third is to mask real image with the discriminative region, so that DPRAN(as shown in figure. 2) can optimazed on the most fake part.

There are three main contribution in this paper:

- discriminators based on patch for producing discriminative region;
- reviser based on GANs to provide constructive revisions for generator
- building a DRPAN model

In detail, semantic segmentation graph is put in the generator, the fake image is the output. Then the discriminator is to find and extract the discriminative region in the fake image with sliding window. The corresponding real image mask with the discriminative region producing a fake-mask real sample. Last, the sample is put in the reviser to judge the image whether fake-mask real or real image. The whole architecure is shown in figure. 2.

2. Method

Figure. 3 shows the process of how to improve the quality of synthesized image. With DRPAN's training, the discriminative of fake-mask real image varies improving the quality of synthesized image and scoring higher.

Figure. 4 shows the output results of score map on different quality levels(fake and real) of images by a pre-trained PatchGAN. The score maps of the fake samples are almost dark with lower score while of the real samples are brightest with highest scores.

Given an input image with resolution $w_i \times w_i$, and processed by the patch discriminator to be a probability score map with size $w_s \times w_s$. Suppose we want to obtain the discriminative region at $w^* \times w^*$, the size of sliding window w for score map can be calculated by

$$w = w^* \times w_s / w_i \tag{1}$$

Then our DRPnet will find the discriminative square patch on score map with the center coordinates (x_c, y_c) and length w, so the scaleτ between the input image and output score map is

$$\tau = \frac{w_i - w^*}{w_s - w} \tag{2}$$

The center coordinates (x_c^*, y_c^*) of discriminative region will be calculated by

$$\begin{cases} x_c^* = \tau \times x_c, \\ y_c^* = \tau \times y_c. \end{cases} \tag{3}$$

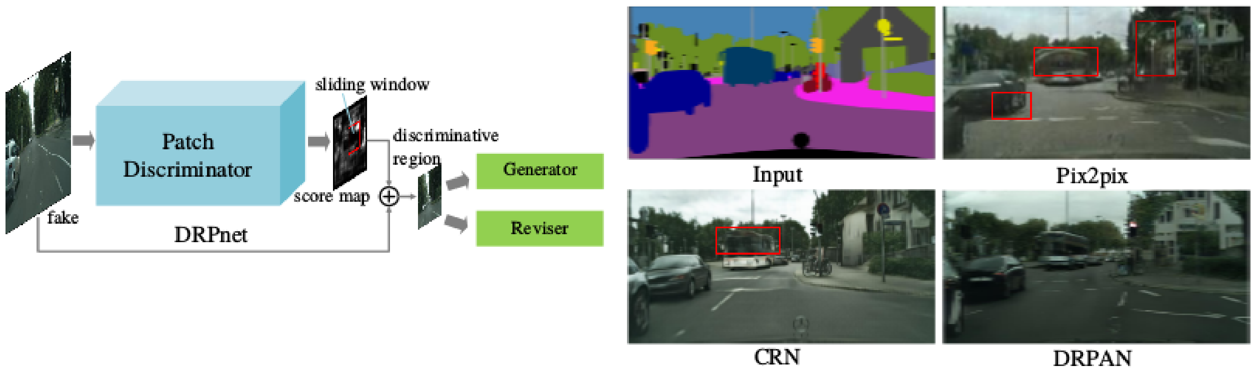

Figure 1. **Left**: Their Discriminative Region Proposal network (DRPnet). **Right**: Synthesized samples compared with previous works on Cityscapes validation dataset [1]. The regions within red window show obvious artifacts or deformation. Their method can synthesize images with clear structure and vivid details. [4]

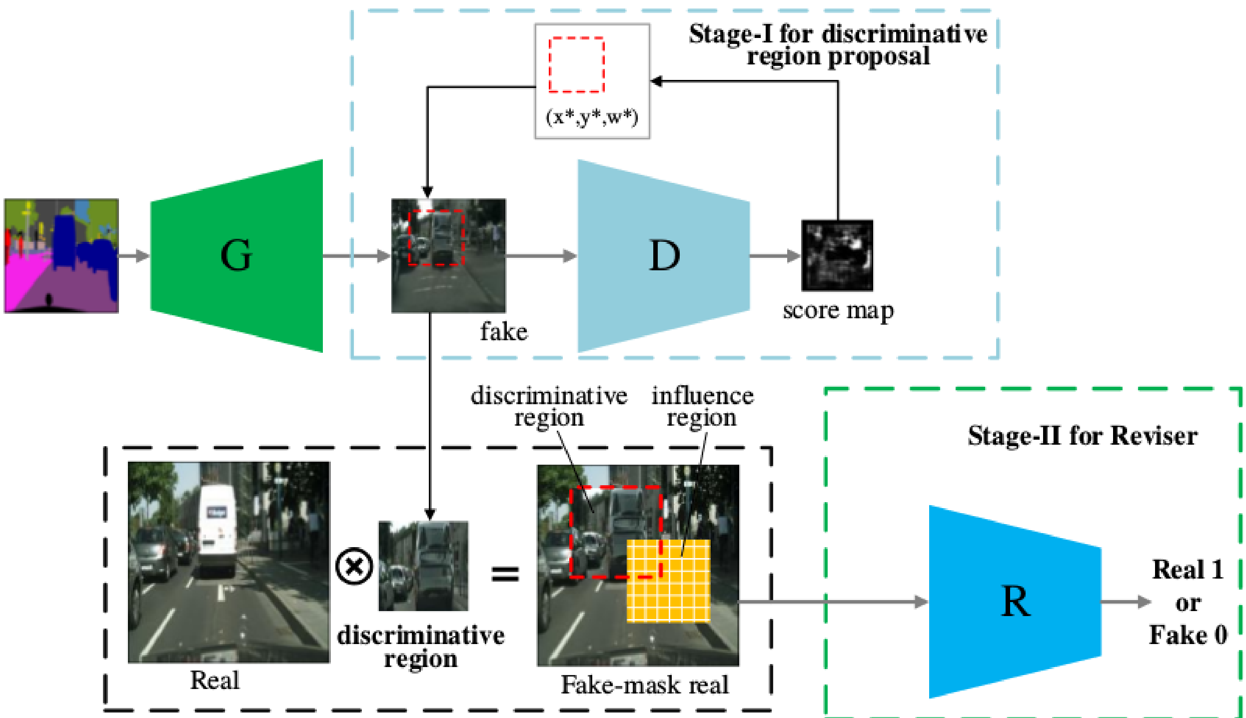

Figure 2. The overall network architecture and data flow of our proposed Discriminative Region Proposal Adversarial Network (DRPAN), which is composed of three components: a generator, a discriminator, and a reviser, and is a unified model for image-to-image translation tasks. [4]

Finally, the discriminative region d r produced by DRPnet can be expressed as

$$d_r = F_{DRPnet}(x_c^*, y_c^*, w^*) \tag{4}$$

3. Conlusion

Thanks to the carefully reading of this paper, I learnt a lot about the procedure of GAN training, the working mechanism and concrete understanding of this paper. There are quite some difficulty in the understanding of the equations,

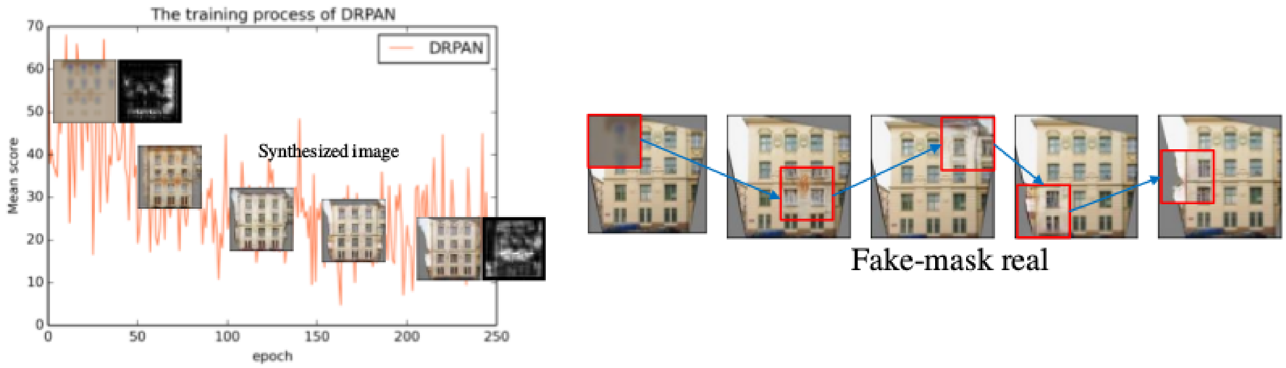

Figure 3. The training process of DRPAN on facades dataset [3]. Left: The plotting curve shows mean value of score map on synthesized samples. Right: Step by step synthesis on different discriminative regions. [4]

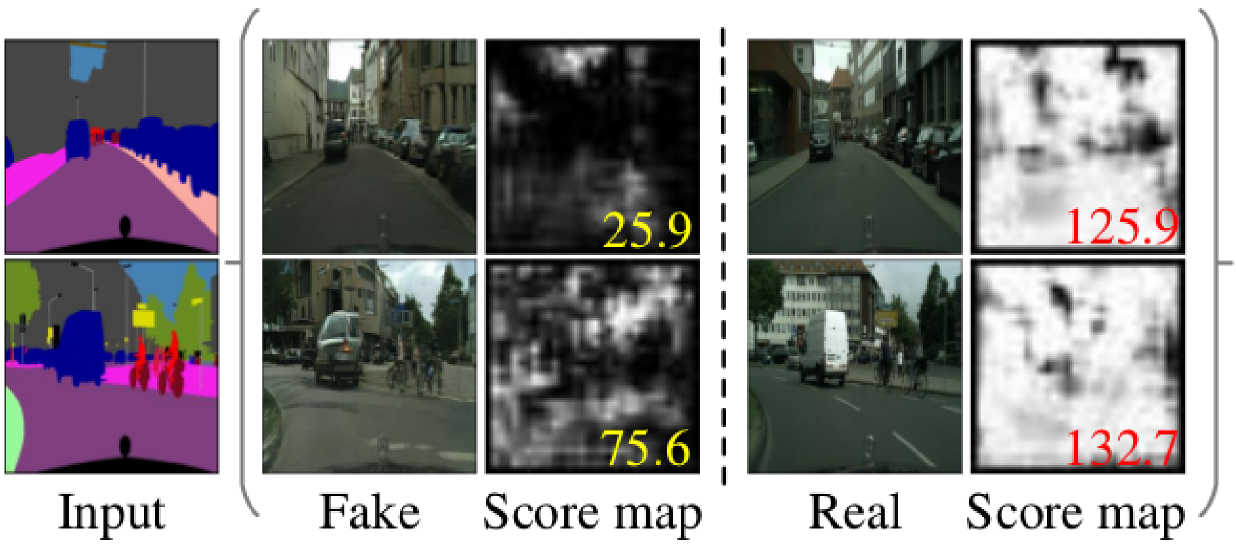

Figure 4. The output results of score map on different quality levels (fake and real) of images by a pre-trained PatchGAN. The darkest regions on score maps mean the lowest quality, indicating that patch-based discriminators can be explored for discriminative region proposal. [4]

I still need some time to catch.

References

[1] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The cityscapes dataset for semantic urban scene understanding. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 2

[2] N. Kodali, J. Abernethy, J. Hays, and Z. Kira. How to train your dragan. *arXiv preprint arXiv:1705.07215*, 2017.

[3] R. Tyleek and R. ra. Spatial pattern templates for recognition of objects with regular structure. In *Global Conference on Psychology Researches*, 2013. 3

[4] C. Wang, H. Zheng, Z. Yu, Z. Zheng, Z. Gu, and B. Zheng. Discriminative region proposal adversarial networks for high-quality image-to-image translation. *arXiv preprint arXiv:1711.09554*, 2017. 2, 3