

# Discriminative Region Proposal Adversarial Networks for High-Quality Image-to-Image Translation

Yufeng Jiang

## Abstract

*Generative Adversarial Networks (GANs) is one of the most promising methods for unsupervised learning in complex distribution in recent years. It has made some technologies like image-to-image develop quickly. However, it still has a great challenge at generating images with high-quality, especially at high-resolution and photo-reality. In this paper, authors present a novel Discriminative Region Proposal Adversarial Network (DRPANs) for high-quality image-to-image translation. Authors explore the patch discriminator to find the most artificial image patch by sliding windows with the output score map to solve the problem of artifacts and blur problems based on GAN. Meanwhile, authors also map the image patch onto the fake image synthesized by generator as discriminative region. Then, authors mask the real image corresponding to the fake image with the discriminative region to obtain the fake-mask image. The purpose of reviser is that it distinguishes the real from the fake-mask real for producing realistic details and improves to generate high-quality image.*

## 1. Introduction

Authors consider that people think the synthesized image is fake because it contains local artificals, but they still can recognize it by only gazing about 1000ms [1]. People can get a realist scene from the coarse structure and fine details, and understand the corresponding of things around it. Motivated by this ability, the goal of this work is to develop an image-to-image translation system for high-quality image synthesis with clear structure and realistic details.

People has make many efforts to develop an automatic image-to-image translation system. The straightforward approach was to optimize on pixel-wise space with L1 or L2 loss [3, 4]. However, both of them suffer from blur problem. From this consideration, authors decided to use GAN to do image-to-image translation. Although using GAN still encounters with the artifacts and unsmooth color distribution problems, and it is even hard to generate high-resolution photo-realistic images because of the high dimension distribution [5].

In this paper, authors decompose the procedure of image-to-image translation task into three steps. (i) They generate a image with a global structure but some local artifacts through GAN. (ii) They extract the most fake region from the generated image with the DRPnet shown at Fig. 1. (iii) they implement image inpainting on the most fake region for more realistic result to obtain the most realistic image.

From these reasons mentioned before, authors explore a framework based on patch-wise discriminator to predict the discriminative score map and use sliding windows to find the most artificial region. Then, the proposed discriminative region is used to make the corresponding real sample and output as “fake-mask real”. Finally, with the goal of generating realistic details, they propose a reviser to distinguish the real image to the “fake-mask real”.

## 2. Method

Their image-to-image translation model called Discriminative Region Proposal Adversarial Networks (DRPANs) is composed of three components: a generator, a discriminator and a reviser. The discriminator explores PatchGAN to construct Discriminative Region Proposal network (DRP-net, see Fig. 1) to extract the discriminative region for producing fake-mask real sample. The overall network architecture and data flow are illustrated in Fig. 2.

Their process of improving the quality of synthesized image. With the training of their DRPAN, the discriminative region of fake-mask real image (right) varies. And the quality of synthesized image (left) has been improved and obtains the brighter score map (the first and the last). Although it is still difficult to distinguish the real from the synthesized image after training many times, DRPAN proposed by them can obtain the high-quality image through optimized by reviser at some details.

## References

- [1] Q. Chen and V. Koltun. Photographic image synthesis with cascaded refinement networks. In *ICCV*, 2017. 1
- [2] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and Schiele. The cityscapes

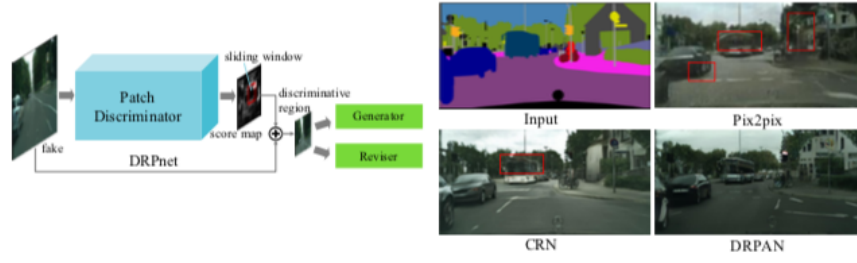


Figure 1. **Left:** Our Discriminative Region Proposal network (DRPnet). **Right:** Synthesized samples compared with previous works on Cityscapes validation ataset [2]. The regions within red window show obvious artifacts or deformation. Our method can synthesize images with clear structure and vivid details.

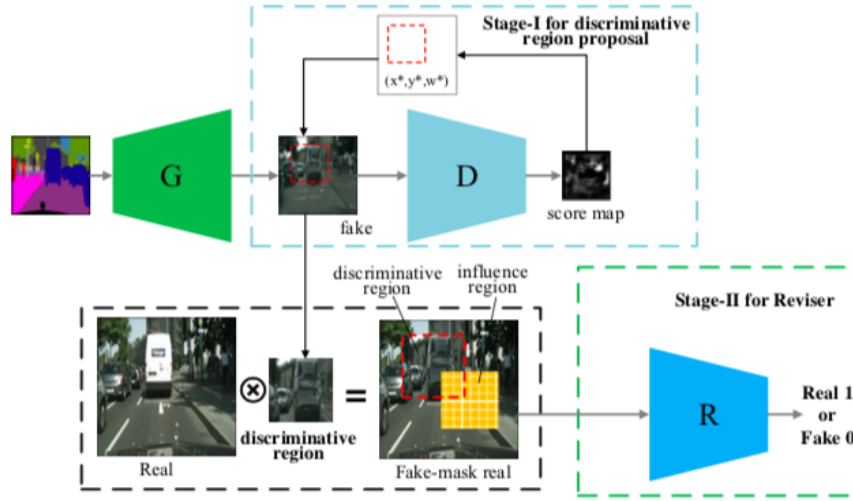


Figure 2. The overall network architecture and data flow of our proposed Discriminative Region Proposal Adversarial Network (DRPAN), which is composed of three components: a generator, a discriminator, and a reviser, and is a unified model for image-to-image translation tasks.

- dataset for semantic urban scene understanding. In *CVPR*, 2016. 2
- [3] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE TPAMI*, 38(2):295–307, 2016. 1
- [4] J. Long, E. Shelhamer, and T. Darrel. Fully convolutional networks for semantic segmentation. In *CVPR*, 2015. 1
- [5] A. Odena, C. Olah, and J. Shlens. Conditional image synthesis with auxiliary classifier gans. In *ICLR*, 2016. 1