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

Wenjie Niu

August 3, 2018

This writing is also about the paper of *Discriminative Region Proposal Adversarial Networks for High-Quality Image-to-Image Translation*. And this part mainly analyse the experiment, which I spent more than two days to understand what's the meaning. Therefore, the experiment part is worth considering.

1. Experiment

To evaluate the performance of proposed method on image-to-image translation tasks, they deploy a variety of experiments about different levels of translation tasks to compare our method with state-of-the-arts. And for different tasks, they also use different evaluation metrics including human perceptual studies and automatic quantitative measures.

1.1. Evaluation metrics

Image quality evaluation. PSNR, SSIM [7] and VIF [5] are some of the most popular evaluation metrics in low-level computer vision tasks such as deblurring, dehazing and image restoration. So for de-raining and aerial to maps tasks, thhey adopt PSNR, SSIM, VIF and RECO [1] to qualify the performance of results.

Image segmentation evaluation metrics. We use standard metrics from Cityscapes benchmark [2] to evaluate real to semantic labels task on Cityscapes dataset, including per-pixel accuracy, per-class accuracy, and Class IOU.

Amazon Mechanical Turk (AMT). AMT [3, 10, 8] is adopted in many tasks as a gold metric to evaluate how real the synthesized images, and we use it as evaluation metric for semantic labels to photo and maps to aerial tasks.

FCN-8s score. The intuition of using an off-the-shelf classifiers for automatic quantitative measurement is that if the generated images are realistic, classifiers trained on real images will be able to classify the synthesized image correctly as well [3]. We use the FCN-8s score [4] to evaluate semantic labels to real task on Cityscapes dataset. The FCN-8s model trained on Cityscapes segmentation tasks is taken from [3].

1.2. Why DRPAN?

The PatchD is efficient to discover the most fake or real region (Fig 1) from the image but is limited to improve these regions with fine details for that PatchD is hard to capture the high dimension distribution. In this case, It's be proposed a DRPnet (explore the strength of PatchD) for discriminative region proposal and design a reviser to gradually remove visual artifacts, and thus reduce it to lower dimension estimation problem. This can be seen as a top-down procedure which is different from other gradually bottom-up image generation method [9]. Fig 2 shows the necessity of our proposed DRPAN for high-quality image-to-image translation, which illustrates that continue training PatchD is no help to reduce artifacts even with a L1 loss for balance, and DRPAN with only L1 loss can smooth the artifacts but not very sharp in details, while DRPAN with reviser exceeds the PatchDs performance with less visual artifacts. The combination of reviser and L1 loss can reduce these artifacts ignored by PatchD. It's be found that fake-mask operation can improve the fluency of whole image in certain samples (*e.g.*, the connection between door and wall). So DRPAN with fake-mask is implemented in the following experiments.

2. Conlusion

Thanks to the carefully reading of this paper, I learnt a lot about the procedure of GAN training, I understand the method of the paper,which is very strange for me before. The entire procedure is complex and it worth reading and considering again and again.

References

[1] V. Baroncini, L. Capodiferro, E. D. D. Claudio, and G. Jacovitti. The polar edge coherence: A quasi blind metric for video quality assessment. In *ESPC*, 2015. 1

[2] 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 *CVPR*, 2016. 1

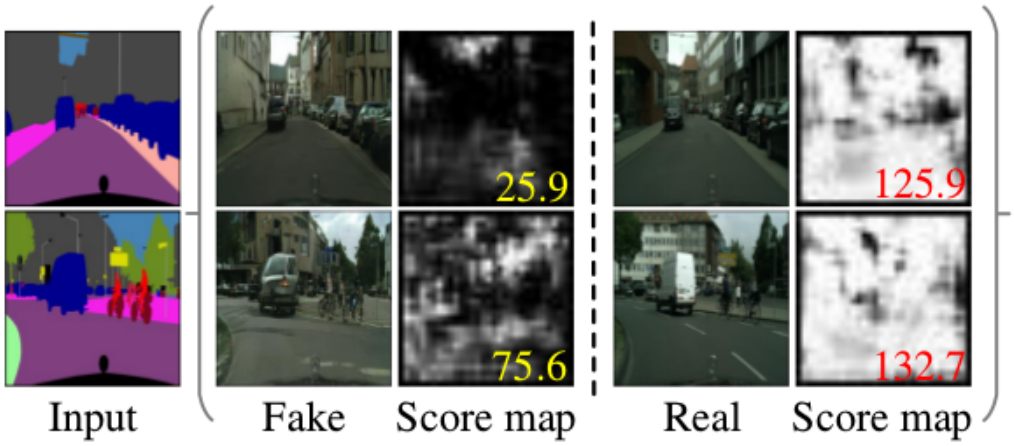

Figure 1. The output results of score map on different quality levels (fake and real) of images by a pre-trained PatchGAN. The darkest regions on score maps mean the lowest quality, indicating that patch-based discriminators can be explored for discriminative region proposal. [6].

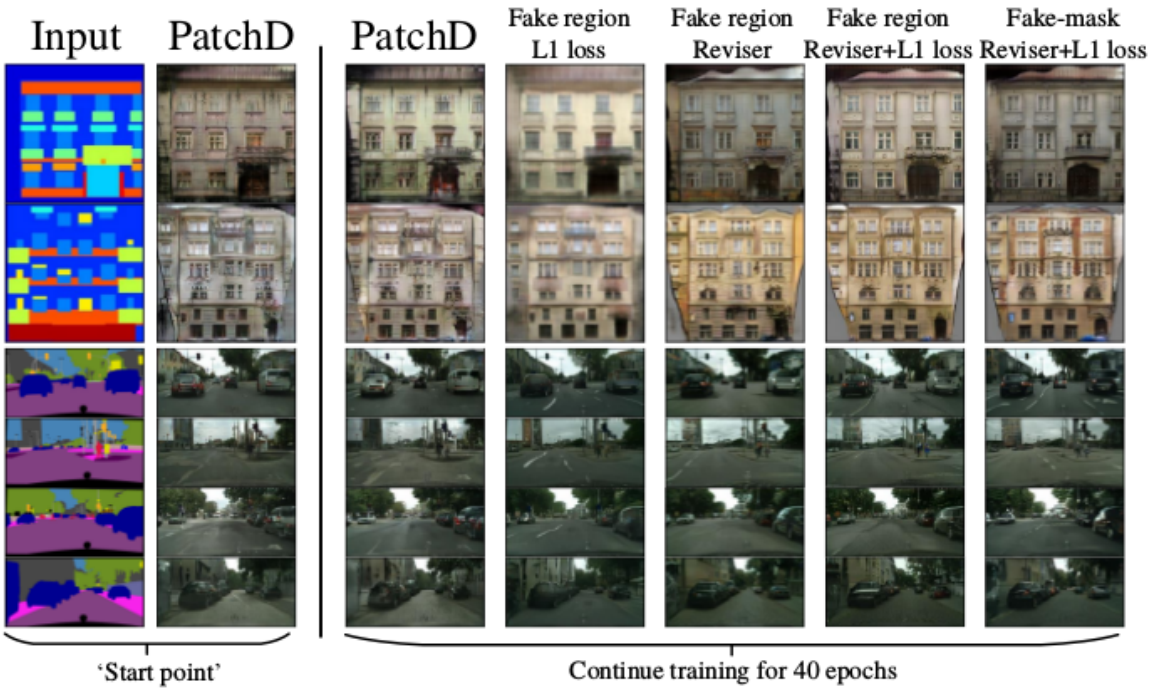

Figure 2. Different methods with various losses produce different quality of results. The second column is the start point of comparison trained by PatchD, and all other models are continued trained for 40 epochs more. These experiments validate the necessary of our DRPnet for discriminative region proposal, our reviser for optimizing generator, and our fake-mask operation for improving synthesis. [6].

[3] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. In *CVPR*, 2016. 1

[4] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, 2015. 1

[5] H. R. Sheikh and A. C. Bovik. Image information and visual quality. *IEEE TIP*, 15(2):430–444, 2006. 1

[6] 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

[7] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE TIP*, 13(4):600–612, 2004. 1

[8] Z. Yi, H. Zhang, P. Tan, and M. Gong. Dualgan: Unsupervised dual learning for image-to-image translation. In *ICCV*, 2017. 1

[9] H. Zhang, T. Xu, and H. Li. StackGAN: Text to photo-realistic image synthesis with stacked generative adversarial networks. In *ICCV*, 2017. 1

[10] J. Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, 2017. 1