Image to Image translation Experiment

Xinhao Chen

chenxinhao@sjtu.edu.cn

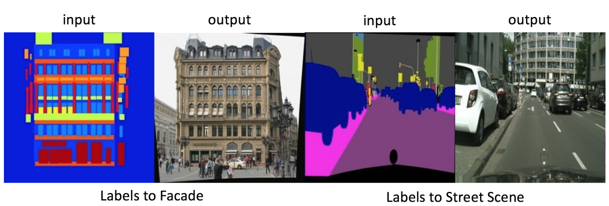
Shanghai Jiao Tong University/ACM Hornored Class

Abstract

Our project is an image-to-image translation problem, and we use Generative Adversarial Networks(GAN) as our solution. Based on the previous work "Image-to-Image Translation with Conditional Adversarial Networks" [1] and "Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks" [3], we made some interesting attempts, and achieved a "not bad" result.

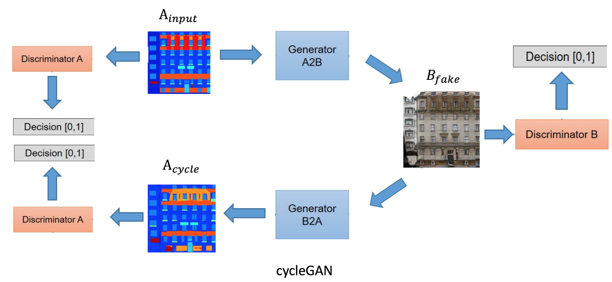
1. Introduction

Specifically, our task is "translating" a semantic image to a real image. And we using "Labels to Street Scene" and "Labels to Facade" from paper "Image-to-Image Translation with Conditional Adversarial Networks" as our experiment dataset.



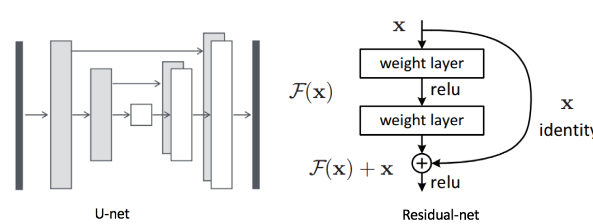
2. Method

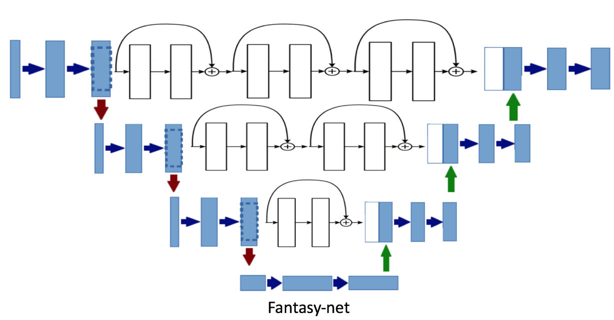
As mentioned before, we using conditional GAN and cycle GAN as our basic model. Different from the normal GAN, the inputs of conditional GAN are not random noise, but the conditional settings, for our task they are the paired image. And base on the conditional GAN, cycle GAN made an innovation on the original conditional GAN, and it's a bi-direction GAN, the conditional GAN only can generate A to B, and measure the different between B_{real} and B_{fake}, but the cycle GAN can generate in both direction, A to B and B to A, so it can achieve a cycle generate, translate A_{input} to B_{fake}, then translate B_{fake} to A_{cycle}, just like two conditional GANs which using cycle consistency to connect each others, so it can not only measure the different between B_{real} and B_{fake}, but also A_{input} and A_{cycle}, which may help us reach a better result in our task.



2.1. Generator

And like all GANs, we need to define our Generator(G) and Discriminator(D), and from the origin implement, we know that they use U-net as the generator of conditional GAN, and residual-net as the generator of cycle GAN, they all use the idea of skip connection, which is very effective in large network training.

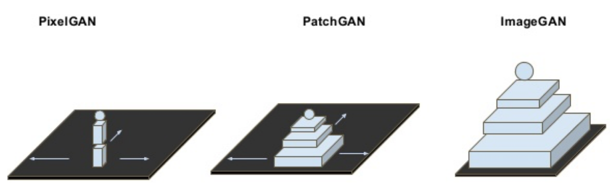




Inspired by this two network, We proposed a new kind skip connection network which can combine the characteristics of these two networks, and its implement is very simple, just change the skip connection of U-net into the multi-level residual-net, this change will not eliminate any part of origin networks, but just provide more choices, intuitively, it can improve the robustness of our network. we call it “Fantasy-net”

2.2. Discriminator

We use the N layer fully convolutional patch GAN in origin paper as our discriminator. The difference between a PatchGAN and regular GAN discriminator is that rather the regular GAN maps from a 256×256 image to a single scalar output, which signifies ”real” or ”fake”, whereas the PatchGAN maps from 256×256 to an $N \times N$ array of outputs X, where each X_{ij} signifies whether the patch ij in the image is real or fake, so it can keep our attention to the structure in local image patches, and more effective for us to determine it is real or fake.

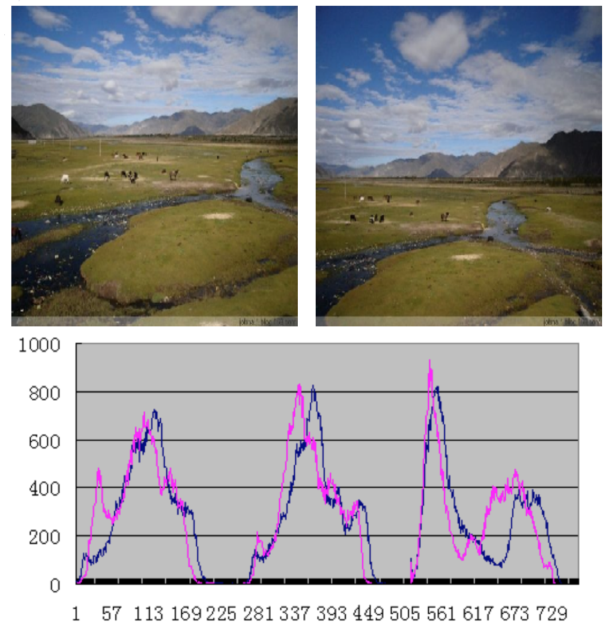


3. Evaluation

Evaluating the quality of the synthesized images is an open and difficult problem. There is no clearly definition of so called style similarity, so is very difficult to accurately evaluate our results. And, finally, we choose the FCN score [2] and similarity of image histogram as our evaluation.

The fully convolutional network(FCN), is a pre-trained model to do the semantic segmentation. And we use that to measure the discriminability of the generated images.The intuition is that if the generated images are realistic, classifiers trained on real images will be able to classify the synthesized image correctly as well, we adopt the popular FCN-8s architecture for semantic segmentation.

And the image histogram is a type of histogram that acts as a graphical representation of the tonal distribution in a digital image. It plots the number of pixels for each tonal value. so that we can use the difference of the tonal distribution to measure the similarity of our fake image and real image.



4. Experiment and Result

In our experiments, we tried six models in two different datasets, detailed information is in following table:

Model	Generator	Discriminator	Dataset
Cycle GAN	U-net	Patch GAN	Houses
	Cycle-net		
Conditional GAN	Fantasy-net		Cityscapes

4.1. Result

Here is our final result, (FCNs) / (similarity of image histogram)

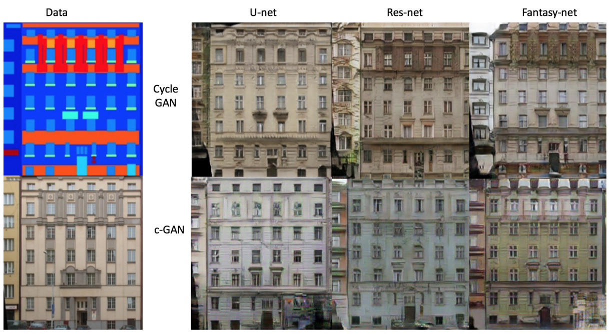
Houses Dataset	U-net	Res-net	Fantasy-net
Cycle GAN	18.56/34.33	17.98/33.56	19.22/42.55
Conditional GAN	20.24/48.83	21.05/49.02	20.87/48.16

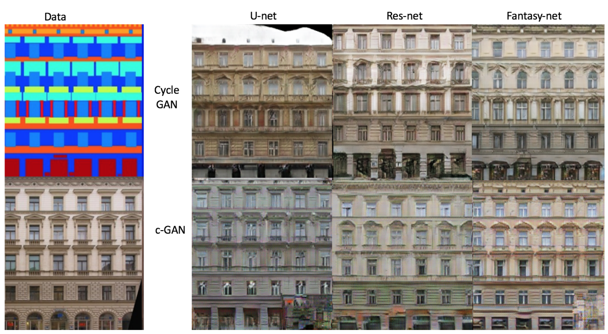
Cityscapes Dataset	U-net	Res-net	Fantasy-net
Cycle GAN	24.83/50.51	25.44/51.91	24.92/49.52
Conditional GAN	26.65/55.22	26.12/54.78	25.71/53.74

The FCN score is pretty low, it means our generator model is not real enough to recognized by the FCN segmentation. But the similarity of the histogram is not bad.

This means that the fake image generated is similar to the real one in the color space.

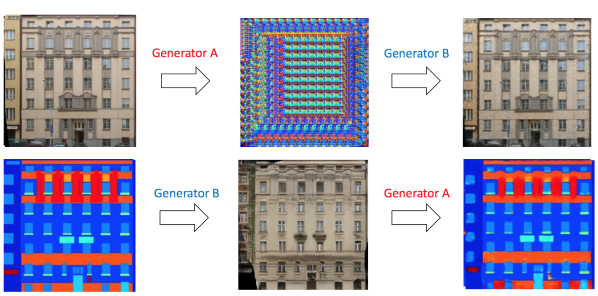
Although the final result is different in evaluation values, it is still very hard for the human eye to distinguish which picture is better. I can only say that different models have learned different styles, and all of the generated images are partially obscured, and a little messy, but overall, it is still a very good result.





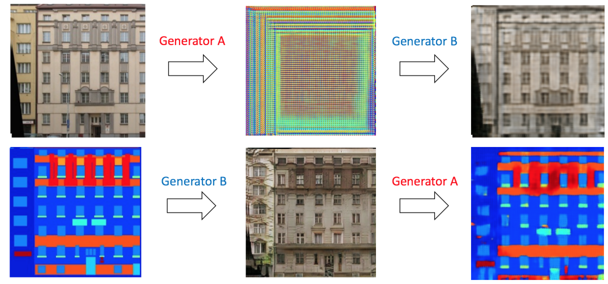
4.2. Model collapse!

In fact, we also found some problems in our experiments, part of our cycle model is collapse.



Cycle-GAN with U-net generator

The GeneratorA and GeneratorB not only learned how to change the style of the picture, but also learn to encode and decode the image. For the GeneratorA, if the input is a real image, it will encode the input into something similar to mosaic, and when the input is a fake image, it will translate it into semantic image. In this point, they both learned to identify input is real or fake, and process it in different way.



Cycle-GAN with Res-net generator

After confirm that our program is implemented without any problems, we think this phenomenon may cause by two reason.

First is because the coefficient of cycle loss is too large, and the model tend to make the result of cycle generate better.

And the second reason the difficulty of two directions generate is different, translate a house image into a semantic image is much easier than translate a semantic image into a house image, so, maybe we can use different model for generator A and B.

But, the good news is, the model collapse doesnt happen in our our fantasy-net. It work well, and that can prove that our network is more robust than others in some extent.

5. Conclusion

Although our experiments did not have some surprising results, we still very enjoy this project. In fact, the most important thing I learned in this project is the implement of this whole network, the code provided by the papers is very easy to understand and beautiful. Whether GAN model or machine learning, their development really shocked me, so, thanks to this project and course, I have learnt a lot.

At last, thanks to my partner Jinning Li and Prof. John, who helped me a lot in this project and presentation.

References

[1] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. *CVPR*, 2017.

[2] D. Pakhomov, V. Premachandran, M. Allan, M. Azizian, and N. Navab. Deep residual learning for instrument segmentation in robotic surgery. *arXiv preprint arXiv:1703.08580*, 2017.

[3] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint arXiv:1703.10593*, 2017.