Discovering the relationship between part and object with GAN

Chao Wang
chaowangplus@gmail.com

Abstract

We can recongize a cat through seeing a part of its body and we can imagine the vivid image from that. Our visual system have the ability to link part with object. We use conditional generative adversarial networks to discover the relationship between part and whole object. This make it possible for computer to have the ability as we human being have.

1 Introduction

In object recognition task, we can recognize what it is without seeing the whole object. Beyond this, we human being also have the power to draw it in our imgination. How could computer see an object and recognize it? That's a solved problem with the rapid development of convolutional neural network **?**. And surpass human-level performance on imagenet classifiaction. To understand the world is to create. So, how could computer draw an object from a part? And which is the main part of an object? What about the relationship between them? Recently, several works use GAN to discover the relationship of images included paired and unpaired images. However, all their works can only translate in image style especially the parten of color.

2 Generate part from object

To generate a obeject form a part is a problem of map a low resolution and incomplete information to a high resolution object image and whole information. It also can be treated as a conditional probblem and different can be taken as different contional information. The basic work is done with the pix2pix architecture . In this work, they use dropout as the input of noise. It becomes not effictive in this task.

2.1 Objective function

After adding the conditional information to GAN, the obejective function becomes:

$$\underset{G}{min}\,\underset{D}{max}\,V(D,G) = \mathbb{E}_{x,y\sim p_{data}(x,y)}[logD(x,y)] + \mathbb{E}_{z\sim p_z(z)}[log(1-D(G(z)))] \quad (1)$$

Where y is the label or the conditional information of the target x. In this model, y is an object part of whole object. Z is taken place with dropout in the G model layers.

2.2 Network architecture

We use 9 blocks residual network for 256*256 training resolution. It shows impressive results for neural style transfer and super resolution. In discirmative part, we use pathchGAN to classify the image patches are real or fake.

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.

Figure 1: Real part Figure 2: Fake object Figure 3: Real object

Figure 4: From part to object

3 Experiments

To discovery the relationship between part and object, we test conditional GAN in several datasets, including a fine-grained classification datasets-CUB200-2011, and a Large-scale CelebFaces Attributes Dataset (CelebA). On these two datasets, we study which part is more relevant to the whole part. It should get the same result with our visual attention systerm and reconginition knowledge.

3.1 Training details

In the experiment, we adapt the architecture in Johnson . This network contains 9 blocks for 256*256 resolution images and is used as generator. For the discriminator, we use 70*70 PatchGANs , which try to classify whether 70*70 overlapping image patches are real or fake.

To optimize the network, we follow the training settings in . For all experiments, we set $= 10$. We use the Adam solver with a batch size of 1. The network is trained from scratch, and trained with learning rate of 0.0002 for the first 100 epochs and a linearly decaying rate that goes to zero over the next 100 epochs. The experiment is implemented on PyTroch and use a single Nvidia GTX 980Ti for training and inference.

Fig 4 shows the experiment result. The left-most is the real part from the object, and the right-most is the paired real object. We generate object from real part A to fake object B. It is presented in the middle column. It can be taken as a reasonable result, for we get the real-like image from image part.

3.2 Analysis of the generator result

From the recongination task, we can get different distinguish information from different part. While the same as generation, in Fig 8, from the head of Gull we can get more realistic image sample compared to the part of leg. However, we notice that even the background of the object can get an image sample.It is not the result we want and could be seen as an mode collapse problem.

Figure 5: Real part Figure 6: Fake object Figure 7: Real object

Figure 8: The different image part with the realtionship to the same object.

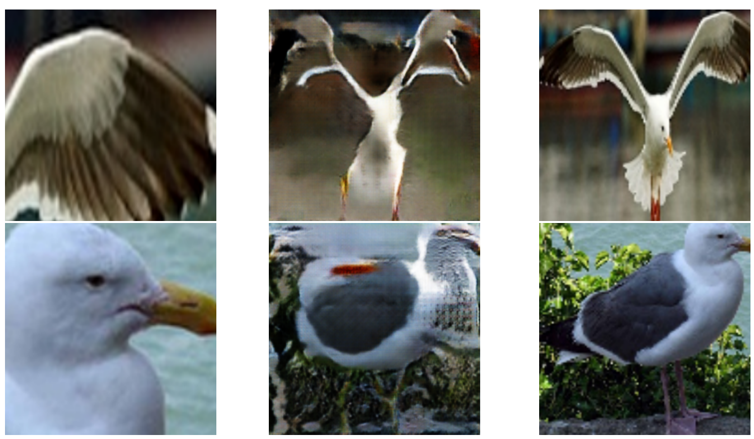

Figure 9: Real part Figure 10: Fake object Figure 11: Real object

Figure 12: The different image part with the realtionship to the same object.

There are some failure examples in the result. The uper one is out of shape and details. We should consider the mode missing problem in the training process. Because only several examples in the datasets have the flight state.We also need some experiments of training on more classes.

4 Conclusion

The result of this experiment suggests that different part of an object has a different conditional information of the whole object. We will study the power of adding extra noise information to conditional GAN with the conditional image patch.

Acknowledgments

Report of SpringCV2017