Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks

In real life, we couldn’t make stone into gold. However, in digital life, we developed a novel method to automatically “translate” an image from Claude Monet’s style into the real style. This image-to-image translation method is based the algorithm called Generative Adversarial Networks. This work was completed by Jun-Yan Zhu in Berkeley AI Research laboratory and publish in IEEE International Conference on Computer Vision (ICCV), 2017.

It can be widely applied to different image transfer situation. Besides collection style transfer, that is, transferring input images into artistic styles of Monet, Van Gogh, Ukiyo-e, and Cezanne, we can also achieve object transfiguration between horses and zebras or between apple and oranges. More fascinating, driving applications such as translation between driving scenes in different style, day to night and Computer Graphics to real become possible. So now, despite never having seen a side by side example of a Monet painting next to a photo of the scene he painted, we can make masterpieces as we wish. We are all painters.

Although this algorithm does great things, the principle isn’t hard to understand. There are two representations of given scenes as photos and Monet’s painting denoted domain X and domain Y. The author may train a mapping G: X to Y. G represents generation here. If we initial G at first, the algorithm can produce plenty lots of images which are most fake because of the arbitrary G. The author designed a discriminator which can distinguish a generated image whether fake or real. In this way, G can generate the photos we want. However, this step often fails. In order to improve the ability of G, we ask G to regenerate from Y to X, so the algorithm can not only translate the photos into Monet’s painting, but also translate Monet’s painting to the photos.

Besides visual evaluation of photos, the author adopt the “FCN score” to evaluate the results. The FCN predicts a label map for a generated photo. This label map can then be compared against the input ground truth labels.

The method outperforms baselines such as pix2pix method, CoGAN method,

SimGAN method, BiGAN/ALI, and Feature loss plus GAN method, and the regeneration from Y to X behaves well. The author observed that the reconstructed images were often close to the original inputs, even in cases where one domain represents significantly more diverse information.

Although the method can achieve compelling results in many cases, there is little success when tasks require geometric changes. The method can’t make minimal changes so it can’t translate a dog’s face to a cat’s face. What’s worse, there is a lingering gap between the results achievable with paired training data and those achieved by our unpaired method.