Image-to-Image Translation using CNN and Cycle-Consistent Adversarial Networks

Xuejiao Li, Wenxin Wei, and Ip Chun Chan
CS230 Deep Learning, Stanford University

Introduction

Image style transfer has been popular for producing samples for visualizing interior design, computer games items and create drawings. Using CNN and CycleGAN, we can not only transfer input images into artistic styles of Monet, but also make object transfiguration such as changing between a horse and a zebra, an apple and orange and even changing the background environment into a snowing scene.

Data

We gather our own unique dataset for both neural style transfer with CNN and Cycle-consistent adversarial networks from google internet, ImageNet and Flickr.

- **Neural style transfer with CNN**: We choose to use images directly from the internet, and use some painting images as our style image. We then scale our images to 400*300 pixels.
- **CycleGAN**: Depending on different applications, we choose to use different dataset as follows:
  - Making snow on the image: Use dataset provided by Flickr with the tag yosemite.
  - Horse to Zebra and Apple to Orange: Use dataset provided by ImageNet that contains horse, zebra, apple and orange as our training set.
  - Monet painting to photo: Gather Monet paintings both from the internet and Wikiart.org; and photo images both from the internet and Flickr.

  We scale the images to 256*256 pixels. For test set, we randomly download some images from the internet to gather the result.

Methods

- **Neural style transfer**: we used VGG16 convolutional neural network model and combined the content of one image with the style of another image. The cost function is:

$$J(G) = \alpha\, J_{content}(C, G) + \beta\, J_{style}(S, G)$$

- **CycleGAN**: We used 2 loss functions and the network architecture proposed by Johnson et al.
  - Architecture: Two stride-2 convolutions, several residual blocks, and two fractionally strided convolutions with stride 1/2
  - Adversarial Loss:

$$\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = \mathbb{E}_{y\sim p_{\text{data}}(y)}[\log D_Y(y)] + \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log(1 - D_Y(G(x))]$$

  - Cycle Consistency Loss:

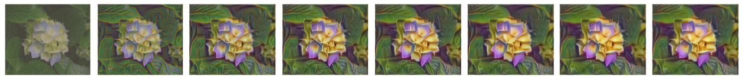
$$\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y\sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1].$$

Experiments and Results

- **Neural style transfer with CNN**:

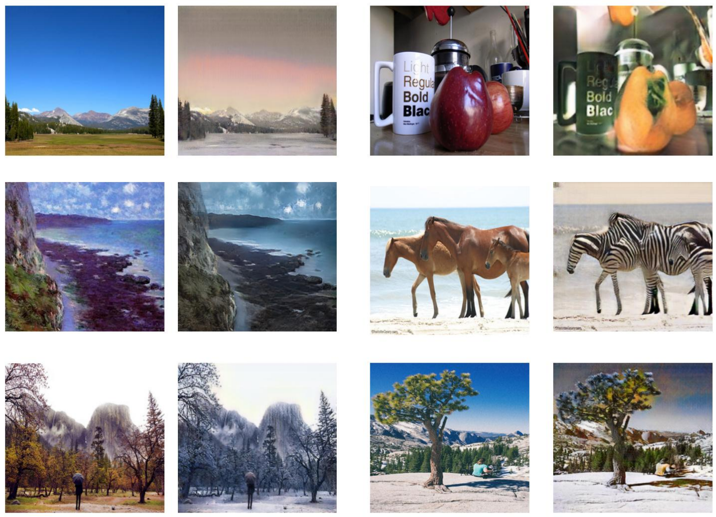
Content Image

Style Image



Generated Images

- **Cycle GAN**:



Discussion

Our method is able to successfully transfer the snow to the test image while preserving the original content. However, in some of the cases, (ex: apple to orange), the algorithm is not able to recognize the right object. It will treat similar object (ex: treat red lamp as apple) and do the style transfer. With further research, we may be able to develop an algorithm that can accurately recognize an object.

Further Work

We observe a lingering gap between the results achievable with paired training data and those achieved by our unpaired method. In some cases, this gap may be very hard to close: for example, our method sometimes permutes the labels for tree and building in the output of the photos→labels task. In the future, we will try some form of weak semantic supervision to resolve this ambiguity. Integrating weak or semi-supervised data may lead to substantially more powerful translators, still at a fraction of the annotation cost of the fully-supervised systems.

Reference

F. Luan, S. Paris, E. Shechtman, and K. Bala. Deep photo style transfer. arXiv preprint arXiv:1703.07511, 2017.
J. Zhu, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. Proceedings of the IEEE International Conference on Computer Vision, 2017