Stylization of Images using CycleGAN

Wenlong Huang
La Jolla Country Day School
(Concurrently Enrolled in COGS 181)
PID: U07648297
San Diego, USA
wenlong.huang@ljcds.org

***Abstract*—Image-to-Image translation with the goal of learning the mapping between an input image and an output image has been considered a difficult problem in computer vision. Aided by the advancement of deep learning in recent years, researchers have developed generative models that can achieve extraordinary result in image-to-image translation problems. However, most methods proposed were based on the condition that dataset with paired images were given. Paired datasets are expensive and often time not available in real life applications. In 2017, Zhu*, Park* et al. proposed CycleGAN, a Generative Adversarial Network that has a new objective function to make sure that the trained images are cycle-consistent with themselves, an approach that enables effective training even when paired datasets are not available. Their goal is to learn a mapping G: X => Y and a mapping F: Y => X at the same time and couple them to push F(G(X)) is indistinguishable from X. In this project, we collect our own training set from Flickr and introduce new applications of CycleGAN, namely adding snow to regular images and stylizing traditional cars with race car paintings. We adopt Zhenliang He's Tensorflow implementation of CycleGAN in this project and observe considerable qualitative results in both applications.**

I. INTRODUCTION

Before Krizhevsky et al. proposed AlexNet and won the 2012 ImageNet Large Scale Visual Competition (ILSVRC), it was even intractable for programs to learn detailed features of good representation of images because human experts needed to design features using their domain knowledge for computers to learn. Aided by the advancement of deep learning since then, researchers have developed successful neural network structures that can not only achieve super-human results in image classification problems but also show significant improvement in image generation. Among the generative models proposed in the recent years, Generative Adversarial Network (Goodfellow et al.) has been one of the most popular ones. An unsupervised approach, GAN consists of a generator network and a discriminator network. While the generator is trained to generate images provided with training data, the discriminator is also trained at the same time striving to discriminate the real data given in the training set from the fake data generated by the generator. This approach does not require labels for the training images since these two adversarial networks provide labels, whether each image is real or fake, for each other. And during training, it is shown to be able to learn good representation of images.

When applied to image-to-image translation problem, regular GAN usually requires paired examples for training. For example, Pix2Pix uses conditional adversarial networks in an entirely supervised setting and achieves notable result. However, paired datasets are usually not available in most settings, thus limiting the applications of such similar networks. To resolve this issue, Zhu*, Park* et al. proposed CycleGAN, a Generative Adversarial Network that has a new objective function to make sure that the trained images are cycle-consistent with themselves. And they observed superior results against those of many other prior methods including BiGAN, CoGAN, feature loss GAN, and SimGAN; it even matches Pix2Pix that is trained in a complete supervised setting with paired examples. In their paper, they also demonstrated its extraordinary qualitative results in many applications like collection style transfer, object transfiguration, season transfer, photo enhancement, etc.

People that live in places where there is usually no snow often aspire to see snow during winter and Christmas. Hence, going up the mountains to find snow and play in the snow becomes a popular activity for families in these places.

Car lovers like going to every possible car shows that they can find. And some of them are fond of seeing cars with different and interesting paintings. Those who love race cars would even paint their cars once in a while to get a sense of refreshment.

Inspired by these two ideas, in this paper, we further explore the potential of CycleGAN by applying it to the two datasets that we collect from Flickr. One dataset we collect consists of photos of roads and cityscapes that are taken not during winter and photos of them that are taken during winter in which there is snow. Our goal is thus to use CycleGAN to stylize regular photos with snow. And the other dataset we collect consists of photos of normal cars and race cars with interesting paintings. Our goal for this dataset is to use the same network to stylize normal cars with race car paintings.

II. METHODS

We fully adopt the method and the structure used in original CycleGAN paper.

Our formulation for the loss function includes both the adversarial loss, as seen in most GANs' training, and the cycle consistency loss.

For the adversarial loss, instead of maximizing the log likelihood of the data which is intractable to compute due to its high nonlinearity, we estimate the likelihood function of the data by using a convolutional neural network, a 70 x 70 PatchGAN filter (Isola et al., Li et al., Ledig et al.) that is applied convolutionally to the entire image. The advantage of such network is that it has fewer parameters and can be applied to images with arbitrarily size. The objective can be stated as:

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

(Picture from the original CycleGAN paper)

For the cycle consistency loss, we adopt the formulation in the original CycleGAN paper. Because the goal is to recover the input image, we use L1-norm to calculate the distance between the second-time generated image and the original input, namely:

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

(Picture from the original CycleGAN paper)

The reason we use L1-norm instead of L2-norm is that L1-norm has been shown to exhibit better result for comparison between colors and tones while L2-norm is better in distinguishing shape and distortion (Johnson et al.).

Hence, the full objective is:

$$\begin{aligned}\mathcal{L}(G, F, D_X, D_Y) =& \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) \\ &+ \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) \\ &+ \lambda\mathcal{L}_{\text{cyc}}(G, F),\end{aligned}$$

(Picture from the original CycleGAN paper)

Where λ is a hyperparameter that controls the relative importance of the cycle consistency loss. And our goal is to solve:

$$G^*, F^* = \arg\min_{G,F}\max_{D_x, D_Y} \mathcal{L}(G, F, D_X, D_Y).$$

(Picture from the original CycleGAN paper)

For the generator, as in the original paper, we adopt the architecture from Johnson et al. This network contains two stride-2 convolutions, several residual blocks, and two fractionallystrided convolutions with stride 1/2. Again, as in the original paper, we also use instance normalization to normalize the training batch of one image every time.

During the experiment, we set λ to be 10 and learning rate to be 0.02. As in the original paper, we keep the same learning rate for the first 100 epochs and linearly decay to zero over the next 100 epochs.

III. EXPERIMENTS

To the best of our knowledge, there is no dataset that we can adopt directly for this experiment. Therefore, we decide to collect our own data for the experiments.

We use Flickr API to collect the images. For the first dataset used to stylize images with snow, we use keywords like 'snow', 'snowfall', 'freeway', 'street', 'house', 'campus', etc. For the second dataset that stylize regular cars with race car paintings, we use keywords including 'race car', 'car sedan', 'car suv'. For each dataset, there are approximately 1000 images in each of the two classes for training and 200 images in each class for testing.

Some of the training images are shown below:

(Car2Race_Car Dataset, where the first row contains images of regular cars and the second row contains images of race cars)

(Regular2Snow Dataset, where the first row contains images of regular images without snow and the second row contains images with snow)

After 20 hours of training for each dataset on a single machine equiped with NVIDIA GTX 1070, we obtain relative success in both datasets. Some of the selected generated images are shown below:

(Car2Race_Car Dataset, where the first, second and third columns are input images of regular cars, generated images with race car paintings, and reconstructed images)

(Regular2Snow Dataset, where the first, second and third columns are input images without snow, generated images with snow, and reconstructed images)

IV. CONCLUSION

Due to the nature of the problem of image-to-image translation, we are not able to evaluate the results quatitatively. However, by examing the generated samples, the results of the experiments are relatively successful.

Compared between the results of these two datasets, however, it is noticable that the network is better at stylizing cars with race car paintings than stylizing images with added snow. This is reasonable probably because there is less variation in shape distortion in the Car2Race_Car dataset because all the images are of cars. In the Regular2Snow dataset, however, images have much more variations. For example, there are images of roads, houses, and cityscapes. Therefore, 1000 training images are probably not enough for the network to learn these variations in order to stylize the images with added snow. Potential methods for improvement include fine-tuning a network that has been trained in large datasets like ImageNet instead of training from scratch, exploring a larger hyperparameter space, and getting more training data.

Building on top of the original CycleGAN work, this paper further explores the potential of CycleGAN in different applications.

REFERENCES

[1] Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems. (2012) 1097–1105

[2] Olga Russakovsky*, Jia Deng*, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg and Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge. IJCV, 2015.

[3] J. Zhu*, T. Park*, P. Isola, A. Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. arXiv preprint arXiv: 1703.10593v3, 2017.

[4] Krizhevsky, A. Learning Multiple Layers of Features from Tiny Images. 2009

[5] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros. Imageto-image translation with conditional adversarial networks. In CVPR, 2017.

[6] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In ECCV, pages 694–711. Springer, 2016.

[7] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cun- ´ ningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image superresolution using a generative adversarial network. arXiv preprint arXiv:1609.04802, 2016.

[8] C. Li and M. Wand. Precomputed real-time texture synthesis with markovian generative adversarial networks. ECCV, 2016.

[9] I. Goodfellow. Nips 2016 tutorial: Generative adversarial networks. arXiv preprint arXiv:1701.00160, 2016.

[10] J. Donahue, P. Krahenb ¨ uhl, and T. Darrell. Adversar- ¨ ial feature learning. arXiv preprint arXiv:1605.09782, 2016.

[11] V. Dumoulin, I. Belghazi, B. Poole, A. Lamb, M. Arjovsky, O. Mastropietro, and A. Courville. Adversarially learned inference. arXiv preprint arXiv:1606.00704, 2016.

[12] M.-Y. Liu and O. Tuzel. Coupled generative adversarial networks. In NIPS, pages 469–477, 2016.

[13] A. Shrivastava, T. Pfister, O. Tuzel, J. Susskind, W. Wang, and R. Webb. Learning from simulated and unsupervised images through adversarial training. arXiv preprint arXiv:1612.07828, 2016.