draw2pix: Generative Networks for Bad Artists

Cedrick Argueta
Stanford University
450 Serra Mall, Stanford, CA 94305
cedrick@cs.stanford.edu

Kevin Wang
Stanford University
450 Serra Mall, Stanford, CA 94305
kwang98@stanford.edu

1. Introduction

We are investigating unpaired image-to-image translation with generative adversarial networks. This work was demonstrated in [4]. Work done in the same lab [2] demonstrates how we can use paired images in training to create an image-to-image mapping from one domain to another. The specific image-to-image task that we're aiming to implement is this, which demonstrates a couple examples of sketch-to-photo translations trained using the pix2pix model. We aim to do something similar, using CycleGAN to instead train on unpaired images. The relaxation that unpaired training gives us allows for easier creation of datasets and combinations of domains – we can simply swap a whole domain and retrain rather than find image-to-image pairs for that specific translation task. Our project differs from [5] in that we aren't doing multi-modal image translation, just from one domain to another. We are essentially implementing the demo from [2] with an unpaired dataset and architecture matching that of [4]

2. Related Work

describe some of these, how they relate
The original Generative Adversarial Network (GAN) proposed by Goodfellow et al. [1] proposes a network that learns an approximation of a distribution of data p_X. It does so by training two networks - a generator network and a discriminator network. These networks are trained adversarially, i.e. the discriminator's objective is to discern between real images and the generator's images, and the generator's objective is to generate images that fool the discriminator.

Generative networks are not limited to simple distributions. Work done by Liu et al. [3] demonstrated the ability of generative networks to learn a joint distribution between variables, through the use of multiple generator and discriminator networks. This can be used to generate pairs $x_1 \sim p_{X_1}, x_2 \sim p_{X_2}$ of images that belong to the joint distribution just by using the marginal distributions.

Zhu et al. [4] goes a step further by enforcing cycle consistency between pairs created, i.e. by ensuring that generated images can be run through another generator to create the paired image. CycleGAN effectively learns a mapping function $G : X \to Y$ for images in domains X and Y.

The networks discussed earlier are performing unsupervised iamge-to-image translation in that no labels are given to pair images from domain X and domain Y together. When these labels are given, supervised methods like those described by Isola et al. [2] are possible through the use of conditional GANs.

3. Problem Statement

describe image-to-image translation why do unpaired: easier for dataset collection, labels are expensive because of this, allows for more general mappings

4. Technical Approach

We aim to use CycleGAN [4] for unpaired image-to-image translation. More precisely, we will train networks to learn functions $G : X \to Y$ and $F : Y \to X$. This will allow us to take an image $x_1 \in X$ and generate its pair $y_1 \in Y$, and vice versa. We let the true distributions of the domains be p_X and p_Y. Just as in [1], we will use discriminators D_X and D_Y as adversarial networks to the generators, which we will train to distinguish between elements in X and p_X and Y and p_Y, respectively. CycleGAN uses two losses per generator function. The first loss is adversarial, where the loss function for G and its discriminator D_Y is:

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

. and the objective is $\min_G \max_{D_Y} \mathcal{L}_{GAN}$. We can apply this without loss of generality to F and its discriminator D_X. Because it is possible for a network to map an input image x to multiple images in Y, we use a cycle consistency loss to guarantee that there is a unique mapping. Specifically, we wish to enforce that for images x and y $F(G(x)) \approx x$ (forward cycle consistency) and $G(F(y)) \approx y$ (backward cycle consistency). The loss used is:

$$\mathcal{L}_{cyc} = \\ \mathbb{E}_{x \sim p_X}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_Y}[\|G(F(y)) - y\|_1] \tag{2}$$

.

Then the full loss function is:

$$\mathcal{L}(G, F, D_X, D_Y) = \\ \mathcal{L}_{GAN}(G, D_Y, X, Y) + \mathcal{L}_{GAN}(F, D_X, Y, X) + \lambda \mathcal{L}_{cyc}(G, F) \tag{3}$$

, where λ is a parameter that determines the importance of cyclic consistency loss vs. adversarial loss. Minimizing this loss through

$$G^*, F^* = \arg \min_{G,F} \max_{D_X, D_Y} \mathcal{L}(G, F, D_X, D_Y) \tag{4}$$

yields G^* and F^*, our optimal mapping functions.

5. Dataset

describe sketchy dataset

6. Baseline Results

describe cogan method, results

References

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

[2] P. Isola, J. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. *CoRR*, abs/1611.07004, 2016.

[3] M. Liu and O. Tuzel. Coupled generative adversarial networks. *CoRR*, abs/1606.07536, 2016.

[4] J. Zhu, T. Park, P. Isola, and A. A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. *CoRR*, abs/1703.10593, 2017.

[5] J. Zhu, R. Zhang, D. Pathak, T. Darrell, A. A. Efros, O. Wang, and E. Shechtman. Toward multimodal image-to-image translation. *CoRR*, abs/1711.11586, 2017.