CycleFlow: Project Progress Report

Alex Kolchinski
yakolch@stanford.edu

Andrew Kondrich
andrewk1@stanford.edu

1 Introduction

1.1 Problem and related work

Image-to-image translation is the task of mapping images between two image domains, like mapping a landscape painting to a photographic representation of the same scene or vice-versa. In the supervised setting, where training data includes labeled pairs of images, systems like pix2pix [**?**] have achieved great success in learning domain-to-domain mappings. More recently, promising results have been demonstrated in the unsupervised setting as well, where only unpaired samples are available from the two domains which the system must learn to map between [**? ? ?**]. These models generally take an approach with paired GANs [**?**], where in a setting with two domains, each of the two generators learns one of the two mapping directions, and each of the two discriminators learns to discriminate generated samples from training data for one of the two domains.

While these models achieve good results, they have the drawback of intractable likelihoods, which is inherent to GANs [**?**]. This also means that the training of such adversarially-trained models is less stable than could be achieved by training by maximum likelihood. Another drawback of the paired-GAN approach used in CycleGAN [**?**] is that the cycle consistency loss, which encourages the two generators to learn inverse mappings of each other, actually induces steganographic behavior which discourages the generators from becoming inverses of each other and affects sample quality [**?**].

Flow-based models offer an avenue to addressing these limitations. Over the last few years, models like NICE [**?**], Real NVP [**?**] and Glow [**?**] have made great progress in flow-based generative models, which learn a fully invertible mapping between a latent space with tractable likelihood estimation and sampling and the space of the data. While these models have historically been used for mapping between a data distribution and a latent space, here we demonstate a model that learns a similar invertible mapping to map between two domains in a fully reversible way, alleviating many of the drawbacks of the paired-GAN style models.

Moreover, our use of a mapping which is both invertible and has tractable determinants means that the subsequent step of learning a mapping to a shared latent space with tractable likelihoods and sampling for two paired domains should be a direct next step, allowing not only the usual advantages of single-domain models with tractable likelihoods but also encouraging the shared latent space to learn semantically meaningful disentangled representations, which would be necessary for the inter-domain mapping to be learned in the unsupervised setting.

2 Approaches

We demosntrate two approaches to flow-based inter-domain mapping with unpaired data. The first, CycleFlow, is trained adversarially in a manner similar to CycleGAN [**?**], except that the paired generators are replaced with an invertible flow-based generator. The second, PairFlow, dispenses with discriminators entirely and is trained with maximum likelihood, leveraging a shared latent space for the two domains.

Preprint. Work in progress.

2.1 CycleFlow

Something something adversarial training, replace generator etc.

2.2 PairFlow

After achieving good quality when mapping between two domains, we plan to investigate learning a shared latent space as well, by introducing a Glow-style maximum likelihood objective in which the two domains are both mapped through invertible mappings with tractable determinants to a shared latent space Z, such that $Z = G(X), Y = F(X), Y = F(G^{-1}(Z))$. Training this model with maximum likelihood, possibly with an auxiliary adversarial objective for sample quality as in FlowGAN [**?**], should yield more stable training behavior than the dual-GAN approach, in addition to tractable likelihoods and a latent space which is forced to learn a semantically disentangled representation of the two domains by the requirement of learning a semantically paired representation of the two domains simultaneously.

3 Problem Statement

3.1 Dataset

We begin with cropped and scaled map2sat for quick iteration, and intend to move to full-scale map2sat and Cityscapes-to-semantic labels.

3.2 Expected Results

Relative to CycleGAN, we posit that there will be fewer or even no steganographic artifacts in our samples due to our use of an invertible generator, which does not require an explicit cycle consistency loss to train. We expect that this will yield higher sample quality.

3.3 Evaluation

Our quantitative evaluation will be focused on the Cityscapes dataset, which is the main evaluation dataset used in CycleGAN. We will evaluate our model's performance based on FCN score [**?**], which evaluates how well our generated samples correspond to a pre-trained segmentation network. We will also use other Cityscapes metrics such as per-pixel accuracy, and mean class IOU. We also intend to include a Frechet Inception Distance (FID) score [**?**] for all datasets used. We hope to test for the amelioration of steganography in our samples using techniques described in [**?**], such as observing samples modified with adaptive histogram equalization.

4 Technical Approach

Given two domains X and Y, our goal is to learn a bidirectional generation function $F : X \to Y$ where F is invertible such that $F^{-1} : Y \to X$. We learn this function adversarially through the use of two discriminators, D_X and D_Y, where given $y \in Y$ and $x \in X$, D_X learns to distinguish real samples x from generated samples $F(y)$, and D_Y learns to distinguish real samples $y \in Y$ from generated samples $F(x)$. We include an identity loss term such that $F(y) \approx y$ and $F^{-1}(x) \approx x$.

Our loss

$$\mathcal{L} = \mathcal{L}_{GAN} + \mathcal{L}_{idt}$$

is composed of the GAN loss

$$\mathcal{L}_{GAN} = \mathbb{E}_{y \sim p_{data}(y)}[\log D_Y(y)] + \mathbb{E}_{x \sim p_{data}(x)}[\log(1 - D_Y(F^{-1}(x)))]$$

And the identity loss

$$\mathcal{L}_{idt} = \ell_1(X, F^{-1}(Y)) + \ell_1(Y, F(X))$$

This is similar to the formulation in CycleGAN [**?**], except that we do not require a cycle consistency loss as our model is automatically cycle-consistent.

Our generator is composed of a sequence of additive coupling layers, optionally with squeezing layers interposed, as in NICE [**?**] and Real NVP [**?**]. This will be augmented later as mentioned above. The discriminators, as in CycleGAN, are PatchGANs [**?**].

5 Preliminary Results

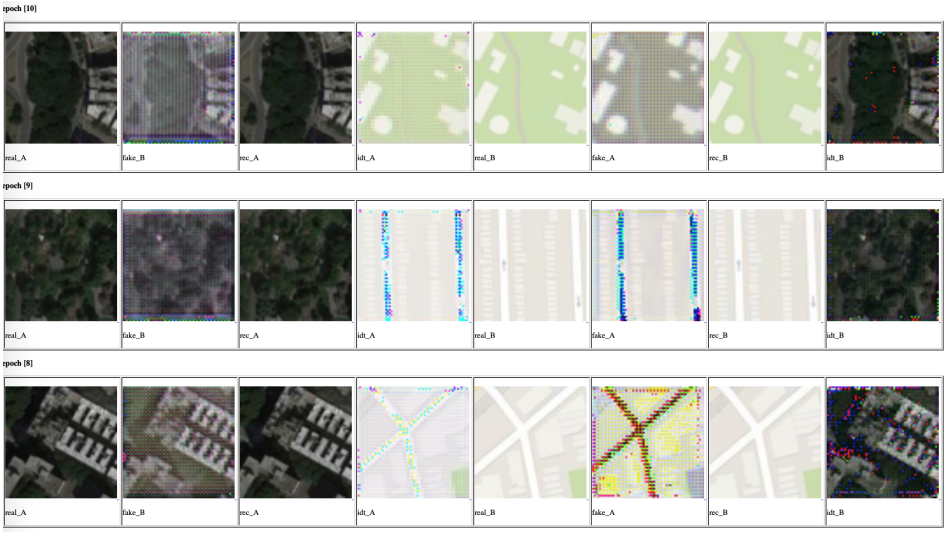

Figure 1: Sample results

Please see the enclosed figure for a representative example of our current performance. The model is implicitly cycle-consistent by nature, so the rec_A and rec_B images created by mapping an image first through one direction of the generator and then through the other (the inverse of the first) are perfect. The model also learns quite quickly to respect the identity loss and keep images largely unchanged when mapped in reverse through the generator. The model also begins to learn to map characteristics of the two domains from one to each other in the fake_A and fake_B images, but clearly the results are far from perfect. We expect that a combination of larger models, more powerful architecture as described above, and longer training will yield better results.