Guiding the One-to-One Mapping in CycleGAN via Optimal Transport[*]

Guansong Lu, Zhiming Zhou, Yuxuan Song, Kan Ren, Yong Yu

Shanghai Jiao Tong University

{gslu, heyohai, songyuxuan, kren, yyu}@apex.sjtu.edu.cn

Abstract

CycleGAN is capable of learning a one-to-one mapping between two data distributions without paired examples, achieving the task of unsupervised data translation. However, there is no theoretical guarantee on the property of the learned one-to-one mapping in CycleGAN. In this paper, we experimentally find that, under some circumstances, the one-to-one mapping learned by CycleGAN is just a random one within the large feasible solution space. Based on this observation, we explore to add extra constraints such that the one-to-one mapping is controllable and satisfies more properties related to specific tasks. We propose to solve an optimal transport mapping restrained by a task-specific cost function that reflects the desired properties, and use the barycenters of optimal transport mapping to serve as references for CycleGAN. Our experiments indicate that the proposed algorithm is capable of learning a one-to-one mapping with the desired properties.

Introduction

Image-to-image translation aims at learning a mapping between a source distribution and a target distribution, which can transform an image from the source distribution to that from the target distribution. It covers a variety of computer vision problems including image denoising (Buades, Coll, and Morel 2005), segmentation (Long, Shelhamer, and Darrell 2015), and saliency detection (Goferman, Zelnik-Manor, and Tal 2012). Along with the recent popularity of deep supervised learning, many algorithms based on paired training data and deep convolution neural networks have been proposed for specific image-to-image translation tasks. Among them, Pix2pix (Isola et al. 2016) proposed an image-to-image translation framework utilizing adversarial training technique to force the translation results being indistinguishable from the target distribution.

In practice, it is usually difficult to collect a large amount of paired training data, while unpaired data can usually be obtained more easily, hence unsupervised learning algorithms have also been widely studied. Particularly, generative adversarial networks (GANs; (Goodfellow et al. 2014))

and dual consistency (He et al. 2016) are extensively studied in image-to-image translation. CycleGAN (Zhu et al. 2017), DiscoGAN (Kim et al. 2017) and DualGAN (Yi et al. 2017) adopt these two techniques for solving unsupervised image-to-image translation where GAN loss is used to ensure the generated images being indistinguishable from real images and cycle consistency loss helps to establish a one-to-one mapping between source distribution and target distribution. In this paper, to simplify the terminology, we will use Cycle-GAN as a representative for these three similar frameworks combining GANs and the idea of cycle consistency.

CycleGAN can establish a one-to-one mapping between two data distributions unsupervisedly with the help of the cycle consistency losses in both directions. However, theoretically, there is no claim on the detailed properties of the mapping established by CycleGAN, which results in a large feasible solution space. Consequently, without meticulously designed network and hyper-parameters, the one-to-one mapping learned by CycleGAN will be a random one within this large space.

For many cross-domain translation tasks, people actually have expected properties on the learned mapping, e.g. in language translation task, people would expect the semantic meaning keeps unchanged. Hence, it will be more satisfactory if we can add explicit constraints on the one-to-one mapping within CycleGAN to control the mapping's properties, so as to meet the requirements of specific tasks.

Among the many potential feasible maps between two data distributions, it is more promising to find the optimal one according to some measure. Optimal transport (OT) aims at finding a transportation plan (Kantorovich 1942) that holds the least cost of transporting the source distribution to the target distribution, given a cost function that specifies the transportation cost between any pair of samples from the two distributions.

It is worth mentioning that the cost function in optimal transport is very flexible. For specific tasks, it is possible to define a cost function to reflect the underlining expectation of the desired mapping properties. For example, given a set of handbags and shoes, if one would like to pair the handbags with the shoes such that they have matched colors, one can specify the cost function to be the distance between their color histograms, and then the optimal transport would find the mapping that has the least overall difference in color dis-

tribution.

In summary, CycleGAN lacks the control of the one-to-one mapping, while optimal transport holds the ability to establish a mapping towards the desired property. However, the optimal transport mapping, i.e., transportation plan, is usually not a one-to-one mapping, but many-to-many instead; that is, we cannot directly use optimal transport to build a desired one-to-one mapping. We thus propose to use optimal transport as a reference to endow CycleGAN with the ability of learning a one-to-one mapping with desired properties.

The contributions of this paper has been summarized as below.

- We study the properties of the one-to-one mapping learned by CycleGAN and verify that under some circumstances the one-to-one mapping learned by CycleGAN is just a random one within the large feasible solution space, which is due to the lack of constraint on the one-to-one mapping established by CycleGAN.

- We propose to use the optimal transport with respect to a task-specific metric to guide CycleGAN on learning a one-to-one mapping with desired properties. Our experiments on several datasets demonstrate the effectiveness of the proposed algorithm on learning a desired one-to-one mapping.

Related Work

Generative Adversarial Networks (GANs), consisting of a generator network and a discriminator network, is originally proposed as a generative model to match the distribution of generated samples to the real distribution, where the discriminator is trained to distinguish generated samples from real ones while the generator learns to generate samples that fool the discriminator. Researchers have been working hard on improving the stability of training and exploiting the capacity of GANs for various computer vision tasks. For instance, (Radford, Metz, and Chintala 2015) proposes a deep convolutional architecture that stabilizes the training; WGAN (Arjovsky, Chintala, and Bottou 2017) proposes to utilize Wasserstein-1 distance (or Earth Mover's distance/EMD) as an alternative metric.

Conditional GANs (cGANs; (Mirza and Osindero 2014; Odena, Olah, and Shlens 2016; Zhou et al. 2017)) proposes to extend GANs to a conditional model by conditioning some extra information, such as class label, on both generator and discriminator in GANs so that it can generate images conditioned on class labels and so on. (Reed et al. 2016) extends cGANs with conditional information being text features. Pix2pix (Isola et al. 2016) proposed a unified image-to-image translation framework based on conditional GANs, with conditional information being images.

In practice, it is always hard to collect a large amount of paired training data, while unpaired data can always be obtained more easily. In order to make better use of unpaired data in real world, CycleGAN (Zhu et al. 2017), DiscoGAN (Kim et al. 2017) and DualGAN (Yi et al. 2017) adopt the idea of dual consistency, which was firstly proposed in language translation (He et al. 2016), together with GANs to

simultaneously train a pair of generators and discriminators for translation in both directions and applied cycle consistency loss on both data distributions, which forces the mapping to be a one-to-one mapping. However, theoretically, there is no explicit constraint on the properties of the one-to-one mapping within CycleGAN, which results in a large feasible solution space and the learned one-to-one mapping being a random one within this space.

Optimal transport (Villani 2008) aims to find a mapping between two distributions that can transport the source distribution to the target distribution with the least transportation cost. In many cases, the mapping between two distributions, where each source point only maps to a single target point (the Monge's problem) does not exist. The modern approach to optimal transport relaxes the Monge's problem by optimizing over plans, i.e., a distribution over the product space of the source distribution space and the target distribution space. (Cuturi 2013) proposes to introduce entropic regularization term into OT problem which turns it into an easier optimization problem and can be solved efficiently by Sinkhorn-Knopp algorithm. (Seguy et al. 2017) proposed a stochastic approach for solving large-scale regularized OT and estimating a Monge mapping as a deep neural network approximating the barycentric mapping of the OT plan.

Method

Given two sets of unpaired images that respectively from domain U and domain V, the primal task of unsupervised image-to-image translation is to learn a generator G_{uv} : $U \to V$ that maps an image $u \in U$ to an image $v \in V$. The modern techniques (Zhu et al. 2017; Yi et al. 2017; Kim et al. 2017) of unsupervised image-to-image translation introduce an extra generator $G_{vu} : V \to U$ that maps an image $v \in V$ to an image $u \in U$ and cycle consistency loss, i.e., $G_{vu}(G_{uv}(u)) \approx u$ and $G_{uv}(G_{vu}(v)) \approx v$, is introduced to regularize the mapping between U and V. As the result, the learned mapping would be a bijection, i.e., a one-to-one mapping. However, as we will discuss in the latter of this section, cycle consistency loss, though helps build a one-to-one mapping, has no control on the properties of the learned one-to-one mapping. In this section, we will also discuss how to add extra constraints on the learning of the one-to-one mapping to chase desired properties.

Preliminary: CycleGAN

In CycleGAN, besides the above-mentioned two coupled generators G_{uv} and G_{vu} that translate images across domain U and V and the cycle consistency losses that regularize the learned mapping to be a bijection, it also introduces an adversarial loss to each generator to ensure translated images are valid samples. More strictly, by playing a minimax game with the discriminator, the adversarial loss forces the generator to match the distribution of generated images with the distribution of real images in the target domain.

Adversarial Loss   In the original GAN (Goodfellow et al. 2014), the discriminator was formulated as a binary classifier outputting a probability. Given a real image distribution
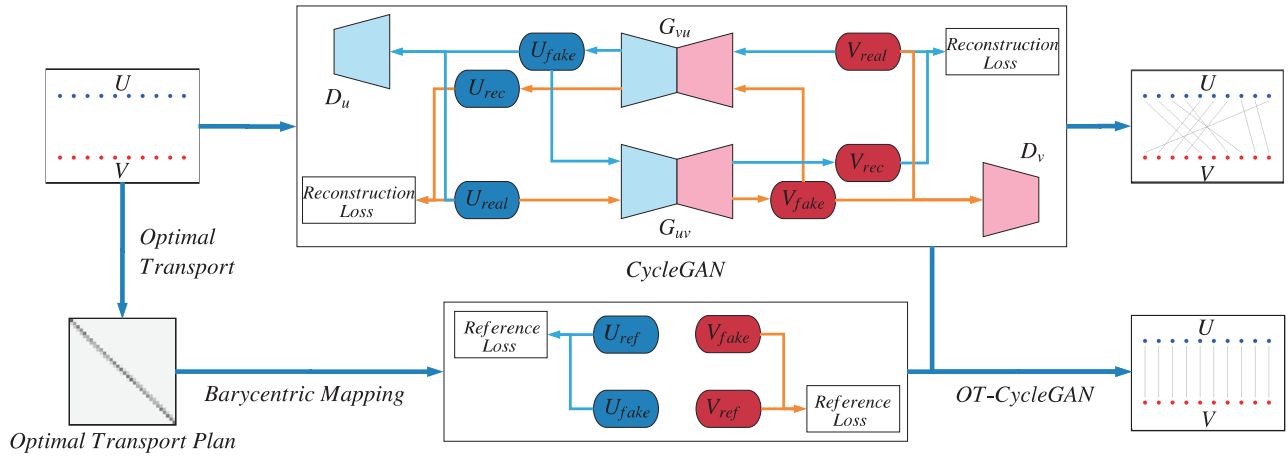
Figure 1: The framework of the proposed method. We use the barycentric mapping of optimal transport, which minimizes the cost of mismatching of a task-specific property, to guide the CycleGAN on learning a one-to-one mapping with the desired properties.

ν and the fake image distribution formed by generated samples $G_{uv}(u)$ with $u \sim \mu$, the loss function of original GAN is defined as:

$$L_{gan}(G_{uv}, D_v) = \mathbb{E}_{u \sim \mu}[\log(1 - D_v(G_{uv}(u)))] \\ + \mathbb{E}_{v \sim \nu}[\log D_v(v)], \quad (1)$$

The discriminator D_v learns to maximize $L_{gan}(G_{uv}, D_v)$, that is to distinguish the real samples and the fake samples, while the generator learns to minimize $L_{gan}(G_{uv}, D_v)$ such as to make the generated samples have a low probability of being classified as fake by the discriminator. When D_v is assumed to be optimal, the objective of generator is to minimize the Jensen-Shannon divergence between $G_{uv}(\mu)$ and ν, and the minimum is achieved if and only if $G_{uv}(\mu) = \nu$.

Although GANs have achieved great success in the realistic image generation, training of the original GANs turns out to be very difficult and one has to carefully balance the ability of generator and discriminator. It was showed in (Arjovsky and Bottou 2017; Arjovsky, Chintala, and Bottou 2017) that Jensen-Shannon divergence is ill-defined when the supports of the two distributions are not overlapped. Wasserstein distance is thus introduced (Arjovsky, Chintala, and Bottou 2017) as an alternative metric for evaluating the distance between the real and fake distributions. Wasserstein distance $W(\mu, \nu)$ is defined as the minimal cost of transporting distribution μ into ν. In its primal form, it is formally defined as:

$$W(\mu, \nu) = \inf_{\pi \in \Pi(\mu, \nu)} \mathbb{E}_{(u,v) \sim \pi}[d(u, v)], \quad (2)$$

where $\Pi(\mu, \nu)$ denotes the collection of all probability measures on $U \times V$ with marginals μ on U and ν on V.

Since the infimum in Eq. (2) is highly intractable, in WGAN (Arjovsky, Chintala, and Bottou 2017), the discriminator (critic) is designed to estimate the Wasserstein distance by solving its dual form, with the corresponding objective defined as:

$$L_{wgan}(G_{uv}, D_v) = \mathbb{E}_{v \sim \nu}[D_v(v)] - \mathbb{E}_{u \sim \mu}[D_v(G_{uv}(u))], \quad (3)$$

where the discriminator is constrained as a 1-Lipschitz function. The problem of how to properly enforce 1-Lipschitz has evolved a set of investigations (Gulrajani et al. 2017; Miyato et al. 2018; Petzka, Fischer, and Lukovnicov 2017). In our experiments, these solutions show very similar results and we choose the Gradient-Penalty (Gulrajani et al. 2017) loss for on-the-fly example through the paper, i.e.,

$$L_{gp}(D_v) = \mathbb{E}_{\hat{x} \sim P_{\hat{x}}}[(\|\nabla D_v(x)\|_2 - 1)^2], \quad (4)$$

where $P_{\hat{x}}$ is the distribution of uniformly distributed linear interpolations of $v \sim \nu$ and $x \sim G_{uv}(\mu)$.

Cycle Consistency Loss   Training G_{uv} with respect to the adversarial loss forces the distribution of $G_{uv}(u)$ to match with the distribution ν. However, this actually does not build any relationship between the source domain and the target domain. Without paired data, traditional approaches build the relationship between the domain data via predefined similarity function (Bousmalis et al. ; Shrivastava et al. 2017; Taigman, Polyak, and Wolf 2016) or assuming shared low-dimensional embedding space (Liu, Breuel, and Kautz 2017; Aytar et al. 2017). In CycleGAN series (Zhu et al. 2017; Kim et al. 2017; Yi et al. 2017), a dual task of translating data from domain V to domain U is introduced and cycle consistency is encouraged as a regularization.

Specifically, cycle consistency requires any image u in domain U can be reconstructed after applying G_{uv} and G_{vu} on u in turn and any image v in domain V can be reconstructed after applying G_{vu} and G_{uv} on v reversely. That is, $G_{vu}(G_{uv}(u)) \approx u$, $G_{uv}(G_{vu}(v)) \approx v$. The cycle consistency loss can be formulated as follow:

$$L_{rec}(G_{uv}) = \mathbb{E}_{u \sim \mu}[\|G_{vu}(G_{uv}(u)) - u\|], \\ L_{rec}(G_{vu}) = \mathbb{E}_{v \sim \nu}[\|G_{uv}(G_{vu}(v)) - v\|], \quad (5)$$

in which we adopt L_2 distance to measure the distance between the original image and the reconstructed image.

The One-to-one Mapping in CycleGAN

In CycleGAN, the adversarial losses applied on two generators help to establish the mappings between domain U and domain V in both directions, as it forces the generated images to be within the target domain. Meanwhile, the cycle consistency losses help to relate these two mappings and force them to be two one-to-one mappings, as it forces different samples in the source domain to be mapped to different samples in the target domain (otherwise, the consistency loss would be large). Therefore, CycleGAN would establish a bijective mapping between domain U and domain V, which is also mentioned in DiscoGAN (Kim et al. 2017) and CycleGAN (Zhu et al. 2017).

It is promising that CycleGAN can find a one-to-one mapping between two data distributions unsupervisedly. But theoretically, there exists a large number of one-to-one mappings between two data distributions. For example, the number of possible one-to-one mapping between two discrete data distributions with each containing n discrete data points is the factorial of n, i.e. $n!$. And all these one-to-one mappings are perfect in terms of CycleGAN's objective.

Since there is no extra control on the properties of the mapping, as long as it is one-to-one, the learned one-to-one mapping with CycleGAN would theoretically be a random one in this large feasible solution space.

For verification, we conducted experiment across two synthetic datasets A and B, each consists of 32 images in the resolution of 64x64, with each image contains one vertical line at a different position as showed in Figure (2a). The resulting mapping learned with CycleGAN is showed in Figure (2b). As we can see, images with the vertical line in different positions in A is mapped to images in B without any order. Furthermore, this one-to-one mapping changes, given different initializations and hyper-parameters.

Guiding CycleGAN with Optimal Transport

As discussed above, the one-to-one mapping learned by CycleGAN can be random in the large feasible solution space. However, in many practical applications, we would expect certain feature getting matched in the learned mapping. For example, when the two domains are different languages, one may expect the semantic information of characters keeps unchanged after translation. Without any additional control, the one-to-one mapping function learned by CycleGAN, in theory, will fail to achieve this with a very high probability (approaching one as the number of term increasing).

Here we propose to make use of the controllability in optimal transport to endow CycleGAN with the ability of learning one-to-one mapping with desired properties.

Optimal Transport (OT)   According to Kantorovich formulation (Kantorovitch 1958), the typical optimal transport problem can be defined as finding a mapping function π between two distribution μ and ν, which is optimal with respect to cost function $c(x, y)$, and it can be formulated as follows:

$$\pi^* = \underset{\pi \in \Pi(\mu, \nu)}{\operatorname{argmin}} \int_{\mu \times \nu} c(u, v)\pi(u, v)dudv, \qquad (6)$$

where $\Pi(\mu, \nu)$ denotes the collection of all probability measures on $U \times V$ with marginals μ on U and ν on V, as in the primal form of Wasserstein distance. In fact, Wasserstein distance is a special form of optimal transport with the cost function $c(u, v)$ required to be a distance (a proper metric), while in optimal transport, $c(u, v)$ can be any cost function. Another difference is that, as an adversarial objective, Wasserstein distance is conducted between the distribution formed by $G_{uv}(u)$ and the target distribution ν, while the optimal transport is conducted between the source distribution μ and the target distribution ν. And here we will focus more on the optimal transport plan, instead of the optimal cost.

Reflecting the Desired Properties with OT   Given two distribution μ and ν, CycleGAN builds a one-to-one mapping $\pi(u, v)$ between μ and ν. As we discussed previously, the one-to-one mapping might be a random one in the feasible solution space. However, in the specific tasks, people actually have an expectation on the outcome of the learned one-to-one mapping, e.g. pixel-level distance or average hue difference was expected to be low, the outline or semantic meaning was expected to be unchanged and so on.

One way to model the expectation is to define a task-specific cost function $c(x, y)$ and then the satisfaction degree of the expectation, if it is defined to be the averaged satisfaction degree of all pairing, can be modeled as the transport cost $\int_{\mu \times \nu} c(u, v)\pi(u, v)dudv$, Eq. (6). It follows that, given a task-specific cost function $c(x, y)$, in terms of the optimal transport, the best mapping is the π^*.

We thus propose to solve the optimal transport problem under the task-specific cost function and use the optimal transport plan π^* as a reference to build the one-to-one mapping in CycleGAN.

Optimal Transport Plan as Reference   Given an arbitrary cost function, the optimal transport plan is usually a many-to-many mapping, i.e. $\pi(u|v)$ and $\pi(v|u)$ is usually not a Dirac delta distribution. Therefore, it is not feasible in cross-domain translation tasks, and some previous work (Perrot et al. 2016; Seguy et al. 2017) attempt to use the Barycenter instead. The Barycenter of u in the source distribution μ is defined to be a sample $v_b^*(u)$ in target domain V that has the minimal transport cost to its transport targets $\pi(v|u)$:

$$v_b^*(u) = \underset{v_b}{\operatorname{argmin}} \int_{v \sim \pi(v|u)} c(v_b, v)dv, \qquad (7)$$

However, the Barycenter is not guaranteed to lie in the distribution ν, which in practice behaves as blurring images.

We thus proposed that, instead of directly using the optimal transport plan or the Barycenter, we train a CycleGAN and use the Barycenter of the optimal transport plan as a reference to guide the establishment of its one-to-one mapping. Given a proper weight on this regularization, CycleGAN will be able to learn a one-to-one mapping that basically follows the optimal transport plan, while at the same time, makes each translated sample lies in the target distribution under the supervision of adversarial loss. Our algorithm can then be separated into two steps:

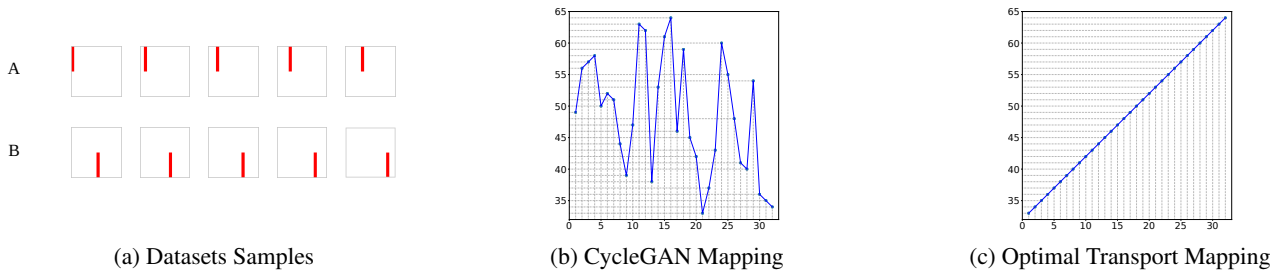| (a) Datasets Samples | (b) CycleGAN Mapping | (c) Optimal Transport Mapping |

Figure 2: Synthetic experiments: CycleGAN learns a one-to-one mapping between datasets A and B, however, the learned mapping is out-of-order. Defining the cost function to be the squared Euclidean distance of the locations of their vertical lines, the optimal transport is capable of mapping the images in dataset A to B in sequence. This illustrates the randomness of the one-to-one mapping established via CycleGAN and at the same time show the ability of optimal transport to build a desired mapping, given task-specific cost function. The x- and y-axis ticks in sub-figure (b) and (c) indicate the images with the specified locations of the vertical line in domain A and B respectively.

	CycleGAN	Optimal Transport	Nearest Neighbor
Controlling	N	Y	Y
Mapping	One-to-One	Many-to-Many	N/A
Generalization	Y	N	N

Table 1: Comparison among CycleGAN, optimal transport, and nearest neighbor. The nearest neighbor and optimal transport are capable of controlling the mapping with respect to a given metric between two samples. However, the mapping build via nearest neighbor does not form a joint distribution, i.e. may collapse to a subset, and optimal transport usually builds a many-to-many mapping, which is not adequate in cross-domain translate. And also, the optimal transport plan does not generalize to out-of-distribution samples.

- Firstly, given two distributions and a task-specific cost function, we learn an optimal transport plan between the two distributions, and we evaluate the Barycenter $v_b^*(u)$ and $u_b^*(v)$ for each sample in the two distributions.

- Secondly, we train a CycleGAN model using these Barycenters as references to the two cross-domain generators. The corresponding reference loss is defined as follows:

$$L_{ref}(G_{uv}) = \mathbb{E}_{u \sim \mu}[\|G_{uv}(u) - v_b^*(u)\|],$$
$$L_{ref}(G_{vu}) = \mathbb{E}_{v \sim \nu}[\|G_{vu}(v) - u_b^*(v)\|]. \tag{8}$$

The full objective of our algorithms can be formulated as:

$$\begin{aligned} L(G_{uv}, G_{vu}, D_u, D_v) = {} & L_{wgan}(G_{uv}, D_v) + L_{wgan}(G_{vu}, D_u) \\ & - \lambda_{gp} \cdot (L_{gp}(D_u) + L_{gp}(D_v)) \\ & + \lambda_{rec} \cdot (L_{rec}(G_{uv}) + L_{rec}(G_{vu})) \\ & + \lambda_{ref} \cdot (L_{ref}(G_{uv}) + L_{ref}(G_{vu})), \end{aligned} \tag{9}$$

where G_{uv} and G_{vu} are optimized to minimize the objective, while D_v and D_u are optimized to maximize the objective. We will later refer to this model as OT-CycleGAN.

Discussions

As discussed in the previous sections, in the sense of establishing a mapping between two data distributions, Cycle-GAN and optimal transport both have strengths and weaknesses. This motivates us to use the barycenters of optimal transport mapping to serve as the references of CycleGAN, so as to combine the strengths of the two models to establish a one-to-one mapping with (mostly) minimized mismatching cost over task-specific properties between two data distributions.

Another difference between CycleGAN and optimal transport is that optimal transport establishes a mapping between samples in both datasets mathematically. Under the circumstance of two discrete datasets, it cannot **generalize** to out-of-distribution samples. In contrast, CycleGAN learns the mapping function between two distributions via two neural networks and thus has the ability to generalize to out-of-distribution samples. When the two discrete datasets hold the same number of unduplicated samples, a perfect one-to-one mapping actually may also exist in optimal transport. Under such conditions, CycleGAN helps optimal transport generalized to out-of-distribution samples.

Besides optimal transport, **nearest neighbor** might also come to mind for controlling the mapping to have matched properties. With the nearest neighbor algorithm, every sample in the source distribution will be mapped to the nearest one in the target distribution. However, nearest neighbor is a local algorithm, and without considering the global status, the mapping established via nearest neighbor might collapse to a subset in the target domain or even a single point. For example, source domain is a set of real numbers whose range is [0, 31] while the range of target domain is [32, 63] and the cost function is specified as the squared difference. In this case, the nearest neighbor would map all samples in the source domain to the 'leftmost' one in the target domain *i.e.* 32. In comparison, optimal transport will map the whole source domain to the whole target domain in sequence. We summarize the discussion among CycleGAN, optimal transport and nearest neighbor in Table (1).

Experiments

In order to demonstrate the effectiveness of our proposed algorithm for learning a one-to-one mapping between two data distributions with desired properties, we conduct sev-

eral image-to-image translation experiments between different datasets, and we compare the translation results of our algorithm with CycleGAN. Details of our experimental setting are as follows.

Network Architecture

In our experiments, we adopted the architecture of autoencoder (Hinton and Salakhutdinov 2006) in both of our generators. The encoder is composed of a set of stride-2 convolution layers with a 4x4 filter, while the decoder is composed of several stride-2 deconvolution layers with 4x4 filter. Each convolution layer in the encoder or deconvolution layer in the decoder is followed by a normalization layer except the first and the last one. We use WGAN-GP loss instead of the original GAN loss in our experiments. The architecture of discriminator (critic) is designed to be the similar as the decoder, except that we eliminate all normalization layers.

Optimization Details

We use network simplex algorithm (Damian, Comm, and Garret 1991) for solving the optimal transport problem between two data distributions as linear programming. Due to the lack of computation power, we use L2 barycenter instead of accurate barycenter to obtain the barycentric mapping out of the previously-obtained optimal transport plan, which can be simplified as the weighted sum of mapped samples. We use Adam (Kingma and Ba 2014) optimizer with $\beta_1 = 0.5$, $\beta_2 = 0.999$. We train our model for 3000 epochs with an initial learning rate of 0.0002 and linearly decayed it to zero. λ_{gp} is set as 10, λ_{rec} is set in the range of [100, 800] and λ_{ref} is set in the range of [50, 300]. We train critic for 5 steps and generator for 1 step in turn.

Experiment: Car-to-Chair

We conduct our first experiment between a car dataset (Fidler, Dickinson, and Urtasun 2012) and a chair dataset (Aubry et al. 2014). Both datasets consist of images of 3D rendered objects with varying azimuth angles and the value of azimuth angle of each image is provided by the dataset. Figure (3b) shows the translation results of CycleGAN between these two datasets. As we can see, as the images of car vary in azimuth angle in order, the translation results are random samples in the target domain.

OT Barycenter   By using the azimuth angle of each image provided by each dataset and specifying the cost function between each image to be the squared difference of azimuth angle, we are able to find an optimal transport plan that can transport the car distribution to the chair distribution with the least overall azimuth angle difference. Additionally, as there is more than one image at each azimuth angle, we further use the Euclidean distance between the average RGB color of each image (exclude the white background) to as subsidiary cost function, such to find an optimal transport plan that can further minimize the overall color difference. In summary, the task-specific cost function in this experiments is formulated to be:

$$c(x,y) = d_{angle}(x,y) + \lambda_{color} \cdot d_{color}(x,y). \quad (10)$$

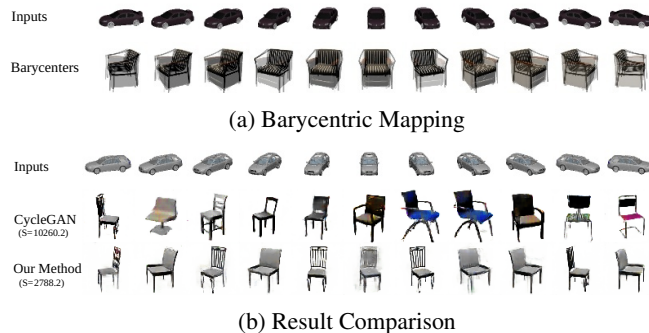The samples of resulting barycentric mapping are illustrated in Figure (3a).

(a) Barycentric Mapping

(b) Result Comparison

Figure 3: Car-to-Chair Experiments.

OT-CycleGAN Result   Figure (3b) shows translation results of our algorithm. The resulting mapping of OT-CycleGAN successfully matches the azimuth angles and colors of the generator's input and output. We also evaluate the **mismatching degree** $S = \int_{\mu \times \nu} c(u,v)\pi(u,v)dudv$ for each method. As listed in Table (2), OT-CycleGAN achieves a much lower mismatching degree.

Azimuth-Angle Mapping Analysis   We plot the overall azimuth-angle mapping to provide a global comparison between CycleGAN and OT-CycleGAN. As we can see in Figure (4), the resulting azimuth-angle mapping with CycleGAN is fairly random, while the OT-CycleGAN mostly matches the azimuth-angle of input and output. The azimuth-angle of translated image is obtained via finding its nearest neighbor in the training set. It worth mentioning that here we ignored the color attribute, therefore, the result is a superposition over images of different colors.
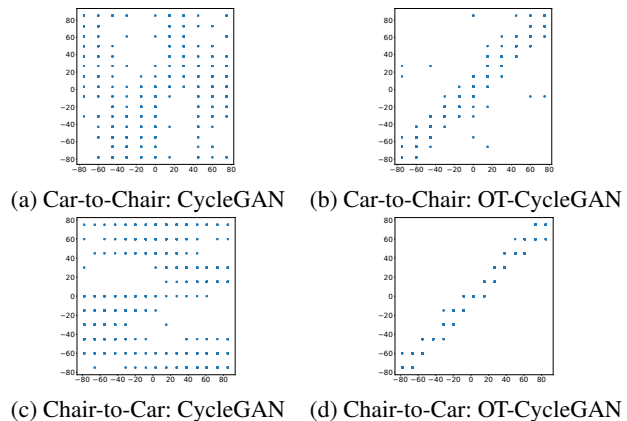

(a) Car-to-Chair: CycleGAN   (b) Car-to-Chair: OT-CycleGAN

(c) Chair-to-Car: CycleGAN   (d) Chair-to-Car: OT-CycleGAN

Figure 4: Azimuth angle mapping of Car-to-Chair.

Experiment: Shoes-to-Handbags

In this experiment, we performed image-to-image translation between a shoes dataset (Yu and Grauman 2014) and

	CycleGAN	OT-CycleGAN				
		$\lambda_{ref} = 50$	$\lambda_{ref} = 100$	$\lambda_{ref} = 200$	$\lambda_{ref} = 300$	$\lambda_{ref} = 500$
Mismatching Degree ($\times 10^4$)	1.026	0.5634	0.3393	**0.2788**	0.2865	0.3023

Table 2: Comparison between CycleGAN and OT-CycleGAN in terms of mismatching degree.

(a) Barycentric Mapping

(b) Result Comparison

Figure 5: Shoes-to-Handbags experiments.

Figure 6: Tuning the parameter λ_{ref}.

a handbags dataset (Zhu et al. 2016). Figure (5b) shows the translation results of CycleGAN between these two datasets. As we can see, the translation results are of an obvious color difference from the source samples.

OT Barycenter   In this experiment, we would like to establish a one-to-one mapping that matches the color of the handbags with the color of shoes. As the color in each image of these two datasets is much more complex than the previously used car and chair datasets, it would be inaccurate to use the average color to represent the color information of each image, we thus adopted a color histogram to represent color information of each image. We use the Wasserstein distance $W(\mu, \nu)$ between two histograms μ and ν as the cost function, with $d(u, v)$ being the Euclidean distance of two color bins u and v in Lab color space. Samples of the resulting barycentric mapping are showed in Figure (5a).

OT-CycleGAN Result   Figure (5b) illustrates the mapping function learned by our method (OT-CycleGAN). Compared with the original CycleGAN, the mapping established by our algorithm is significantly better, in terms of whether the color distributions match each other, in both visual and quantitative metric S.

Reference Weight

One important parameter in OT-CycleGAN is λ_{ref}, i.e. the weight of OT reference loss to CycleGAN. Ideally, if λ_{ref} is extremely large, the resulting mapping will be identical to the barycentric mapping of OT, while if the λ_{ref} is extremely small, the reference loss will not take effect and the result will be similar to CycleGAN, which is evidenced in Figure 6. More results are summarized in Table (2), and we can see there exists a pretty large range of λ_{ref} where OT-CycleGAN is able to learn a satisfactory mapping.

Discussion

U-Net architecture is mostly used in image-to-image translation tasks, though it tends to connect the pixel information between the input and output and achieved many satisfactory results, it, however, does not theoretically guarantee the relationship between source and target and thus may require extensive tuning if a special property is wanted. Our method, in contrast, can directly specify which properties to be matched.

Conclusion and Future Work

We have presented OT-CycleGAN where an optimal transport mapping is used to guide the one-to-one mapping established by CycleGAN. With the proposed algorithm, one can control the learned one-to-one mapping in CycleGAN via defining a task-specific cost function that reflecting the desired mapping properties.

Specifically, we demonstrate that there is no controllability on the properties of the learned one-to-one mapping in CycleGAN, and optimal transport can provide a mapping that minimizing the overall cost of mismatching of expected properties, given a task-specific cost function. Since the optimal transport mapping is usually not one-to-one, we propose to use the Barycenters of learned mapping as references to guide the training of CycleGAN to form a one-to-one mapping with desired mapping properties.

Experiments conducted on several benchmark datasets have shown that the mapping function learned by vanilla CycleGAN can be quite messy and the guiding of optimal transport can significantly improve the mapping in terms of the task-specific properties.

In the main-body and experiments, we mainly focused on image-to-image translation, as it is the most successful application of CycleGAN. We hope the detailed analysis of the properties of CycleGAN and optimal transport would also benefit further investigation on cycle consistency loss and unsupervised cross-domain translation. OT-CycleGAN is a general framework for establishing one-to-one mapping

with desired properties and we plan to investigate more related tasks in the further.

References

Arjovsky, M., and Bottou, L. 2017. Towards principled methods for training generative adversarial networks. In *ICLR*.

Arjovsky, M.; Chintala, S.; and Bottou, L. 2017. Wasserstein gan. *arXiv preprint arXiv:1701.07875*.

Aubry, M.; Maturana, D.; Efros, A. A.; Russell, B. C.; and Sivic, J. 2014. Seeing 3d chairs: exemplar part-based 2d-3d alignment using a large dataset of cad models. In *CVPR*, 3762–3769.

Aytar, Y.; Castrejon, L.; Vondrick, C.; Pirsiavash, H.; and Torralba, A. 2017. Cross-modal scene networks. *PAMI*.

Bousmalis, K.; Silberman, N.; Dohan, D.; Erhan, D.; and Krishnan, D. Unsupervised pixel-level domain adaptation with generative adversarial networks.

Buades, A.; Coll, B.; and Morel, J.-M. 2005. A non-local algorithm for image denoising. In *CVPR*, volume 2, 60–65. IEEE.

Cuturi, M. 2013. Sinkhorn distances: Lightspeed computation of optimal transport. In *NIPS*, 2292–2300.

Damian, K.; Comm, B.; and Garret, M. 1991. *The minimum Cost Flow Problem and The Network Simplex Method*. Ph.D. Dissertation, Dissertation de Mastère, Université College Gublin, Irlande.

Fidler, S.; Dickinson, S.; and Urtasun, R. 2012. 3d object detection and viewpoint estimation with a deformable 3d cuboid model. In *NIPS*.

Goferman, S.; Zelnik-Manor, L.; and Tal, A. 2012. Context-aware saliency detection. *PAMI* 34(10):1915–1926.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative adversarial nets. In *NIPS*.

Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; and Courville, A. 2017. Improved training of wasserstein gans. *arXiv preprint arXiv:1704.00028*.

He, D.; Xia, Y.; Qin, T.; Wang, L.; Yu, N.; Liu, T.; and Ma, W.-Y. 2016. Dual learning for machine translation. In *NIPS*, 820–828.

Hinton, G. E., and Salakhutdinov, R. R. 2006. Reducing the dimensionality of data with neural networks. *science* 313(5786):504–507.

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

Kantorovich, L. V. 1942. On the translocation of masses. In *Dokl. Akad. Nauk. USSR (NS)*, volume 37, 199–201.

Kantorovitch, L. 1958. On the translocation of masses. *Management Science* 5(1):1–4.

Kim, T.; Cha, M.; Kim, H.; Lee, J. K.; and Kim, J. 2017. Learning to discover cross-domain relations with generative adversarial networks. *arXiv preprint arXiv:1703.05192*.

Kingma, D. P., and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Liu, M.-Y.; Breuel, T.; and Kautz, J. 2017. Unsupervised image-to-image translation networks. In *NIPS*, 700–708.

Long, J.; Shelhamer, E.; and Darrell, T. 2015. Fully convolutional networks for semantic segmentation. In *CVPR*, 3431–3440.

Mirza, M., and Osindero, S. 2014. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.

Miyato, T.; Kataoka, T.; Koyama, M.; and Yoshida, Y. 2018. Spectral normalization for generative adversarial networks. *arXiv preprint arXiv:1802.05957*.

Odena, A.; Olah, C.; and Shlens, J. 2016. Conditional image synthesis with auxiliary classifier gans. *arXiv preprint arXiv:1610.09585*.

Perrot, M.; Courty, N.; Flamary, R.; and Habrard, A. 2016. Mapping estimation for discrete optimal transport. In *NIPS*, 4197–4205.

Petzka, H.; Fischer, A.; and Lukovnicov, D. 2017. On the regularization of wasserstein gans. *arXiv preprint arXiv:1709.08894*.

Radford, A.; Metz, L.; and Chintala, S. 2015. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*.

Reed, S.; Akata, Z.; Yan, X.; Logeswaran, L.; Schiele, B.; and Lee, H. 2016. Generative adversarial to image synthesis. In *ICML*, volume 3.

Seguy, V.; Damodaran, B. B.; Flamary, R.; Courty, N.; Rolet, A.; and Blondel, M. 2017. Large-scale optimal transport and mapping estimation. *arXiv preprint arXiv:1711.02283*.

Shrivastava, A.; Pfister, T.; Tuzel, O.; Susskind, J.; Wang, W.; and Webb, R. 2017. Learning from simulated and unsupervised images through adversarial training. In *CVPR*, volume 2, 5.

Taigman, Y.; Polyak, A.; and Wolf, L. 2016. Unsupervised cross-domain image generation. *arXiv preprint arXiv:1611.02200*.

Villani, C. 2008. *Optimal transport: old and new*, volume 338. Springer Science & Business Media.

Yi, Z.; Zhang, H. R.; Tan, P.; and Gong, M. 2017. Dualgan: Unsupervised dual learning for image-to-image translation. In *ICCV*, 2868–2876.

Yu, A., and Grauman, K. 2014. Fine-grained visual comparisons with local learning. In *CVPR*, 192–199.

Zhou, Z.; Rong, S.; Cai, H.; Zhang, W.; Yu, Y.; and Wang, J. 2017. Activation maximization generative adversarial nets. *arXiv preprint arXiv:1703.02000*.

Zhu, J.-Y.; Krähenbühl, P.; Shechtman, E.; and Efros, A. A. 2016. Generative visual manipulation on the natural image manifold. In *ECCV*, 597–613. Springer.

Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv preprint*.