Energy-guided Entropic Neural Optimal Transport
(Project proposal at Selected Topics of Data Science course)

Petr Mokrov
Skolkovo Institute of Science and Technology
Moscow, Russia
petr.mokrov@skoltech.ru

Abstract

Energy-Based Models (EBMs) are known in the Machine Learning community for the decades. Since the seminal works devoted to EBMs dating back to the noughties there have been appearing a lot of efficient methods which solve the generative modelling problem by means of energy potentials (unnormalized likelihood functions). In contrast, the realm of Optimal Transport (OT) and, in particular, neural OT solvers is much less explored and limited by few recent works (excluding WGAN based approaches which utilize OT as a loss function and do not model OT maps themselves). In our work, we bridge the gap between EBMs and Entropy-regularized OT. We present the novel methodology which allows utilizing the recent developments and technical improvements of the former in order to enrich the latter. We plan to validate the applicability of our method on toy 2D scenarios as well as standard unpaired image-to-image translation problems.

1 Introduction

Generative modeling is an emergent area of Machine Learning which has already presented unprecedented success in various applied spheres dealing with images, texts, sound, and videos. There are several problem setups arising in the context of generative modeling, and for most of them, there have been proposed machine learning techniques efficiently solving the corresponding matters. One can remember the StyleGANv3 model [3], the Dalle-E2 model [6], and the emerging ChatGPT dialogue system. In contrast to the cases above, the problem of unpaired data translation or, in particular, unpaired image-to-image translation is not yet fully explored and enjoys special attention in the machine learning community. In a few words, given two datasets of images from different domains, the unpaired image-to-image translation task is to map images from the source domain to the target domain, preserving the content of the corresponding image. The reference example exercises for evaluating unpaired image-to-image algorithms include the vangogh2photo and Yosemite summer2winter datasets [9].

2 Problem Statement and Related Works

The main tricky point with unpaired image-to-image translation comes from the ambiguity due to the unpaired nature of the problem. Given an image from the source domain, there is no ground truth counterpart image from the target domain which can be used for supervised training of a machine learning model. In order to leverage the difficulty, several methods have been proposed. The first group of methods [9], [2] is based on Generative Adversarial Networks equipped with specifically-designed cycle-consistency loss [9]. The data mappings learned by GAN-based models look plausible, but the complex adversarial objective with additional problem-specific loss intricates their application for real-world high-dimensional problems. Moreover, the theoretical justification of these methods is limited. The same problem with theoretical justification is presented for the second group of methods [7], [8] utilizing Energy Based Models setup. The idea is to recover the energy functions (unnormalized probability density functions) of the source and target domain distributions and map source images to the target domain using the stochastic analog of gradient ascent algorithm, known as Langevin dynamics (known also as ULA algorithm).

One of the most promising and emerging research direction in the sphere of unpaired image-to-image translation is connected with Optimal Transport theory (OT). The key and natural idea coming from the OT is to explicitly prescribe some cost function between images (i.e., a function that takes two data points as input and outputs how far these points are from each other) and look for a map from the source domain to target domain which minimizes the average of costs needed to translate source images to target images. The OT-based scheme was exploited in [5], [4]. Optimal Transport theory both provides solid theoretical background of the methods and mitigates the ambiguity of the unpaired image translation problem. However, the adversarial nature of the min-max objective arising in the Neural Optimal Transport solver [5] still limits the applicability of the approach, especially in the case of a nontrivial cost function.

3 Proposed method

In order to avoid the min-max optimization yet preserve the theoretical foundation, we propose to combine Energy Based models (EBMs) with Optimal Transport theory. Our method is based on Entropic regularized Optimal Transport formulation. Thanks to the Entropic regularization, the energy function we learn turns out to be the Lagrange multiplier of the target domain-constrained optimization problem. Compared to previously discussed approaches which operate with several neural networks simultaneously in a nontrivial way, our approach deals with a single neural network without auxiliary ones. Moreover, our proposed methodology allows utilizing the recent developments and technical improvements of EBMs in order to solve rather specific and hard Optimal Transport problem.

We validate our method empirically on toy 2D setups and plan to evaluate our algorithm's performance on large-scale image datasets. For 2D setups, we use discrete OT solvers for obtaining ground-truth solution, and for image datasets we plan to utilize FID metric to estimate the quality of our approach.

4 Experiments

In what follows, we demonstrate the performance of our method on toy 2D scenario and Colored MNIST images transformation benchmark [1]. Note, that our aim here is to show that the proposed Energy-guided EOT methodology actually works. We leave technically-saturated adjustment of our approach for practically-important large-scale applications for future work. Both in 2D and image cases the cost function is chosen to be squared halved l_2 norm: $c(x,y) = \frac{1}{2}\|x-y\|_2^2$.

Our code is written in PyTorch. The experiments are conducted on a single GTX 1080Ti. The actual neural network architectures as well as practical training setups are disclosed in the corresponding subsections.

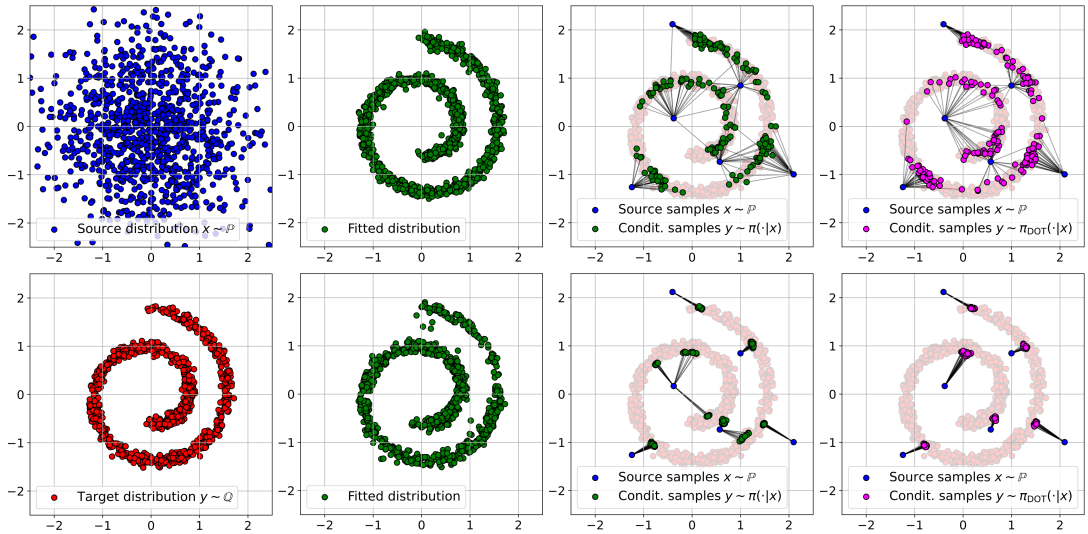
4.1 Toy 2D experiment

We apply our method for 2D *Gaussian→Swissroll* distributions modification task and demonstrate the qualitative results on Figure 1 for Entropy regularization coefficients $\varepsilon = 0.1, 0.001$. For this experiment, we parameterize the potential f_θ as MLP with two hidden layers and *LeakyReLU*$(\texttt{negative\_slope}= 0.2)$ as the activation function. Each hidden layer has 256 neurons.

Figure 1b shows that our method succeeds in transforming source distribution $\mathbb{P}$ to target distribution $\mathbb{Q}$ for both Entropy regularization coefficients. Note, that Figures 1c and 1d show, that our model learns conditional plans $\pi^*(y|x)$ similar to reference ones, given by discrete OT solver.

References

[1] Nikita Gushchin, Alexander Kolesov, Alexander Korotin, Dmitry Vetrov, and Evgeny Burnaev. Entropic neural optimal transport via diffusion processes. *arXiv preprint arXiv:2211.01156*, 2022.



(a) Input and target distributions $\mathbb{P}$ and $\mathbb{Q}$.

(b) **Our** fitted distributions; $\varepsilon = 10^{-1}$ (above), $\varepsilon = 10^{-3}$ (below).

(c) **Our** fitted conditional plans $\pi(\cdot|x)$; $\varepsilon = 10^{-1}$ (above), $\varepsilon = 10^{-3}$ (below)

(d) Discrete conditional plans $\pi_{\text{DOT}}(\cdot|x)$; $\varepsilon = 10^{-1}$ (above), $\varepsilon = 10^{-3}$ (below)

Figure 1: Performance of Energy-guided EOT on *Gaussian→Swissroll* 2D setup.

[2] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. Multimodal unsupervised image-to-image translation. In *Proceedings of the European conference on computer vision (ECCV)*, pages 172–189, 2018.

[3] Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Alias-free generative adversarial networks. *Advances in Neural Information Processing Systems*, 34:852–863, 2021.

[4] Alexander Korotin, Daniil Selikhanovych, and Evgeny Burnaev. Kernel neural optimal transport. In *The Eleventh International Conference on Learning Representations*, 2023.

[5] Alexander Korotin, Daniil Selikhanovych, and Evgeny Burnaev. Neural optimal transport. In *The Eleventh International Conference on Learning Representations*, 2023.

[6] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 2022.

[7] Yang Zhao and Changyou Chen. Unpaired image-to-image translation via latent energy transport. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16418–16427, 2021.

[8] Yang Zhao, Jianwen Xie, and Ping Li. Learning energy-based generative models via coarse-to-fine expanding and sampling. In *International Conference on Learning Representations*, 2021.

[9] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.