Assignment 2

Nikita Andreev, team name: happyhippogan

November 2024

1 Introduction

The task was to train a generative based model for image generation task for MNIST data. The structure of generator was fixed, while the discriminator was free to change. As the main purpose of this project is not the scoring best possible metric, but to utilize some additional approaches and techniques for GANs. In my work I would like to use the Unbalanced Optimal Transport Modeling which is a recent trend in generative modeling. More precisely, I am using the paper "Generative modeling through the semi-dual formulation of unbalanced optimal transport" (https://arxiv.org/abs/2305.14777).

2 Main approach

Basically, through optimal transport we could perform the unconditional generation of images. For the dataset, we learn its distribution and can produce the simillar images. Unbalanced Optimal Transport described in the paper is an extension of the original dual formulation of the optimal transport problem. It is formulated as

$$Cost_{ub}(\mathbb{P},\mathbb{Q}) = \inf_{\pi \in \Pi(\mathbb{P},\mathbb{Q})} \left[\int_{\mathcal{X}\times\mathcal{Y}} c(x,y)d\pi(x,y) + D_{\Psi_1}(\pi_0|\mathbb{P}) + D_{\Psi_2}(\pi_1|\mathbb{Q}) \right]$$

where $D_f(a|b)$ stands for the Csiszar divergence associated with f is a generalization of f-divergence for the case where a is not absolutely continuous with respect to b. c - some cost function, for example MSE.

Precisely, similarly to classic optimal transport, the semi-dual form of the **UOT problem** can be reduced to the following objective:

$$\inf_{f_w} \left[\int_{\mathcal{X}} \Psi_1^* \left(-\inf_{T_\theta} \left[c\left(x, T_\theta(x)\right) - f_w\left(T_\theta(x)\right) \right] \right) d\mathbb{P}(x) + \int_{\mathcal{Y}} \Psi_2^* \left(-f_w(y) \right) d\mathbb{Q}(y) \right]$$

where Ψ_1^* and Ψ_2^* are some functions, for example $softplus$, $\exp(x)$ or $\exp(x)-1$, and T (Generator) is a transport map, f (discriminator) is a potential. However, in general Ψ^* should be a non-decreasing, differentiable, convex function.

Authors not only suggest new formulation, but also utilize some techniques from GAN training strategies such as regularization (which is gradient penalty, learning schedule). And generally, the usage of the divergences is a way of regularizing the objective for both distributions.

The regularization loss for potential (discriminator) was used in the form of:

$$L_{reg} = \lambda ||\nabla_x f_w(y)||_2^2$$

3 Suggested improvements

Here, I would like to list the amount of improvements which I tried in my experiments:

- Usage of UOTM formulation
- gradient regularization
- changed discriminator architecture (removed last activation + changed the activation function)
- tried several Ψ^* options
- tried Exponential Moving Average for the generator (the common tools for weight averaging for generative models)

4 Results of experiments

Due to some time and computational limitations I have only tried several setups of my approaches, but still can do some conclusions.

Method	**FID score**
GAN baseline	29.78
UOT baseline (60 epochs)	35.6
UOT baseline (100 epochs)	46.8
UOT(100 epochs) + Ψ^**=exp*	37.1
UOT + Ψ^**=exp + EMA (50 epochs)*	143
UOT + Ψ^**=exp + EMA (90 epochs)*	136

Table 1: The results of experiments.

Basically, the results are not great. I did not get the improvement with the usage of UOT approach. I suppose that there are 2 key things here. First, the amount of hyperparameters of the model which we need to select and this is extremely difficult with restricted computational resources. Secondly, the models itself are pretty weak as we only use linear layers, so the usage of the unbalanced ot approach is not effective because the model itself (generator) can not learn much information and the distribution of data. So, perhaps if we use convolutional architectures of the models the results would be much better but here we are limited and the model just is not able to learn the data well enough.

Moreover, I tried the EMA technique which is well-known but its effect was negative which is quite strange because for GANs it is usually helpful. So, I think it is again because of the architecture of the models and their limitations.

5 Conclusion

As I worked alone, I did not try more approaches, but I think that my approach is unique and the study that was conducted is quite useful as we see that even strong methods (such as UOT) for weak architectures could be non-effective because of the limitations in architectures. One can try to get as much as possible from it, but when scaling to real world tasks and traditional architectures the results won't be the same so it is better to conduct a study with the real approaches which were approved on other datasets but not trying to hit the best score with strong limitations.