Improving GAN by neighbor embedding

Tuan-Anh Bui

August 16, 2018

Introduction

Generative Adversarial Networks (GANs) have shown their powerful in not only generative problems [6] but also in many other fields, such as, domain transfer [1] [2], image super resolution [3], even on several fields that are seem irrelevant, such as, physics [4], steganography [5] etc. In spite of its potential, GANs still suffer several drawbacks, that are mode collapse, unstable in training and unrealistic output, especially with high-dimensional data.

In this project, I would like to propose a method that efficiently solves the mode collapse problem. This method is inspired from two works, the first one is the discovery from the paper [7] about the relationship between the mode collapse problem and the distance constraint between latent space and data space. And the second idea from t-sne visualization method [9] that can help to maintain the distribution between high-dimensional space and low-dimensional space. A preliminary experimental result is also provided to show the feasibility of my approach.

Contribution to the Discipline

Meanwhiles, GANs have shown their impressive performance for learning generative models of arbitrarily complex data distributions. The reason behind the success of GANs is that GAN can implicitly learn the hidden distribution of data p(x) without any effort or knowledge about that field even image, text and speech. GANs introduce a minimax game between the generator and discriminator which try to reach two opposite aims. Specifically, generator tries to generate a more realistic sample but discriminator tries to distinguish which sample is real or fake. By virtue of that strategy, the generator is able to learn and improve the quality of the output. However, despite having a smart strategy, GANs still have some critical weaknesses. The first one is unstable in training, for instance, GANs objective loss does not guarantee that generator can achieve the global optimal due to employing *log* function, which actually leads to the unstable training problem of GANs. With the aim of overcoming that challenge, the recent works [10] [11] replace the KL divergence objective with the new Earth Move distance that can efficiently reduce the instability. The second problem is to generate realistic samples, especially in high resolution image. The work suggested in [6] successfully handles with that problem by introducing a progressive framework that learn from low-level of resolution to higher resolution. Nevertheless, the last and most serious problem, which is mode collapse, has yet had an effective solution.

The main contribution of this problem is to propose a novel approach that can efficiently deal with the mode collapse problem. My strong belief is that if the mode collapse is solved properly, GANs can generate more diversity space, which could possibly achieve a large impact on a wide range of applications.

Theoretical Framework and Methods

Mode collapse from latent point of view

Mode collapse is easily observed in data space as the generator often generate similar shapes or objects, that made the generated space is less diversity than the original dataset. However, that view in data space is still not sufficient for understanding the cause of mode collapse. The research in [7] is the first work that considers the mode collapse problem in the other point of view, e.g, latent space. Concretely, by leveraging the independent classifier, the authors map each class index of generated images $G(z)$ to each prior sample z in 2D latent space. That idea is inspired by how human performs the classification task, which is assigning a specific label to an arbitrary sample. By this way, the more collapse model can be distinguished by the less balance probability of each class. The more important view is that if there are a large enough numbers of samples (as Monte Carlo test) the boundary of each class can be seen as the corresponding area in latent space. Hence, this view will lead to an important mismatch problem, which is two far data points in latent space is likely projected to similar data points in data space. This hypothesis motivates me to an idea whether we can control the relative distance among data points and their corresponding points within both data space and latent space. However, this idea still requires much effort for further investigation.

Visualizing high dimensional space

t-sne is the famous method for data visualization that works well for not only a specific dataset but various type of the high-dimensional datasets. This method stands out other visualization methods such as Principal Components Analysis (PCA) [12], Multidimensional Scaling (MDS) [13], Stochastic Neighbor Embedding (SNE) [14] because it can not only preserve the local structure of data space (the relation inside each natural cluster) but also the global structure of data space (the relation between each natural cluster). The method's objective is forcing the similarity between the distribution of data points in high-dimensional space and their projected data points in low-dimensional space. This idea inspired me about how to keep the relationship between data space and latent space in GANs problem.

Improving Gan by neighbor embedding

The project's idea and t-sne method share a similar goal. If t-sne method tries to preserve as much of the significant structure of the high-dimensional data as possible in the low-dimensional map, my idea is trying to preserve the same relationship between the source space and the target projected space. This projection can be both scenarios, from high-dimensional space to low-dimensional space in case projector is the Generator, and from low-dimensional space to high-dimensional space in case of Encoder. Similar the approach in t-sne, this project's aim can be done by add the regularization term called "neighbor embedding" to objective loss of generator in GANs as below:

$$\mathcal{L}_\theta(G) = \mathcal{L}^o_\theta(G) + \mathcal{L}_\theta(S) \tag{1}$$

Where $\mathcal{L}^o_\theta(G)$ is the ordinary generator objective loss in GANs model such as $\mathcal{L}^o_\theta(G) = E_z(log(1-\sigma(D(G(z)))))$. The regularization term $\mathcal{L}_\theta(S)$ means the dissimilarity between two joint distributions p_{ij}, q_{ij} where each distribution represents the neighbor distance distribution as in [9]. The dissimilarity measure by the Kullback-Leibler divergence as equation 2.

$$\mathcal{L}_\theta(S) = KL(P||Q) = \sum_i \sum_j p_{i,j} \log \frac{p_{i,j}}{q_{i,j}} \tag{2}$$

$p_{i,j}$ is joint, symmetric distribution of source space (which is latent space in case projector is the Generator). $q_{i,j}$ is joint, symmetric distribution of target space (which is data space in case projector is the Generator). The most different to t-sne is that t-distribution was represented for both source distribution and data distribution to slow the generator convergence by smaller gradient.

$$p_{ij} = \frac{(1+||z_i - z_j||^2)^{-1}}{\sum_{k\neq l}(1+||z_k - z_l||^2)^{-1}} \tag{3}$$

$$q_{ij} = \frac{(1+||G(z_i) - G(z_j)||^2)^{-1}}{\sum_{k\neq l}(1+||G(z_k) - G(z_l)||^2)^{-1}} \tag{4}$$

The two noticeable improvements of this method to the most similar method as in [7] are that firstly, this method does not require an auto-encoder, therefore, can avoid the weaknesses of autoencoder, not only that, this method can apply to both GANs and AutoEncoder GANs architectures such as GAAN, VAEGAN [8]. Secondly, this method takes advantage of t-sne method to preserve the relationship between two spaces more efficiently and in more make sense way.

Baseline result

I have executed some preliminarily experiments to test the feasibility of my method. The result can see in table 1 where GAAN-NE and GAN-NE mean for my method apply to both autoencoder (GAAN) and GAN. This result seem promising when improve not only the state of the art model GAAN but also improve the performance of GANs.

References

[1] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros. Image-to-Image Translation with Conditional Adversarial Networks, https://arxiv.org/abs/1611.07004.

[2] Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros. Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. https://arxiv.org/abs/1703.10593.

Table 1: Comparing FID score to the state of the art (Smaller is better). Methods with $*^1$, $*^2$ indicate that we follow the experimental setup of [cite] and [cite] receptively. h if using hinge loss

	CelebA	Cifar-10	Cifar-10	Cifar-10 (h)	STL-10 (h)
NS GAN	58.0	58.6	-	-	-
GAN-GP	-	-	37.7	-	-
LSGAN	53.6	67.1	-	-	-
WGAN-GP	26.8	52.9	40.2	55.1	-
BEGAN	38.1	71.4	-	-	-
VAEGAN	27.5	58.1	-	-	-
SN-GAN	-	-	29.3	25.5	47.5
GAAN	23.7	45.6	28.23	22.95	36.1
GAAN-NE	-	-	27.1	22.3	32.0
GAN-NE	-	-	27.0	21.0	31.8

[3] Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, Wenzhe Shi. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, https://arxiv.org/abs/1609.04802.

[4] Luke de Oliveira, Michela Paganini, Benjamin Nachman. Learning Particle Physics by Example: Location-Aware Generative Adversarial Networks for Physics Synthesis https://arxiv.org/abs/1701.05927

[5] Denis Volkhonskiy, Ivan Nazarov, Boris Borisenko, Evgeny Burnaev. Steganographic Generative Adversarial Networks. https://arxiv.org/abs/1703.05502.

[6] Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen. Progressive Growing of GANs for Improved Quality, Stability, and Variation, https://arxiv.org/abs/1710.10196.

[7] Ngoc-Trung Tran, Tuan-Anh Bui, Ngai-Man Cheung. Generative Adversarial Autoencoder Networks. https://arxiv.org/abs/1803.08887

[8] Anders Boesen Lindbo Larsen, Sren Kaae Snderby, Hugo Larochelle, Ole Winther. Autoencoding beyond pixels using a learned similarity metric. https://arxiv.org/abs/1512.09300

[9] Laurens van der Maaten, Geoffrey Hinton. Visualizing Data using t-SNE. Journal of Machine Learning Research, 2008.

[10] Martin Arjovsky, Soumith Chintala, Lon Bottou. Wasserstein GAN. https://arxiv.org/abs/1701.07875.

[11] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, Aaron Courville. Improved Training of Wasserstein GANs. https://arxiv.org/abs/1704.00028.

[12] H. Hotelling. Analysis of a complex of statistical variables into principal components. Journal of Educational Psychology, 1933.

[13] W.S. Torgerson. Multidimensional scaling I: Theory and method. Psychometrika, 1952.

[14] G.E. Hinton and S.T. Roweis. Stochastic Neighbor Embedding. In Advances in Neural Information Processing Systems, volume 15, pages 833840, Cambridge, MA, USA, 2002. The MIT Press.