Improving GAN by neighbor embedding

Tuan-Anh Bui

July 29, 2018

Introduction

Generative Adversarial Networks (GANs) have shown these powerful in not only generative problems [6] but also in many other fields such as domain transfer [1] [2], image super resolution [3], even some fields that seem completely irrelevant such as physics [4], steganography [5] etc. Although it's powerful, GAN still has several open questions, cristical weekeness for further improvement that is mode collapse, unstable when training and unrealistic output especially with high resolution images.

In this project, I would like to propose a method that efficiently solving mode collapse problem. This method 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. I also provide some quick experiments to show the possibility of this method.

Contribution to the Discipline

Introduction about gan contribution: GANs have shown the huge impact to research not only in files of images processing but other fields such as text, speech. The reason behind successful 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, speech. GANs introduce a minimax game between the generator and discriminator which those try to get two opposite aims, generator tries to generate a more realistic sample but discriminator try to distinct which sample is real or fake. By this strategy, the generator can learn and improve the quality of output. But although having a smart strategy, GANs still have some critical weakness. The first one is unstable in training that the GANs objective does not guarantee that generator can achieve the global optimal, actually the log objective function also is source of unstable problem of GAN. Fortunately, the recent work [10] [11] by replacing the KL divergence objective with the new Earth Move distance that can efficiently reduce the instability. The second problem is the realistic of generated samples, especially in high resolution image. The recent work [6] also successful handle with this problem by introducing a progressive framework that learn from low-level of resolution to higher resolution. But for the last and the most serious problem - mode collapse, currently have yet a effective solution.

The main contribution of this project is to propose a new method that can effectively handle the mode collapse problem. If successful, this method can help not only the GANs but any applications of GANs can generate a more diversity space, that I think will have a large impact on other research.

Theoretical Framework and Methods

Mode collapse from latent point of view

Mode collapse can be easily seen in data space as the generator can generate not all but some similar shapes or objects. It can be understood that the diversity of generated dataset has been modified which less diversity than the original dataset, it also can be seen as entropy $H(x|z)$ has smaller than H(x).

But this view in data space is not useful for understanding why mode collapse happened. The paper [7] is the first method that view mode collapse in another point of view: latent space. By using the independent classifier, the authors mapping each class index of generated images $G(z)$ to each prior sample z in 2D latent space. That can be understood as a human tries to classy each class that each sample belong to. And the result is that the more collapse model is which have more samples in one class than other class (or actually more than the probability of this class in training dataset). The more important view is that if have large enough samples (as Monte Carlo test) the boundary of each class can be seen as the corresponding area in latent space. Then this view will lead to an important mismatch problem is that two far data points in latent space but projected to similar (or near in some metrics) data points in data space. This point led me to an idea about that somehow if we can keep the relative between the distance between data points in data space and the distance of their corresponding data points in latent space, we can handle the mode collapse problem. But how?

Visualizing high dimensional space

t-sne is the well-known method for data visualization into 2D or 3D map. This method stand 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 (relation inside each natural cluster) but also the global structure of data space (relation between each natural cluster). This method can be done by keeping 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 GAN problem.

Improving Gan by neighbor embedding

My idea and t-sne method sharing 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-dimensional space (it can be both low-dimensional space as in Generator and high-dimensional space as in Encoder) and target-dimensional space (high if Generator, low if Encoder). Therefore, I just add the term called "neighbor embedding regularization" to objective loss of generator in GANs as below:

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

Where $L^o_\theta(G)$ is the normal generator objective loss such as $L^o_\theta(G) = E_z(log(1-\sigma(D(G(z)))))$. The regularization term $L_\theta(S)$ meaning the similarity between two joint distribution p_{ij}, q_{ij} where each distribution meaning the neighbor distance distribution as in [9]. We will minimize the Kullback-Leibler divergence as equation 2 to force the generator to generate the samples that satisfied the distance constraints.

$$L_\theta(S) = KL(p_{ij}||q_{ij}) \tag{2}$$

$p_{i,j}$ is joint, symmetric distribution of source space (in case Generator is latent space). $q_{i,j}$ is joint, symmetric distribution of target space (in case Generator is data space). One different between t-sne is that I use both t-distribution on both source space and data space to slow the generator by smaller gradient when t-distribution instead of Gaussian distribution as t-sne.

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

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

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

Baseline result

I have done some quick experiemnts to test the posibility 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.

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, and R if using ResNet model.

	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

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.

[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.