GAN with Autoencoder and Importance Sampling

Gahye Lee
Department of Computer Science and Engineering
Kyunghee University, Republic of Korea
waldstein94@gmail.com

Seungkyu Lee
Department of Computer Science and Engineering
Kyunghee University, Republic of Korea
seungkyu@khu.ac.kr

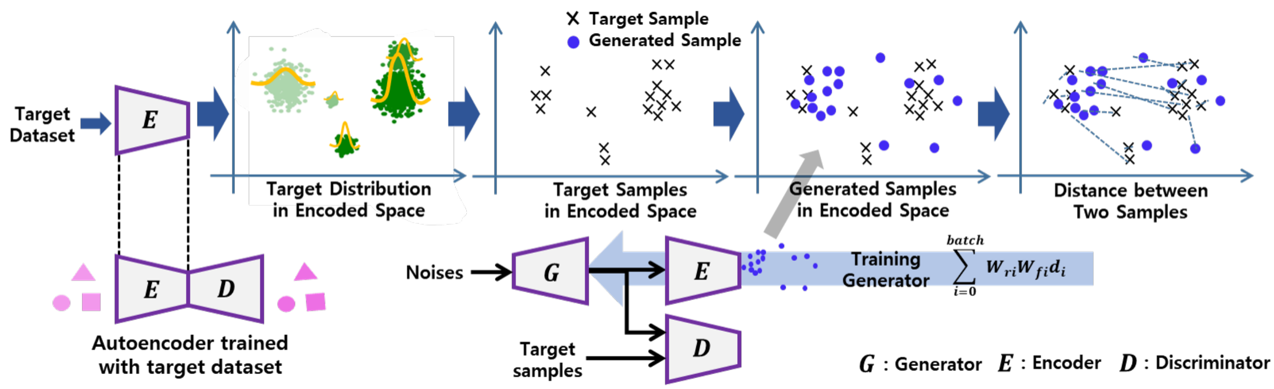

Figure 1: Proposed network and training steps: Simple 2-dimensional data distribution is used to visualize training steps. Target and generated samples are represented by cross and solid circle respectively.

ABSTRACT

Deep generative model such as generative adversarial networks (GAN) has shown impressive achievements in computer graphics applications. GAN is trained to learn the distribution of target data and is able to generate new samples similar to the original target data. However, most GAN based networks encounter mode collapse problem resulting in the generation of samples only from a single or a few modes of target data distribution. In order to address mode collapse problem, we propose to adopt autoencoder to learn target data distribution in encoded space. An importance sampling scheme is used to collect fake and real data samples in the encoded latent space and calculate the similarity of two distributions in real data space. Experimental evaluation compared to state-of-the-art method on synthetic and MNIST datasets shows the potential of our approach in reducing mode collapse problem and generating samples from diverse aspect of target data.

CCS CONCEPTS

• **Computing methodologies** → **deep generative model**; **GANs**;

KEYWORDS

deep generative model, generative adversarial nets(GAN), mode collapse, autoencoder

SA '18 Posters, December 04-07, 2018, Tokyo, Japan

ACM ISBN 978-1-4503-6063-0/18/12.
https://doi.org/10.1145/3283289.3283333

ACM Reference format:
Gahye Lee and Seungkyu Lee. 2018. GAN with Autoencoder and Importance Sampling. In *Proceedings of SA '18 Posters, Tokyo, Japan, December 04-07, 2018*, 2 pages.
https://doi.org/10.1145/3283289.3283333

1 INTRODUCTION

Recently, deep generative model such as generative adversarial networks (GAN) [Goodfellow et al. 2014] has been successfully applied in diverse computer graphics methods such as human face generation, 3D objects and stylized image synthesis, etc. GAN is composed of two deep networks: generative and discriminative networks. Discriminator helps generator capture the true data distribution by learning the dissimilarity between generated samples and true samples iteratively. On the other hand, adversarial structure of GAN suffers fro.m mode collapse problem resulting in the generation of samples only from a single or a few modes of target data distribution. This is because the generator is trained to resemble certain instance rather than entire data distribution. During the training step, if generator starts to make any good sample from one mode of target data, discriminator stops to drive the generator learn target data. Therefore, it fails to recover the entire distribution of target data. Many researchers have tried to address the problem. Unrolled GAN [Metz et al. 2016] proposes unrolling optimization of the discriminator objective to train generator and VEEGAN [Srivastava et al. 2017] employs a reconstruction network to map data distribution to a Gaussian for estimating implicit probability. In this paper, we propose an improved GAN consists of generator, discriminator and autoencoder. The main idea of our approach is that autoencoder trained with target real dataset is able to represent the distribution of real dataset in the encoded latent space. We train our generator to produce fake samples resembling the distribution of true dataset by minimizing weighted distance between paired fake and real samples in encoding space. In this step an importance sampling scheme is used to collect fake and real data samples in the encoded latent space.

2 PROPOSED METHOD

In general, high dimensional target data distribution cannot be explicitly represented. In order to represent target data distribution in encoded space, we attach another network E that is pre-trained autoencoder with the target data. This pre-trained Encoder is able to provide explicit low dimensional representation of target data preserving important features of from modes of real target data. Figure1 shows overall framework of our network and training steps.

$$\min_{G} -\mathbb{E}_{z\sim P_z} \log D(G(z)) + \frac{1}{n_{batch}} \sum_{i=1}^{batch} W_{ir} W_{if} ||X_{ir} - X_{if}||^2 W_l \tag{1}$$

We expand the generator loss of original GAN adding second term in (1) of encoder. With all samples of true dataset, its distribution is estimated using KDE (Kernel Density Estimation). And then, we obtain good samples of number of batch size representing true data distribution. X_{ir} denotes i-th real sample in encoded space. Likewise, X_{if} denotes i-th generated sample produced by generator in encoded space. For optimal generated and real sample pair matching in distance calculation, we calculate distances among all real samples to all generated samples and assign pair of X_{ir} and X_{if} one by one minimizing average distance. W_{ir} and W_{if} are two types of importance weights. W_{ir} indicates how much real sample is well described by generated samples over past N training iterations. W_{if} indicates how much current generated sample is well represent the distribution of data that current generator create. W_{if} is also estimated from accumulated generated samples of past N training iterations. W_l is weight adjusting the contribution of two losses: original GAN loss and added encoder loss, first and second terms in equation (1) respectively. Consequently, our objective function helps generator detect missing modes and achieves better convergence to whole distribution.

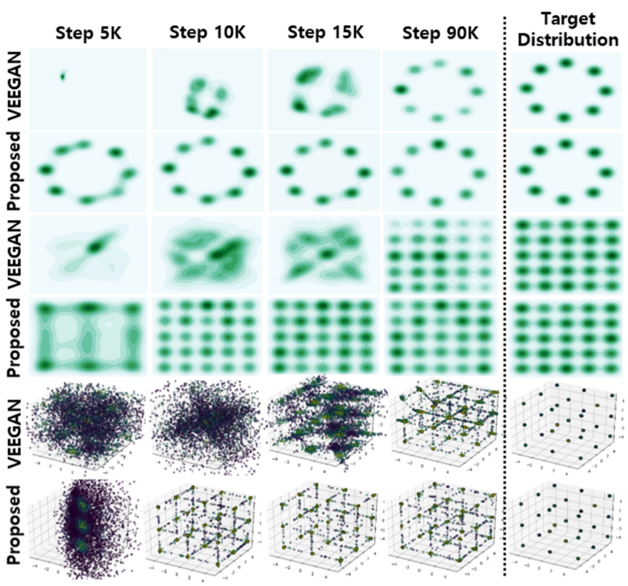

Figure 2: Experimental results compared to VEEGAN method: 2D Ring, 2D Grid, and 3D Cube testset. Proposed method converges faster than the state-of-the-art method.

Table 1: Quantitative Evaluation: Number of modes found, HQS(High Quality Sample), and JSD(Jensen-Shannon divergence) between real and generated sample distributions

	METRIC	VEEGAN(std)	Proposed(std)
2D Ring	Modes(Max 8)	8(0)	8(0)
	% HQS	60.4(0.005)	**85.5**(0.006)
	JSD(real‖fake)	0.19(0.005)	**0.172**(0.004)
2D Grid	Modes(Max 25)	24.1(0.31)	**25**(0)
	% HQS	65.4(0.006)	**82.5**(0.004)
	JSD(real‖fake)	0.21(0.004)	**0.12**(0.004)
3D Cube	Modes(Max 27)	26.6(0.5)	**27**(0)
	% HQS	43.3(0.007)	**80.0**(0.005)
	JSD(real‖fake)	0.31(0.005)	**0.125**(0.004)

3 EXPERIMENTS

We compare our network with original GAN and VEEGAN [Srivastava et al. 2017]. For both qualitative and quantitative evaluation, we use mixture of 8 2D Gaussians located in a ring, 25 2D Gaussians located in a grid and 27 3D Gaussians building a cube. For quantitative evaluation, we employ three metrics: (1) Number of modes found, (2) HQS (High quality sample), and (3) JSD (Jensen-Shannon divergence).

REFERENCES

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2014. Generative adversarial nets. In *Advances in neural information processing systems*. 2672–2680.

Luke Metz, Ben Poole, David Pfau, and Jascha Sohl-Dickstein. 2016. Unrolled generative adversarial networks. *arXiv preprint arXiv:1611.02163* (2016).

Akash Srivastava, Lazar Valkoz, Chris Russell, Michael U Gutmann, and Charles Sutton. 2017. Veegan: Reducing mode collapse in gans using implicit variational learning. In *Advances in Neural Information Processing Systems*. 3308–3318.

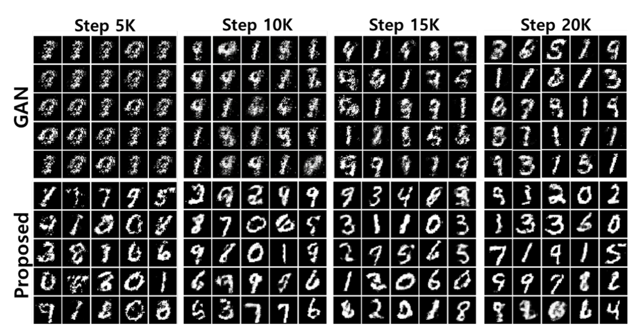

Figure 3: Experimental results on MNIST dataset. First row is original GAN and second row is proposed method.