An Overview and Comparative Analysis on Major Generative Models

Zijing Gu
zig021@ucsd.edu

Abstract

The amount of researches on generative models has been grown rapidly after a period of silence due to insufficient data. The model itself tends to assume data distributions that are too ideal for real world, or, to use Bayesian probability which is too slow to train. However, the proposal of Generative Adversarial Nets (GAN) put generative models under spotlight. This paper will present a brief overview on several important generative models including VAE, WGAN-GP, WINN, and DCGAN. These models will be trained on typical generative modeling datasets including CIFAR-10 and CelebA, and the results will be compared qualitatively and quantitatively.

1 Introduction

The Deep Generative model recently attracts lots of attention due to the promising progress achieved. Both Variational Auto-Encoder (VAE) [Kingma and Welling, 2013] and Generative Adversarial Networks [Goodfellow et al., 2014] can learn to generate realistic images. VAE consists of an encoder and a generator cooperatively to approximate the complex data distribution. However, the model puts oversimplified parametric assumptions on generated conditional data distribution, leading to the consequence of blurry images. GAN overcomes this problem by removing the parametric assumptions and instead using a powerful neural network to dynamically discriminate the differences between real images and the generated ones. Recently GAN has achieved lots of successes and can produce realistic images efficiently without blurs. However, the elegant framework of GAN also suffers problems such as mode collapse and unstable training procedures. The improved work has been proposed to target those disadvantages.

In order to stabilize training procedure, DCGAN [Radford et al., 2015] takes advantage of a set of constraints on the architectural topology of Convolutional GANs, which achieves the goal in most settings. Wasserstein GAN (WGAN) [Arjovsky et al., 2017] leverages the Wasserstein distance to produce a value function which has better theoretical properties than the original. However, WGAN requires that the discriminator must lie within the space of 1-Lipschitz functions through which the authors enforce the weight clipping. The applied weight clipping leads to undesirable behaviors, e,g., most values only lie in the clipping-limited area. WGAN-GP [Gulrajani et al., 2017] uses gradient penalty to overcome the weight-clipping problem and achieves better performances. Although WGAN helps stabilize GAN's training by changing the distance function and brought better theoretical properties, the original iterative-update-framework applied in GAN is still one of the causes of the unstable training procedure. WINN [Lee et al., 2017] does not only leverage the Wasserstein distance to stabilize training procedure but also innovatively unifies the discriminator and generator contained in original GAN into one. Although this architecture does not seem to have a much better performance compared with GAN, it does provide a novel perspective in exploring more elegant and stable architectures for the adversarial generative framework.

32nd Conference on Neural Information Processing Systems (NIPS 2018), Montréal, Canada.

2 Method

In this section, I will summarize the methodologies of several famous generative models including GAN, DCGAN, WGAN, VAE and WINN.

2.1 GANs

The original GAN framework [Goodfellow et al., 2014] contains two neural networks: one generator and one discriminator. The generator tries to generate fake images from random noise and fool the discriminator. The discriminator is trained to distinguish the fake images from the true samples. In this adversarial manner, the generator will ideally learn the true generating distribution. An min-max game is played by GAN with value function:

$$\min_G \max_D V(G,D) = \mathbb{E}_{p_{data}(x)}[\log(D(x))] + \mathbb{E}_{p_z(z)}[1 - \log(D(G(z)))]. \tag{1}$$

However, the training process of GAN is not stable because of the min-max game and gradient vanishing problem. DCGAN [Radford et al., 2015] stabilizes the training process of GAN and improves the fidelity of generated images by designing a deep convolutional neural network structure.

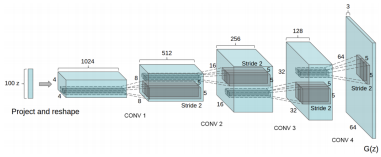


Figure 1: DCGAN structure.

WGAN [Arjovsky et al., 2017] designs a new objective which is much easier to train and enjoys Lipschitz property that makes the gradient stable. Gulrajani et al. [2017] further improved upon WGAN by introducing penalty on the gradients.

2.2 Introspective Neural Networks

Despite the different network structures and distance metrics, all these GAN models share the same adversarial regime. However, a recent method WINN [Lee et al., 2017] proposed a totally different regime in which introspective neural networks are used. The innovation of this network is that it is both a generator and a discriminator. It improves upon the previous proposed INN [Lazarow et al., 2017] method by incorporating a Wasserstein distance. In this way, a single CNN can achieve good generating performance with 20 times reduction in model size.

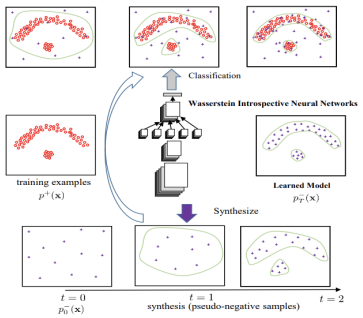


Figure 2: WINN structure.

The underlying theory is Bayes rule

$$p(x|y=1) = \frac{p(y=1|x)}{p(y=-1|x)} p(x|y=-1) \tag{2}$$

when $p(y=1|x) = p(y=-1|x)$. We can then iteratively update the posterior of the negative class by

$$p_t^-(x) = \frac{1}{Z_t} \frac{q_t(y=1|x)}{q_t(y=-1|x)} p_{t-1}^-(x). \tag{3}$$

Table 1: Inception scores (higher is better) on CIFAR-10. Results are from [Lee et al., 2017]

Method	score
Real data	$11.95 \pm .20$
WGAN	$5.88 \pm .07$
WGAN-GP	$7.86 \pm .07$
DCGAN	$6.16 \pm .07$
WINN-single	$4.62 \pm .05$

The conditional q_t can be estimated by a CNN classifier which classifies true samples from fake images generated from $p^-_{t-1}(x)$. As $t \to \infty$, q_t will be more accurate and so is the posterior $p^-_t(x)$, which will approaches $p(x|y=1)$.

2.3 VAE

Different from GAN framework mentioned above, VAE has two procedures: inference and generation. They cooperated with each other to improve the generation performance simultaneously. It also uses the pre-defined metric which is different with GAN, to evaluate the difference among fake samples and real samples with normal assumption on condition generated data distribution. Compared with GAN framework, it benefits from inference mechanism which makes pairwise comparison possible. Therefore, it will not suffer from mode collapse problem. However, the over-simplified assumption inevitably leads to inaccurate estimation of difference, as a consequence, the blurred images.

3 Experiment

3.1 Quantitative Results

The quantitative comparison is shown in the Table 1. As it can be seen, WGAN-GP achieves the best report compared with other models, which illustrates the effectiveness of Wasserstein distance function and gradient penalty. The effectiveness of the gradient penalty can also be further supported by the comparison between WGAN and WGAN-GP. DCGAN also achieves the competitive performance which can be attributed to the carefully designed neural network architecture. Although the WINN-single which uses the simple version of WINN setting does not provide the best result, it can guarantee the reasonable performance with the over-designed architecture by unifying the discriminator and generator. The robustness brought by WINN is valuable and the idea of model framework can inspire other researchers to further bring more possibilities in structural exploration.

3.2 Qualitative Results

In this subsection, I will further discuss the generated samples produced from the models mentioned above on CIFAR-10 and CelebA to explore the unique designs of them.

The generated samples by WGAN-GP is shown in Fig 3. As can be seen, the WGAN-GP provides very realistic generated samples. By changing the distance function and incorporating gradient penalty, it can provide robust and stable realistic results which can also supported by the Inception score shown in the quantitative subsection. Compared with results of other two models, we can observe the generated samples on CelebA from WGAN-GP are more realistic and show more diversities in the image patterns.

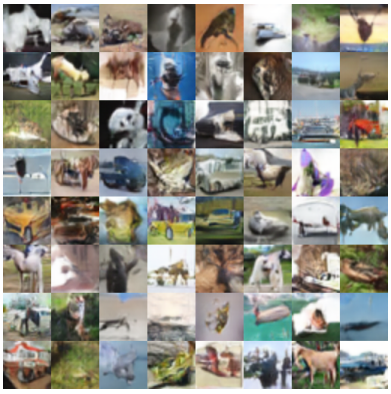
DCGAN provides the carefully-designed architecture to stabilize the training procedure and achieves the competitive results at the same time. From the generated samples shown in Fig 4, it can be seen that DCGAN also provides the realistic images even though some samples are not realistic enough. Although the WGAN-GP and DCGAN both takes advantage of different techniques (distance function and architecture) to stabilize training, GAN framework still suffers the unstable and convergence problem caused by iterative-update-framework. Since the time is limited and the model complexity is large, I only show the results with one cascade. That's why the results in the Fig 5 are not as good as other models'. However, the minimax problem cannot be directly solved by current gradient descent method and iterative-update regime. The WINN provide a very novel idea to unify the discriminator and generator into one, which effectively help solve the problem along with original design of GAN.

(a) CIFAR10

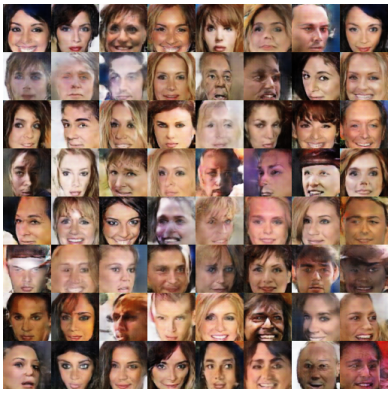
(b) CelebA

Figure 3: WGAN-GP on CIFAR10 and CelebA.

Although the performance of WINN is still limited, this provides very valuable explorations on the basic problem in the GAN research area.

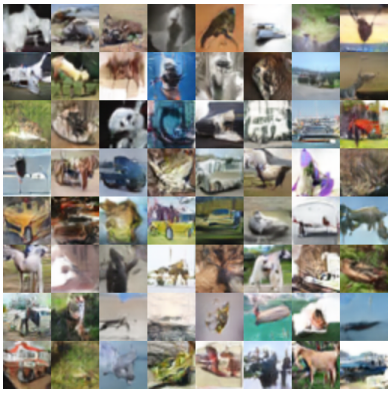


(a) CIFAR10

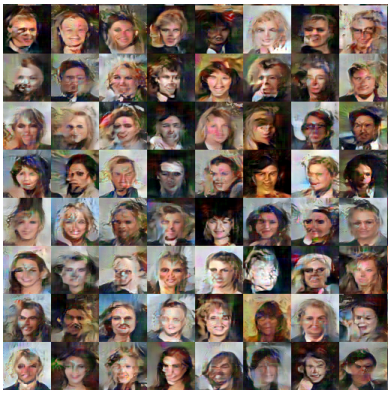


(b) CelebA

Figure 4: DCGAN on CIFAR10 and CelebA.



(a) CIFAR10



(b) CelebA

Figure 5: WINN on CIFAR10 and CelebA.

4 Conclusion

Generative models, being able to generate new data through learning the distribution of training data, have gained more research values during recent years. Efforts to create and refine models that

generate clearer and more realistic images have been made massively. This paper aims to compare four of the important generative models: VAE, WGAN-GP, WINN, and DCGAN both visually and statistically using two image datasets, CIFAR-10 and CelebA. The main ideas of these models, their deficiencies as well as their improvements on one another are briefly addressed to demonstrate general connections between them. The four models are trained individually on two datasets and the generated images are compared side by side. This comparison gives an overview of the different properties of each model and showcases their relationship. It also gives an idea of where we are at in image generating and what we can potentially do to improve our current results.

Acknowledgments

I would like to especially thank my friend Hanbo Li and Yaqing Wang for all the discussions, help and support on this project. I would also like to thank the entire COGS 185 faculty including Professor Zhuowen Tu for introducing this topic and all the related materials.

References

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan. *arXiv preprint arXiv:1701.07875*, 2017.

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

Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. In *Advances in Neural Information Processing Systems*, pages 5769–5779, 2017.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Justin Lazarow, Long Jin, and Zhuowen Tu. Introspective neural networks for generative modeling. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2774–2783, 2017.

Kwonjoon Lee, Weijian Xu, Fan Fan, and Zhuowen Tu. Wasserstein introspective neural networks. *arXiv preprint arXiv:1711.08875*, 2017.

Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.