A MULTI-PLAYER MINIMAX GAME FOR GENERATIVE ADVERSARIAL NETWORKS

Ying Jin, Yunbo Wang, Mingsheng Long (✉), Jianmin Wang, Philip S. Yu, and Jiaguang Sun

School of Software, BNRist, Tsinghua University, China
Research Center for Big Data, Tsinghua University, China
Beijing Key Laboratory for Industrial Big Data System and Application
{jiny18,wangyb15}@mails.tsinghua.edu.cn, {mingsheng,jimwang,psyu,sunjg}@tsinghua.edu.cn

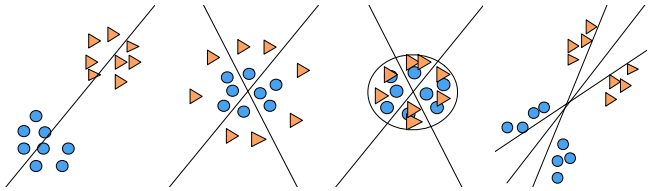
ABSTRACT

While multi-discriminators have been recently exploited to enhance the discriminability and diversity of Generative Adversarial Networks (GANs), these *independent* discriminators may not collaborate harmoniously to learn diverse and complementary decision boundaries. This paper extends the original two-player adversarial game of GANs by introducing a new *multi-player* objective named *Discriminator Discrepancy Loss (DDL)* for diversifying the multi-discriminators. Besides the competition between the generator and each discriminator, there are also competitions between the discriminators: **1)** When training multi-discriminators, we simultaneously minimize the original GAN loss and **maximize** DDL, seeking a good trade-off between the accuracy and diversity. This yields diversified multi-discriminators that fit the generated data distribution to the real data distribution from more comprehensive perspectives. **2)** When training the generator, we **minimize** DDL to encourage the generator to confuse all discriminators. This enhances the diversity of the generated data distribution. Further, we propose a layer-sharing network architecture for the multi-discriminators, which allows them to learn from distinct perspectives about the shared low-level features through better collaboration. It also makes our model more lightweight than existing multi-discriminators approaches. Our DDL-GAN remarkably outperforms other GANs over five standard datasets for image generation tasks.

Index Terms— Generative Adversarial Networks

1. INTRODUCTION

Generative Adversarial Network (GAN) [1] is one of the mainstream techniques that can fit generated data into complicated real data. When being trained towards an adversarial equilibrium (if it exists) in a minimax game, the generator G attempts to fit the real data distribution P_{data}, while a discriminator D attempts to distinguish P_{data} and the generated data distribution P_G. In this two-player game, as long as D manages to distinguish the real from the fake with nonzero probability, it will generate feedback to G through back-propagation to improve its synthesized distribution. However, if D is too weak, as the case in Fig. 1(a), it will lead to *mode collapse* and fail to generate realistic data. A variety of techniques, e.g., weight clipping [2], gradient penalty [3], spectral normalization [4], and self-attention [5], have been introduced to enhance the modeling capability of D. The multi-discriminators framework [6] is an alternative method to strengthen D, where different Ds may focus on different perspectives of P_{data}. Hopefully, an ensemble of Ds can identify the underlying subtle distinctions between P_G and P_{data} and improve G as illustrated in Fig. 1(b) and Fig. 1(c). But such an ideal situation may not be practical, as the diversity of their decision boundaries is not guaranteed explicitly. The multi-discriminators are constructed with homogeneous network architecture and trained for the same task from the same training data. Thus, some of them will generate similar decision boundaries as shown in Fig. 1(d). In the worst case, they may even degenerate to a single discriminator.



(a) 1 weak D (b) 2 Ds (c) 3 Ds **(high diversity)** (d) 3 Ds **(low diversity)**

Fig. 1. Some distributions of real data (**Circle**) and generated data (**Triangle**) of a "winner" generator that successfully deceives the discriminators (shown as black lines). **(b)** shows that there can be a new discriminator that helps to progressively map the generated distribution to the real one; **(c)** shows the benefit of diversifying the multi-discriminators; and **(d)** shows that less diversified multi-discriminators tend to degenerate to a single discriminator in the worst case.

In this paper, we tackle this problem through diversifying the multi-discriminators *explicitly* by introducing a simple yet effective **Discriminator Discrepancy Loss (DDL)**. Accordingly, we present a *multi-player* minimax game, **DDL-GAN**, which unifies the optimization of DDL and the original GAN loss, seeking an optimal trade-off between the accuracy and diversity of multi-discriminators. At some point, the diversity

Copyright notice 978-1-7281-1331-9/20/$31.00 2020 IEEE

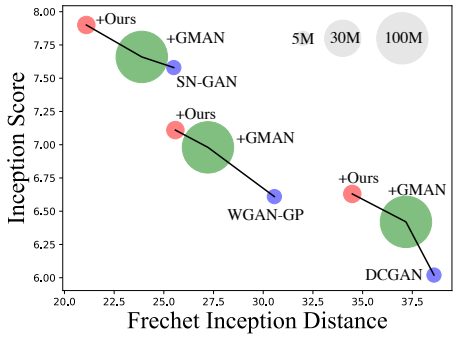


Fig. 2. The proposed DDL-GAN is an effective and efficient method for enhancing GANs with multi-discriminators. All models are trained on CIFAR10. The discs closer to the top left corner are models with better image generation results. The disc radius indicates the size of model parameters.

may be even more important than accuracy, as diversified Ds reveal P_{data} more comprehensively; otherwise, there would be no performance improvement if identical Ds are combined. Competitions not only exist between G and individual Ds but also between each pair of Ds; being trained iteratively, multiple Ds attempt to increase the margins between their decision boundaries by maximizing DDL, while G attempts to make Ds consistent on the generated data by minimizing DDL. Through the minimax game between all players over DDL, we guarantee to reach an *equilibrium* of the diversity, which is important to stabilize the training process.

Further, we propose a **layer-sharing architecture** for the multi-discriminators of DDL-GAN, which has two advantages. First, it encourages harmonious collaboration between different discriminators, thus further enhancing the diversity. Second, it makes the multi-discriminators more memory efficient. As shown in Fig. 2, different from the previous multi-discriminators framework [6], it enables our model to synthesize higher-quality images using much fewer parameters—a comparable model size to the original GAN with single discriminator. We testify the effectiveness of DDL over one synthetic and five standard datasets by successfully applying it to a wide range of GANs with various discriminator techniques.

2. RELATED WORK

In recent years, GANs have achieved great success in several challenging vision and language tasks, such as image-to-image translation [7, 8, 9, 10], image super-resolution [11, 12], and text-to-image synthesis [13, 14]. The conditional GANs enable semantic image synthesis and editing [15, 16, 17, 18, 19, 4, 20, 21, 5, 22, 23]. To address the common *mode collapse* problem, various metrics have been exploited to measure the distance between the generated data distribution and the real data distribution, such as the f-divergence [24], the Wasserstein distance [2], the least squares [25], and the optimal transport [26]. Many architectures have been proposed to enhance the quality of generated images. When GAN was first proposed [1], the generator and the discriminator were both multi-layer perceptrons, which were replaced by deep convolutional networks in DCGAN [27]. LAPGAN uses the Laplacian pyramid framework and builds a cascaded CNN architecture [16]. ProGAN gradually trains one layer at a time instead of training all layers of the generator and discriminator at once, and progressively produces images of higher resolutions [28]. SA-GAN introduces the self-attention mechanism to capture both local and global features of the images [29]. BigGAN adopts the truncation trick to balance the image fidelity and the variety [22]. StyleGAN proposes a style-based generator architecture to generate high-resolution images [30]. Multiple generators are also introduced to mitigate the mode collapse problem [31].

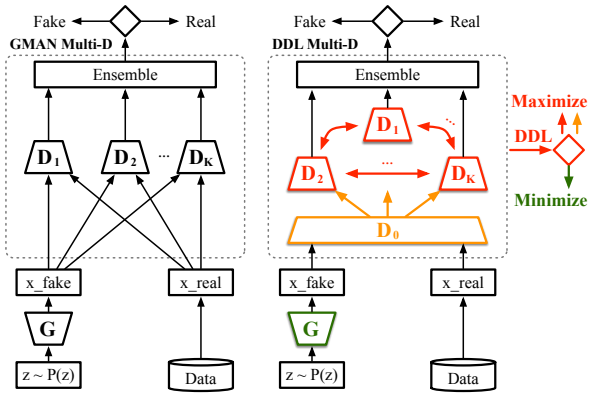
There are many techniques to improve the discriminators in GANs, such as the gradient penalty [3, 32, 33], spectral normalization [4], and regularized information maximization [34]. CatGAN learns a discriminator that separates the generated data into multiple modes [34]. UnrolledGAN builds a computational graph of multiple learning steps of the discriminator, and then back-propagates through all of them when computing the gradient on the generator [35]. D2GAN employs two discriminators to minimize both the KL and reverse KL divergences, thus placing a fair distribution across the data modes [36]. GMAN presents various methods such as boosting to ensemble multiple discriminators [6]. Unlike the existing work, we want the discriminators to be heterogeneous, inspired by the idea of ensemble learning that base learners should be diverse enough. Thus we introduce a new adversarially-learned function to diversify the multi-discriminators, which can be integrated into all above GANs.

3. METHOD

In this section, we present a generic method for explicitly diversifying the multi-discriminators of GANs. We first define the Discriminator Discrepancy Loss (DDL). Next, we formulate the minimax game for training DDL-GAN. Finally, we describe the new architecture of DDL-GAN (Fig. 3).

3.1. Discriminator Discrepancy Loss

Our Discriminator Discrepancy Loss (DDL) is related to the multiple discriminators mechanism studied in GMAN [6]. The efficacy of multiple discriminators has been proven in [37] from a perspective of density ratio estimation in GANs. Suppose P_{data} is the distribution of the real data. We sample a finite set from the population as our training set of empirical distribution $\tilde{P}_{\text{data}}$. The optimum of GANs corresponds to the equilibrium $P_{G*} = \tilde{P}_{\text{data}}$. It indicates that $G(z)$ generates data by mapping samples from a noise space to a finite set of input space, inducing a distribution P_{G*} that is identical to the data distribution $\tilde{P}_{\text{data}}$. Unfortunately, P_{G*} may be incapable of expressing the real distribution P_{data} of complex *multimodal* structures, which makes it necessary to employ multiple discriminators. The ensemble of multiple discriminators can identify the subtle distinctions underlying the real



(a) GMAN (b) Our method

Fig. 3. Two major new contributions of our DDL-GAN over GMAN: first, our method diversifies the predictions of $\{D_k\}_{k=1}^K$ by optimizing DDL adversarially; second, it introduces a layer-sharing architecture to allow $\{D_k\}_{k=1}^K$ to collaborate harmoniously, where D_0 denotes the shared layers.

and generated data distributions more effectively.

However, the ideal situation of GMAN where the K discriminators excel in separate regions of the data space is not always practical, because generating diverse individual discriminators is not easy. The major obstacle lies in the fact that the individual discriminators in GMAN are trained without any *explicit* diversity-enhancing criterion. Specifically, they are trained for the same task of distinguishing real from fake and over different samples from the same training data, and thus they are usually highly correlated. It is inevitable that during training, some of them will become very similar or even degenerate to similar decision boundaries as shown in Fig. 1(d). As this *discriminator degeneracy* problem becomes severe, the multi-discriminator GAN will be degenerated to the vanilla GAN with a single discriminator.

Intuitively, to make the generative model gain from the competition with multiple discriminators, the individual discriminators must be different. Otherwise, there would be no performance gain if identical individual discriminators are combined using the original GAN loss. In light of this, we introduce the **Discriminator Discrepancy Loss (DDL)**, which is computed as the overall distance between the output of each discriminator and the averaged output of all discriminators:

$$L_{\text{DDL}}\left(x;\{D_k\}_{k=1}^K\right)=\frac{1}{K}\sum_{k=1}^{K}\left|\phi\left(D_k(x)\right)-\sum_{k'=1}^{K}\frac{\phi\left(D_{k'}(x)\right)}{K}\right|, \tag{1}$$

where x is either a real image or a generated image, $D_k(x)$ is the output of the k^{th} discriminator, and $|\cdot|$ is the ℓ_1 loss. To make DDL compatible with the general framework of GANs [38], ϕ is chosen as a concave function: $\phi(t)=\log(t)$ for the vanilla GAN [1] and $\phi(t)=t$ for WGAN [2]. DDL is a natural diversity metric for multi-discriminators, and larger DDL indicates more diverse predictions by $\{D_k\}_{k=1}^K$.

3.2. DDL Minimax Game

We apply DDL to a wide range of GANs in a *multi-player* minimax training paradigm. Multi-discriminators $\{D_k\}_{k=1}^K$ are trained to maximize DDL while G is trained to minimize it. DDL can be jointly learned by the original adversarial loss of GAN. The final objective of **DDL-GAN** is formulated as:

$$\begin{aligned}L(\theta_G,\{\theta_D^k\}_{k=1}^K) &= \mathbb{E}_{x\sim P_{\text{data}}}\sum_{k=1}^{K}\frac{\phi\left(D_k\left(x\right)\right)}{K}\\ &+\mathbb{E}_{z\sim P_z}\sum_{k=1}^{K}\frac{\phi\left(1-D_k\left(G(z)\right)\right)}{K}\\ &+\lambda\mathbb{E}_{x\sim P_{\text{data}}}L_{\text{DDL}}\left(x;\{D_k\}_{k=1}^K\right)\\ &+\lambda\mathbb{E}_{z\sim P_z}L_{\text{DDL}}\left(G(z);\{D_k\}_{k=1}^K\right).\end{aligned} \tag{2}$$

We use the unified framework of GANs [38], where ϕ is a concave function. Alternatives include $\phi(t)=\log(t)$ for the vanilla GAN [1] and $\phi(t)=t$ for WGAN [2]. P_{data} denotes the real data distribution and P_z denotes the distribution of noise vector z, θ_G is the parameters of the generator, θ_D^k is the parameters of the k^{th} discriminator, and λ is a coefficient between the DDL and the original GAN loss.

Training $\{D_k\}_{k=1}^K$ to Maximize DDL. Similar to ensemble learning where it is desired that the individual learners should be both accurate and diverse, in our case, diversity is as important as accuracy. Combining only accurate discriminators is often worse than combining some relatively weak ones that can correspond to different regions of the real data space. Thus, we train $\{D_k\}_{k=1}^K$ to enhance the diversity among them by explicitly maximizing DDL. Ultimately, the success of training competitive Ds that can further enhance the training of G lies in achieving a good trade-off between the individual accuracy and diversity of Ds. Along with the original adversarial objective of GANs, the training procedure of $\{\theta_D^k\}_{k=1}^K$ can be formulated as $\{\hat{\theta}_D^k\}_{k=1}^K=\arg\max_{\theta_D^1,\ldots,\theta_D^K}L(\theta_G,\{\theta_D^k\}_{k=1}^K)$.

Training G to Minimize DDL. Notably, we do not maximize DDL monotonously. While $\{D_k\}_{k=1}^K$ attempt to maximize DDL to enhance the diversity of the base discriminators, G is trained to trick all $\{D_k\}_{k=1}^K$. Thus G is adversarially trained to minimize DDL to make all discriminators perform consistently and collaborate harmoniously. Along with the original adversarial objective of GANs, the training procedure of G can be formulated as $\hat{\theta}_G=\arg\min_{\theta_G}L(\theta_G,\{\theta_D^k\}_{k=1}^K)$.

Remark. The proposed DDL method has not been used by existing ensemble learning algorithms, though it is inspired by the general idea of enhancing the diversity of base learners. Essentially, we optimize DDL adversarially in pursuit of a competitive generative model; this is actually not the case in ensemble learning where the ultimate goal is to purely increase the accuracy of the ensemble learner. Such adversarial training process yields more competitive $\{D_k\}_{k=1}^K$ and an all-rounded G that can generate data close to real distribution.

3.3. Layer-Sharing Architecture

Low-level features are significant for image generation. Ideally, we want these discriminators to learn from distinct perspectives about the low-level features. However, multiple homogeneous networks as in GMAN [6] tend to extract redundant low-level features. To address this problem, instead of making each individual discriminator learn from raw pixel values as in GMAN, we make all discriminators share common lower layers and diverge at a mid-level network layer. This layer-sharing architecture potentially enables a harmonic collaboration across all discriminators.

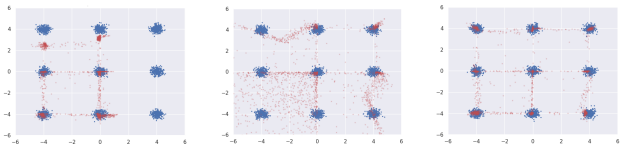
Another consideration is that as K becomes larger, the multi-discriminators in GMAN will be increasingly over-parametrized, leading to training difficulties. This drawback is less prominent in the original GMAN with only 2 to 5 discriminators. However, in our experiments, we find that using more discriminators are generally helpful in enhancing the performance of GMAN, but it increases the model size and worsens the training stability. Sharing lower layers will reduce the number of parameters of the multi-discriminators by an order and make our model more scalable (Fig. 2). We employ a total of N layers for each base discriminator, in which the M lower layers are shared by all discriminators. The architecture details are shown in the supplementary material.

4. EXPERIMENTS

In this section, we compare the DDL-GAN with a wide range of GANs over a synthetic dataset and five real datasets. We also discuss the sensitivity of DDL-GAN to K and λ.

4.1. Warm Up: Toy Dataset

We validate the DDL minimax training procedure on a 2D synthetic dataset with 9 clusters. We apply 4 discriminators for both GMAN and DDL-GAN. Each discriminator consists of a three-layer MLP. As shown in Fig. 4, maximizing DDL alleviates the mode collapse better than GMAN, while optimizing DDL in a minimax game performs best.



(a) GMAN (b) Maximize DDL (c) DDL (Minimax)

Fig. 4. Our DDL-GAN in a minimax training procedure better alleviates the mode collapse than GMAN.

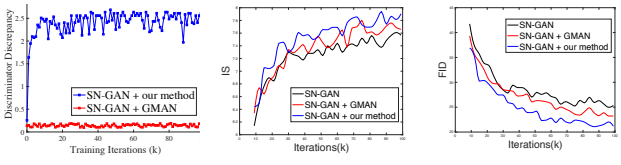
4.2. CIFAR10

On the CIFAR10 dataset [39], we train GMAN with $K = 16$ discriminators based on DCGAN, WGAN-GP (with Gradient Penalty), and SN-GAN (with Spectral Normalization). Then we apply the DDL to the same base networks by adding the adversarial trained DDL penalty. Training and network details are included in the supplementary material.

Table 1 shows the quantitative results on CIFAR10. We use two widely-used metrics: a higher Inception Score (IS) [40] or a lower Frechet Inception Distance (FID) [41] indicates higher fidelity and diversity of the generated images. The GMAN models outperform GANs with a single discriminator, but the improvement over SN-GAN is relatively limited. Our DDL-GAN has a consistent advantage over GMAN across all base networks, even for models with well-designed discriminator structures such as SN-GAN. The last row of Table 1 shows that sharing lower layers is effective upon DDL. Throughout training, our method enhances the diversity of multi-discriminators (Fig. 5(a)), and better IS and FID scores of the generated images (Fig. 5(b) and Fig. 5(c)). It indicates that DDL improves the generator stably across the training procedure by enhancing the diversity of the discriminators.

Model	DCGAN	WGAN-GP	SN-GAN
Vanilla	6.02 / 38.59	6.61 / 30.56	7.58 / 25.50
+ GMAN	6.42 / 37.18	6.98 / 27.22	7.66 / 23.89
+ DDL	**6.63 / 34.48**	**7.11 / 25.58**	**7.90 / 21.01**
+ DDL*	6.37 / 35.16	7.04 / 26.14	7.71 / 23.64

Table 1. IS/FID results on CIFAR10. DDL* is a variant of our method without shared layers between discriminators.



(a) Diversity score (b) IS score (c) FID score

Fig. 5. A comparison between our method and GMAN by (a) the DDL score; (b) IS; (c) FID on CIFAR10 during training.

4.3. STL-10

The STL-10 dataset [42] is more complicated and contains 100,000 training images and 800 testing images. Images are randomly cropped to 48×48. We compare our DDL-GAN and GMAN with both $K = 16$ discriminators. Table 2 shows the IS and FID results averaged over 5 training runs. Our DDL-GAN significantly outperforms GMAN.

Model	WGAN	WGAN-GP	SN-GAN
Vanilla	7.57 / 64.20	8.42 / 55.10	8.79 / 43.20
+ GMAN	7.82 / 54.93	8.72 / 47.26	8.86 / 41.67
+ DDL	**7.92 / 48.05**	**8.94 / 44.80**	**9.21 / 39.68**

Table 2. IS/FID results on the STL-10 dataset.

4.4. CelebA

The CelebA dataset [43] is a face attributes dataset containing 202,599 images of size 218×178. We randomly crop images and resize them to 64×64. We take WGAN as the base model. As above, we equip WGAN with 16 discriminators with the same architecture and then diversify their predictions with DDL. After 40,000 training iterations, we calcu-

Fig. 6. The generated 256×256 images on LSUN-Bedroom.

late the IS and FID scores of the generated images and compare these results to GMAN under a comparable number of model parameters and similar training techniques. Our proposed DDL-GAN improves WGAN and outperforms GMAN by a large margin (Table 3). In the supplementary material, we supply more qualitative results to show the improvement of our method over WGAN across entire training procedure.

CelebA	IS / FID	ImageNet	IS / Intra FID
WGAN	1.67 / 45.17	SN-GAN-Proj	36.8 / 92.4
+ GMAN	1.66 / 41.09	+ GMAN	37.6 / 89.5
+ DDL	**1.75 / 39.15**	+ DDL	**39.7 / 83.7**

Table 3. Results of our method on CelebA and ImageNet.

4.5. ImageNet

ImageNet [44] is a large-scale dataset that can be used for class-conditional image generation. Each image is randomly cropped and resized to 128×128. We train our model over 1,000 classes on ImageNet and compare it with GMAN by applying them to SN-GAN-Proj [20], a competitive model with a well-designed discriminator. Table 3 shows the quantitative results. DDL enhances the quality of the class conditional image generation on the base of SN-GAN-Proj and GMAN, showing that our method can not only improve GANs with weak discriminators, but also those with strong ones.

4.6. LSUN-Bedroom

The LSUN dataset [45] consists of one million images of 10 scene categories and 20 object categories. We use the unlabeled 256×256 bedroom images. We take StyleGAN [30] as the base model, which is competitive in generating high-resolution images. For a fair comparison, we follow StyleGAN to include the Perceptual Path Length [46] as an evaluation metric (lower is better). Results are shown in Table 4 and Fig. 6. We notice that the GMAN method with the same number of Ds (32) is hard to train due to model complexity.

4.7. Ablation Study

We investigate the sensitivity of DDL-GAN to the number of discriminators. It consistently outperforms the baseline models for all tested K values and achieves the best results at $K = 16, 18, 16, 16, 32$ on CIFAR10, STL-10, CelebA, ImageNet, and LSUN. We also evaluate the training coefficient λ in Eq. (2) that balances DDL and the basic GAN loss. We find that small λ leads to similar outputs of multiple Ds, while an excessively large λ is not optimal either (see the supplementary material). The best result is achieved at $\lambda = 1.0$.

Model	FID	Perceptual Path Length	
		Full	End
StyleGAN	3.324	2419.78	1349.88
+ GMAN	2.862	2378.29	1302.09
+ DDL	**2.606**	**2314.87**	**1282.97**

Table 4. Results of our method on LSUN-Bedroom.

5. CONCLUSION

This paper proposed the Discriminator Discrepancy Loss (DDL) to diversify multi-discriminators of GANs. DDL turns the two-player training objective into a multi-player one. Unlike GMAN that seeks the classification accuracy when training the discriminators, we sought a trade-off between the diversity and the accuracy. We diversified the discriminators by maximizing DDL, and alternately trained the generator by minimizing DDL. We also proposed a layer-sharing architecture for the multi-discriminators, which allows these discriminators to learn from distinct perspectives about the low-level features. It enables collaboration across all discriminators and allows GANs with multi-discriminators to be trained more easily. We applied DDL to a wide range of GANs and showed its effectiveness by comparing it with GMAN.

6. ACKNOWLEDGE

This work was supported in part by the Natural Science Foundation of China (61772299, 71690231), and the MOE Strategic Research Project on Artificial Intelligence Algorithms for Big Data Analysis.

7. REFERENCES

[1] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial nets," in *NeurIPS*, 2014. 1, 2, 3

[2] Martin Arjovsky, Soumith Chintala, and Léon Bottou, "Wasserstein generative adversarial networks," in *ICML*, 2017. 1, 2, 3

[3] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville, "Improved training of wasserstein gans," in *NeurIPS*, 2017. 1, 2

[4] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida, "Spectral normalization for generative adversarial networks," in *ICLR*, 2018. 1, 2

[5] Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena, "Self-attention generative adversarial networks," in *NeurIPS*, 2018. 1, 2

[6] Ishan Durugkar, Ian Gemp, and Sridhar Mahadevan, "Generative multi-adversarial networks," in *ICLR*, 2017. 1, 2, 4

[7] Phillip Isola, Jun Yan Zhu, Tinghui Zhou, and Alexei A. Efros, "Image-to-image translation with conditional adversarial networks," in *CVPR*, 2017. 2

[8] Ming-Yu Liu and Oncel Tuzel, "Coupled generative adversarial networks," in *NeurIPS*, 2016. 2

[9] Yaniv Taigman, Adam Polyak, and Lior Wolf, "Unsupervised cross-domain image generation," in *ICLR*, 2017. 2

[10] Jun Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *ICCV*, 2017. 2

[11] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew P Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al., "Photo-realistic single image super-resolution using a generative adversarial network.," in *CVPR*, 2017. 2

[12] Casper Kaae Sønderby, Jose Caballero, Lucas Theis, Wenzhe Shi, and Ferenc Huszár, "Amortised map inference for image super-resolution," in *ICLR*, 2017. 2

[13] Scott Reed, Zeynep Akata, Xinchen Yan, Lajanugen Logeswaran, Bernt Schiele, and Honglak Lee, "Generative adversarial text to image synthesis," in *ICML*, 2016. 2

[14] Scott E Reed, Zeynep Akata, Santosh Mohan, Samuel Tenka, Bernt Schiele, and Honglak Lee, "Learning what and where to draw," in *NeurIPS*, 2016. 2

[15] Mehdi Mirza and Simon Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014. 2

[16] Emily L Denton, Soumith Chintala, arthur szlam, and Rob Fergus, "Deep generative image models using a laplacian pyramid of adversarial networks," in *NeurIPS*, 2015. 2

[17] Augustus Odena, Christopher Olah, and Jonathon Shlens, "Conditional image synthesis with auxiliary classifier GANs," in *ICML*, 2017. 2

[18] Harm de Vries, Florian Strub, Jeremie Mary, Hugo Larochelle, Olivier Pietquin, and Aaron C Courville, "Modulating early visual processing by language," in *NeurIPS*, 2017. 2

[19] Augustus Odena, Jacob Buckman, Catherine Olsson, Tom Brown, Christopher Olah, Colin Raffel, and Ian Goodfellow, "Is generator conditioning causally related to GAN performance?," in *ICML*, 2018. 2

[20] Takeru Miyato and Masanori Koyama, "cgans with projection discriminator," in *ICLR*, 2018. 2, 5

[21] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," in *CVPR*, 2018. 2

[22] Andrew Brock, Jeff Donahue, and Karen Simonyan, "Large scale GAN training for high fidelity natural image synthesis," in *ICLR*, 2019. 2

[23] Tamar Rott Shaham, Tali Dekel, and Tomer Michaeli, "Singan: Learning a generative model from a single natural image," in *ICCV*, 2019. 2

[24] Sebastian Nowozin, Botond Cseke, and Ryota Tomioka, "f-gan: Training generative neural samplers using variational divergence minimization," in *NeurIPS*, 2016. 2

[25] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley, "Least squares generative adversarial networks," in *ICCV*, 2017. 2

[26] Tim Salimans, Han Zhang, Alec Radford, and Dimitris Metaxas, "Improving gans using optimal transport," in *ICLR*, 2018. 2

[27] Alec Radford, Luke Metz, and Soumith Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *ICLR*, 2016. 2

[28] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," in *ICLR*, 2018. 2

[29] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin, "Attention is all you need," in *NeurIPS*, 2017. 2

[30] Tero Karras, Samuli Laine, and Timo Aila, "A style-based generator architecture for generative adversarial networks," in *CVPR*, 2019. 2, 5

[31] Quan Hoang, Tu Dinh Nguyen, Trung Le, and Dinh Phung, "Mgan: Training generative adversarial nets with multiple generators," in *ICLR*, 2018. 2

[32] Naveen Kodali, James Hays, Jacob Abernethy, and Zsolt Kira, "On convergence and stability of GANs," in *ICLR*, 2018. 2

[33] Lars Mescheder, Andreas Geiger, and Sebastian Nowozin, "Which training methods for GANs do actually converge?," in *ICML*, 2018. 2

[34] Jost Tobias Springenberg, "Unsupervised and semi-supervised learning with categorical generative adversarial networks," in *ICLR*, 2016. 2

[35] Luke Metz, Ben Poole, David Pfau, and Jascha Sohl-Dickstein, "Unrolled generative adversarial networks," in *ICLR*, 2017. 2

[36] Tu Nguyen, Trung Le, Hung Vu, and Dinh Phung, "Dual discriminator generative adversarial nets," in *NeurIPS*, 2017. 2

[37] Masatosi Uehara, Issei Sato, Masahiro Suzuki, Kotaro Nakayama, and Yutaka Matsuo, "b-gan: Unified framework of generative adversarial networks," in *ICLR*, 2017. 2

[38] Sanjeev Arora, Rong Ge, Yingyu Liang, Tengyu Ma, and Yi Zhang, "Generalization and equilibrium in generative adversarial nets (gans)," in *ICML*, 2017. 3

[39] Antonio Torralba, Rob Fergus, and William T Freeman, "80 million tiny images: A large data set for nonparametric object and scene recognition," *IEEE transactions on pattern analysis and machine intelligence*, 2008. 4

[40] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen, "Improved techniques for training gans," in *NeurIPS*, 2016. 4

[41] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter, "Gans trained by a two time-scale update rule converge to a local nash equilibrium," in *NeurIPS*, 2017. 4

[42] Adam Coates, Andrew Ng, and Honglak Lee, "An analysis of single-layer networks in unsupervised feature learning," in *AISTATS*, 2011. 4

[43] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang, "Deep learning face attributes in the wild," in *ICCV*, 2015. 4

[44] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *CVPR*, 2009. 5

[45] Fisher Yu, Ari Seff, Yinda Zhang, Shuran Song, Thomas Funkhouser, and Jianxiong Xiao, "Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop," *arXiv preprint arXiv:1506.03365*, 2015. 5

[46] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *CVPR*, 2018. 5