

GAN-based Image Generation Techniques Exploiting Latent Vector Distribution and Edge Loss Methods on Limited Datasets

Yun-Gyeong Song¹, Gun-Woo Kim^{2*}

¹ Department of AI Convergence Engineering, Gyeongsang National University, Jinju, Republic of Korea

² School of Computer Science Gyeongsang National University, Jinju, Republic of Korea

Abstract

Generative Adversarial Networks (GANs) have demonstrated remarkable capabilities in image generation, surpassing the performance of previous image generation models. However, GANs require large training datasets to facilitate proper learning. GANs have inherent problems such as the mode collapse problem, where identical images are generated, and instability problem, where the generator and the discriminator fail to form a successful adversarial relationship. These problems are particularly common when the availability of training data is limited. In this paper, we propose three techniques to address these challenges. Firstly, Common Feature Training (CFT) is introduced to enhance performance by training the Generator to recognize common features, thereby mitigating instability problems. Secondly, Mean Rescaling (MR) is employed to mitigate the mode collapse problem arising from sampling latent vectors with identical means and variances. Thirdly, an edge loss method is implemented, where the edge difference values between real and generated images are added to the GANs loss. This contributes to the classification of shapes, thereby mitigating the mode collapse problem and instability problem. Comparative experimental results illustrate improvements in the highlighted issues, and the performance enhancement is validated by metrics, namely Fréchet Inception Distance (FID) and Inception Score (IS).

Keywords

Limited data, Generative model, Mode Collapse, Instability, Edge Loss, Latent Vector Distribution

1. Introduction

Generative models have demonstrated remarkable performance improvements with the advent of large datasets and deep neural networks. It is used for a variety of tasks, including inpainting and image translation tasks [9,10]. Generative Adversarial Networks (GANs) are a framework for estimating generative models through adversarial learning of generators and discriminators. The discriminator estimates the probability that the input data is real, while the generator generates fake data that mimics the distribution of real data to deceive the discriminator. Due to these characteristics, the discriminator is trained to distinguish real data from fake data, and the generator is trained to produce fake data that closely approximates real data. Training GANs inherently require a large dataset, as insufficient data can lead to instability problem such as a lack of adversarial relationship between the generator and discriminator, and a mode collapse problem, where the generator primarily produces an identical image [1,11].

Extensive research has been conducted to mitigate the mode collapse problem and instability problem in traditional GANs by improving the model's architecture and loss function [2]. However, most of these researches heavily rely on the use of large datasets. The basic solution to mode collapse problem and instability problem is to augment the image or add more datasets [12,13]. Otherwise, without such large datasets, producing diverse and high-quality data becomes a challenge. In this paper, we propose following three techniques to address the problems of

ISE 2023: 2nd International Workshop on Intelligent Software Engineering, December 4, 2023, Seoul

*Corresponding author.

✉ songyg1020@gnu.ac.kr (Y. G. Song); gunwoo.kim@gnu.ac.kr (G. W. Kim)

ORCID 0009-0008-6692-1925 (Y. G. Song); 0000-0001-5643-4797 (G. W. Kim)

© 2023 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

instability and mode collapse. Figure 1 shows the learning process and architecture of our proposed GAN model.

1. Common Feature Training (CFT): We propose the Common Features Training (CFT) to facilitate the Generator's training and establish an adversarial relationship between the Generator and Discriminator to address the instability problem.
2. Mean Rescaling (MR): We propose the Mean Rescaling (MR), which rescales the mean of the sampled vector z by replacing it with a random mean to mitigate the mode collapse problem.
3. edge loss: We propose edge loss to inhibit the fast learning of the discriminator and use GANs + edge loss to classify not only the authenticity of the image but also its shape, which mitigates the problems of mode collapse and instability.

2. Related work

GANs have a different approach to generating high-quality images compared to traditional generative models, and they have demonstrated excellent performance in image generation since the release of Deep Convolutional Generative Adversarial Networks (DCGAN), which was changed to Deep Convolution Nets [1,3]. However, when generating an image with limited data, the generator and discriminator may not form an adversarial relationship due to lack of data, which leads to problems of instability and mode collapse.

To address the mode collapse problem, Diverse and Limited Data Generative Adversarial Networks (DeLiGAN) were proposed [4]. DeLiGAN creates various components through a Gaussian mixture model and reparametrizes the latent vector and latent space, through random sampling. This research follows the approach of modifying the latent distribution to obtain samples in the high-probability region [5]. However, this approach involves the training of mean and variance components, which can be time-consuming and does not address the instability problem.

3. Method

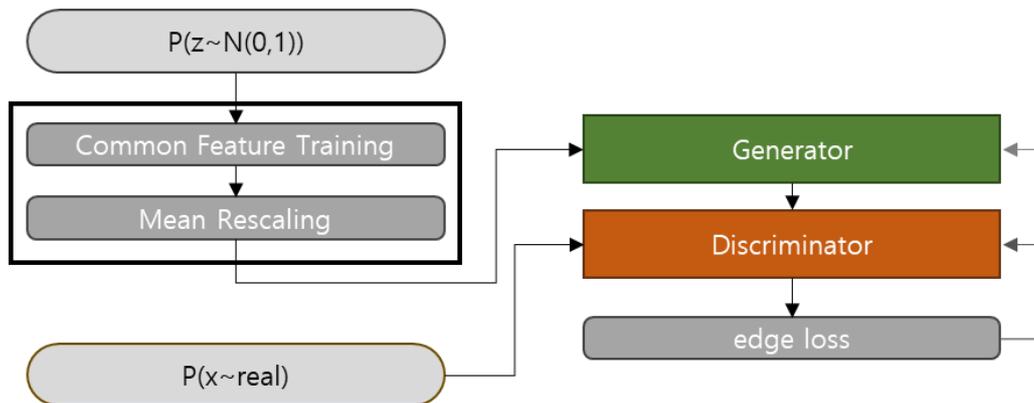


Figure 1: Learning process and architecture of the proposed GAN model.

Latent vectors are randomly sampled from a normal distribution but have entangled features because one feature is related to another [16]. We sample from a random normal distribution, the learning nature of the discriminator will cause the generator to synthesize an identical image that will best fool the discriminator [14]. The mode collapse problem can be mitigated to an extent by reparametrizing the distribution of the latent vector [4]. They can be trained with additional information to direct the generative model and slow down the learning of the discriminator [15]. This mitigates the instability problem.

3.1. Common Feature training

Common Feature Training (CFT) is a technique to enhance the stability of learning. It consists of reconstructing a feature vector z into a distribution that incorporates the common features of the dataset, allowing for the learning of these common features. In this technique, we adopt a hyperparameter called Feature Differences (FD). This hyperparameter reflects the extent to which common features are considered and is initially set to 1. It is incremented by 1 until no instability problem occurs beyond epoch 30, at which point we adopt an appropriate FD value. This helps the generator learn to create an adversarial relationship. The equation of the reconstruction process is as follows:

$$Z = \left(\sum_{i=0}^N x_i w_{1i}, \sum_{i=0}^N x_i w_{2i}, \dots, \sum_{i=0}^N x_i w_{Ni} \right) \quad (1)$$

where Z is a latent vector and x is a random value randomly sampled from a normal distribution. w is the weight of the common feature, $x_i w_{Ni}$ means each feature. Take an input of sampling size N by a normal distribution and run it through a linear layout to get a trained Z of size $N + FD$. In this procedure, each feature's value is learned in such a way that it reduces the loss value, and a variable Z with a common feature is obtained.

FD is a quantification of the degree to which a feature is included, and choosing a high FD value can result in less diverse data because the distribution learns detailed features. Conversely, selecting a low FD value may result in a lack of common features being incorporated. Therefore, it is important to set the correct FD value. Additionally, if the data does not have common features between individual data points, the technique may not be very effective, so prior data analysis is required.

3.2. Mean rescaling

Mean Rescaling (MR) is a technique that involves adding a random average value, ranging from -1 to 1, to the feature vector z after it has undergone CFT. This technique ensures that each latent vector z has a unique average, devised to mitigate the mode collapse problem occurring when the initial latent vectors z have the same mean and variance. The equation of this technique is as follows:

$$z = z + \mu \quad (2)$$

where z is the value of the potential vector generated by the CFT and μ is the mean value added to z and is an arbitrary value between -1 and 1.

MR enables the latent vector z to have the mean of a user-specified range. This technique mitigates the mode collapse problem and overcomes the problem of generating only specific classes of data, enabling the generation of a variety of data categories.

3.3. Edge loss

Most of the instability problem is caused by the faster training of the Discriminator compared to the Generator. To address this problem, we incorporate the similarity of the edges between the generated image and the real image into the loss function. The equation of loss function is as follows:

$$\min_G \max_D V(D, G) = E_{x \sim P_{data(x)}} \left[\log D(x) + \sum (G(z) - x)^2 \right] + E_{x \sim P_{z(z)}} \left[\log (1 - D(G(z))) \right] \quad (3)$$

where x is the real image and $D(x)$ is the probability that it is the real image. $G(z) - x$ is the difference between the image $G(z)$ generated by the Generator and the real image x , which also

allows the Discriminator to learn the shape of the image. The Generator generates images that look like the real image through $1 - D(G(z))$.

Since the discriminator is responsible for distinguishing between real and generated data, it does not recognize the shape of the generated data. However, with formula (3), the discriminator takes shape into consideration. This technique not only ensures the reliability of the data but also allows for a more accurate understanding of the image shape information. This mitigates the mode collapse problem and mitigates the instability problem by slowing down the learning speed of the Discriminator to compete with the Generator.

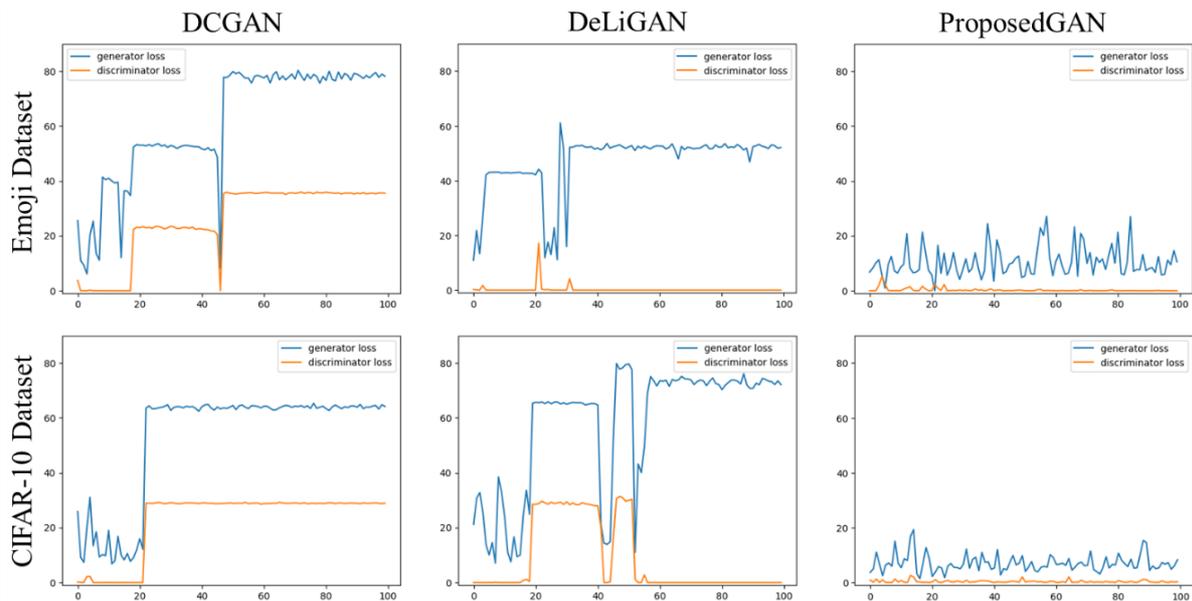


Figure 2: Training loss of the Generator and Discriminator for each model and dataset.

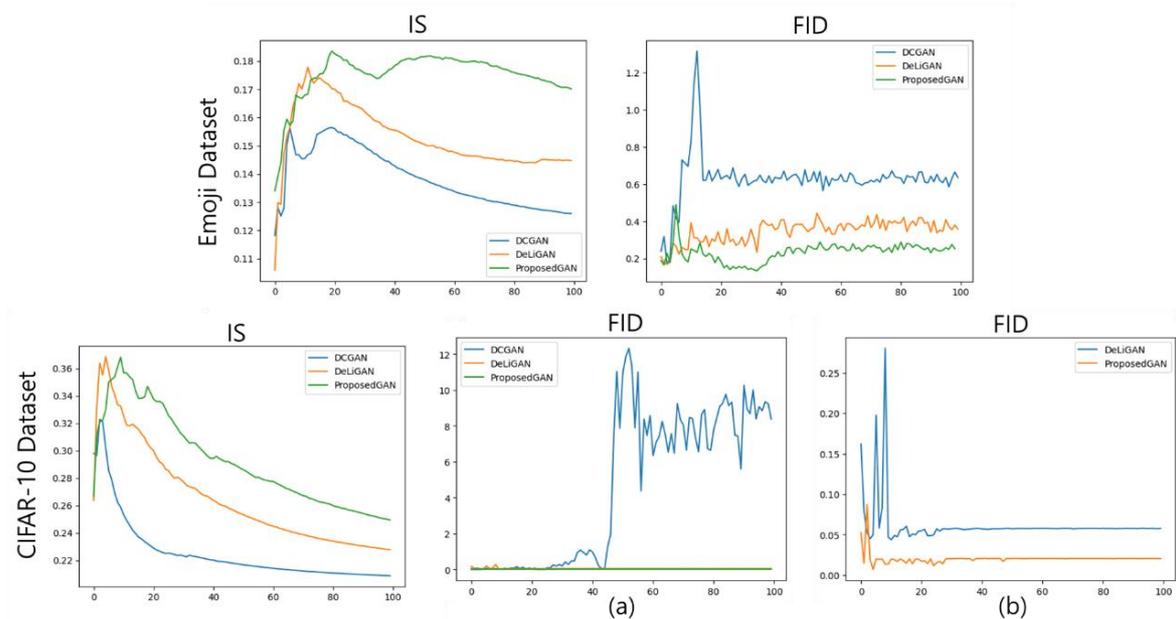


Figure 3: FID and IS graphs for each model and dataset.

Table 1

FID and IS tables for each model and dataset.

MODEL	Emoji		CIFAR-10	
	FID	IS	FID	IS
DCGAN	0.62 ± 0.13	0.13 ± 0.01	4.60 ± 4.21	0.22 ± 0.02
DeLiGAN	0.34 ± 0.05	0.15 ± 0.01	0.06 ± 0.02	0.26 ± 0.03
Proposed GAN	0.23 ± 0.05	0.17 ± 0.01	0.02 ± 0.01	0.29 ± 0.03

4. Experiments

The experimental environment in this paper is as follows: We used PyTorch version 1.12.1, the optimization algorithm is Adam [8], the learning rate is 1e-2, and the batch size is 32. Figure 2 shows the training loss graph for each dataset and model, and we can see that our proposed GAN is more stable in all datasets compared to other models.

4.1. Metrics

For the evaluation, we used Fréchet Inception Distance (FID) [6] and Inception Score (IS) [7]. FID is a metric that estimates the similarity of the distribution of real and generated data and calculates the distance, indicating how similar the two data are. The equation for FID is as follows:

$$FID = |\mu - \mu_w| + tr \left(\Sigma + \Sigma_w - 2 \left(\Sigma \Sigma_w \right)^{\frac{1}{2}} \right) \quad (4)$$

where $|\mu - \mu_w|$ is the feature vector mean of the true image distribution and the generated image distribution. $\Sigma + \Sigma_w$ is the sum of the true and generated image covariance matrices, and $(\Sigma \Sigma_w)^{\frac{1}{2}}$ is the $\frac{1}{2}$ square root matrix of the two covariances. Add $|\mu - \mu_w|$ to the value of Tr to get the FID. The lower the FID, the more similar the generated data is to the real data.

IS is another metric used to evaluate the quality and diversity of generated data, predicting the class of generated data using the Inception Network. The equation for IS is as follows:

$$IS = exp \left(\mathbb{E}_x KL(p(y|x)||p(y)) \right) \quad (5)$$

where $KL(p(y|x)||p(y))$ represents the Kullback-Leibler divergence between the predicted image distribution for image x and the true image distribution. This quantifies the predictive value of the difference between these two probability distributions. The higher the IS, the better the performance in terms of quality and variety of data generated.

The above metrics were used to compare the performance of each model and used to demonstrate the performance of the proposed GAN.

4.2. Dataset

In this paper, we conducted experiments using the Emoji Dataset and CIFAR Dataset. The Emoji Dataset comprises 402 positive classes and 402 negative classes, totaling 804 images. The CIFAR-10 dataset consists of 10 classes (airplane, car, bird, cat, deer, dog, frog, horse, ship, truck), with 6,000 images per class, resulting in a total of 60,000 images. However, for this research, 1,000 images were used.

4.3. Results and Performance

In Figure 3, as well as in Table 1, the IS and FID values for each dataset and model are displayed. Figure 3 reveals that our proposed GAN has the optimal performance, boasting the lowest FID values and the highest IS. In the CIFAR-10 dataset of Figure 3, the FID of DCGAN is overwhelmingly large in Figure 3-(a), so we exclude it and show the FID of DeLiGAN and our proposed GAN in Figure 3-(b).

When compared to DCGAN, our proposed GAN exhibits a 62.90% reduction in FID and a 30.77% increase in IS on the emoji dataset, as well as a 99.57% reduction in FID and a 31.82% increase in IS on the CIFAR-10 dataset. In comparison to DeLiGAN, our proposed GAN shows a 32.35% reduction in FID and a 13.33% increase in IS on the emoji dataset, and a 66.67% reduction in FID and an 11.54% increase in IS on the CIFAR-10 dataset.

Examining the generated images in Figure 4, it becomes visually evident that our proposed GAN outperforms others by generating images that are not only clearer but also more diverse. This underlines that the model, enhanced with the proposed technique and compared against DCGAN and DeLiGAN in experiments, produces exceptionally high-quality images that bear a striking resemblance to real images.

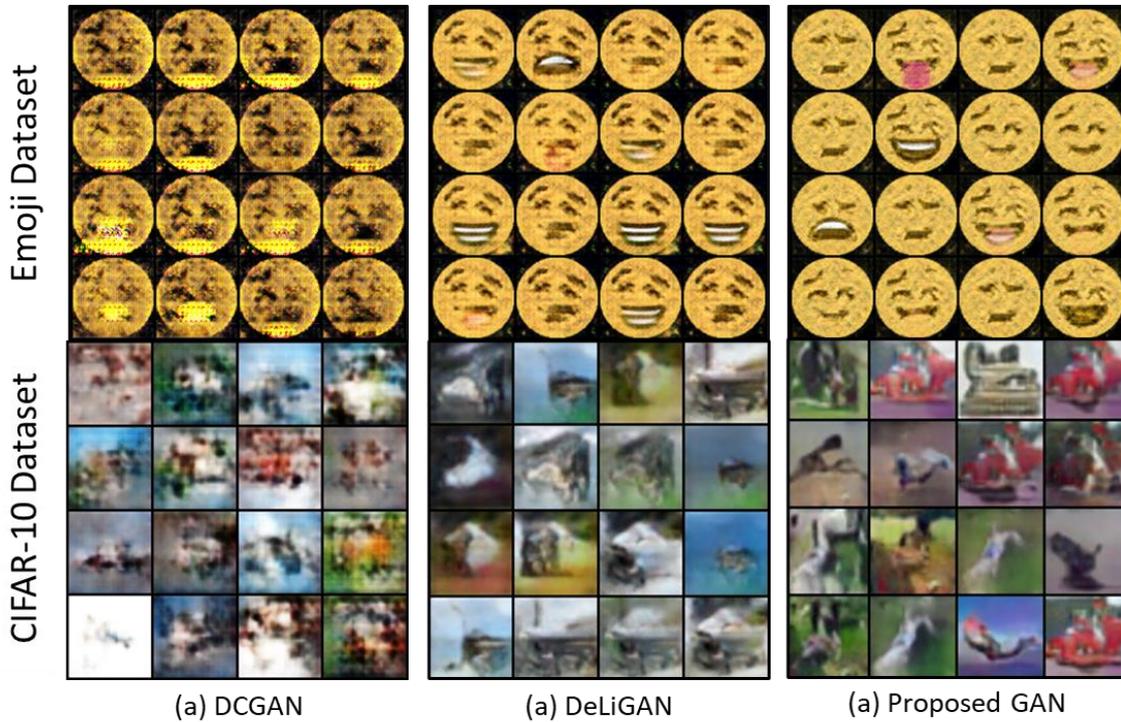


Figure 4: Generative image for each model and dataset.

5. Conclusion

In this paper, we propose Common Feature training (CFT), Mean rescaling (MR), and edge loss to resolve the learning instability problem and mode collapse problem.

Common Feature training (CFT) is intended to train the latent vector z on the overall shape of the data. This causes the generator to learn a common feature to compete with the discriminator, whereby mitigates the instability problem.

Mean rescaling (MR) is a technique to mitigate the mode collapse problem caused by a latent vector z sampled from a distribution with the same mean. Mitigate the mode collapse problem by sampling latent vectors z with different means.

Edge loss is a loss function that adds the difference between the edge of real data and the edge of generative data to GANs loss and does not simply classify whether it is real or generative, but

also learns image shape information to generate various images and slows down D's learning speed. whereby mitigate the mode collapse problem and instability problem. As a result, Figure 2 shows that the proposed GAN trains reliably compared to other models, and Figure 3 shows that our proposed GAN produces higher quality and more diverse images compared to other models.

Acknowledgements

This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education, Science and Technology (NRF-2021R1G1A1006381).

References

- [1] Ian Goodfellow, Jean Pouget Abadie, Mehdi Mirza, Bing Xu, David Warde Farley, Sherjil Ozair, Aaron Courville and Yoshua Bengio, "Generative adversarial nets", *Advances in neural information processing systems*, volume 27, 2014, pp 139-144
- [2] Xudong Mao, Qing Li, Haoran Xie, Raymond Y.K. Lau, Zhen Wang and Stephen Paul Smolley Stephen, "Least squares generative adversarial networks", *Proceedings of the IEEE international conference on computer vision (ICCV)*, 2017, pp 2794-2802
- [3] Alec Radford, Luke Metz and Soumith Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks", in *Proc. Int. Conf. Learn. Representations*, 2016
- [4] Swaminathan Gurumurthy, Ravi Kiran Sarvadevabhatla and R.Ven katesh Babu, "Deligan: Generative adversarial networks for diverse and limited data", In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp 166-174
- [5] Tommi S. Jaakkola and Michael I. Jordan, "Improving the mean field approximation via the use of mixture distributions", *NATO ASI Series D Behavioural Social Sci.*, volume 89, 1998, pp 528-534
- [6] 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", *Advances in neural information processing systems*, volume 30, 2017, pp 6626-6637
- [7] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford and Xi Chen, "Improved techniques for training gans", *Advances in neural information processing systems*, volume 29, 2016, pp 2226-2234
- [8] Diederik P. Kingma and Jimmy Lei Ba, "Adam: A method for stochastic optimization", *arXiv preprint arXiv:1412.6980*, 2014
- [9] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros, "image-to-image translation with conditional adversarial networks", In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)*, 2017, pp 1125-1134
- [10] Jun-Yan Zhu, Taesung Park, Phillip Isola, Alexei A. Efros, "Unpaired Image-To-Image Translation Using Cycle-Consistent Adversarial Networks", In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp 2223-2232
- [11] Martin Arjovsky, Léon Bottou, "Towards Principled Methods for Training Generative Adversarial Networks", In *Proc. ICLR*, 2017
- [12] Connor Shorten, Taghi M. Khoshgoftaar, "A survey on image data augmentation for deep learning", *Journal of Big Data*, volume 6, 2019
- [13] Dan Zhang, Anna Khoreva, "PA-GAN: Improving GAN Training by Progressive Augmentation", In *Proc. ICLR*, 2018
- [14] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, Xi Chen, Xi Chen, "Improved techniques for training gans", *Advances in neural information processing systems* 29, 2016, pp 2234-2242

- [15] Mehdi Mirza, Simon Osindero, "Conditional Generative Adversarial Nets", arXiv preprint arXiv:1411.1784, 2014
- [16] Xi Chen, Yan Duan, Rein Houthoofd, John Schulman, Ilya Sutskever, Pieter Abbeel, "InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets", 30th Conference on Neural Information Processing Systems, 2016, pp 2180-2188