GenCo: Generative Co-training for Generative Adversarial Networks with Limited Data

Kaiwen Cui[1]*, Jiaxing Huang[1]*, Zhipeng Luo[1], Gongjie Zhang[1], Fangneng Zhan[2], Shijian Lu[1]†

[1] School of Computer Science and Engineering, Nanyang Technological University
[2] S-lab, Nanyang Technological University
{kaiwen001, zhipeng001}@e.ntu.edu.sg, {jiaxing.huang, Gongjiezhang, fnzhan, shijian.lu}@ntu.edu.sg

Abstract

Training effective Generative Adversarial Networks (GANs) requires large amounts of training data, without which the trained models are usually sub-optimal with discriminator over-fitting. Several prior studies address this issue by expanding the distribution of the limited training data via massive and hand-crafted data augmentation. We handle data-limited image generation from a very different perspective. Specifically, we design GenCo, a Generative Co-training network that mitigates the discriminator over-fitting issue by introducing multiple complementary discriminators that provide diverse supervision from multiple distinctive views in training. We instantiate the idea of GenCo in two ways. The first way is Weight-Discrepancy Co-training (WeCo) which co-trains multiple distinctive discriminators by diversifying their parameters. The second way is Data-Discrepancy Co-training (DaCo) which achieves co-training by feeding discriminators with different views of the input images. Extensive experiments over multiple benchmarks show that GenCo achieves superior generation with limited training data. In addition, GenCo also complements the augmentation approach with consistent and clear performance gains when combined.

Introduction

Generative Adversarial Networks (GANs) (Goodfellow et al. 2014) have achieved great successes in various image generation tasks such as image-to-image translation (Zhu et al. 2017; Zhan, Zhu, and Lu 2019; Zhan et al. 2021), domain adaptation (Hoffman et al. 2018; Hsu et al. 2020) and image in-painting (Yu et al. 2019, 2021b,a). Nevertheless, high-fidelity image generation requires large amounts of training samples which are laborious and time-consuming to collect. Data-limited image generation, which aims to generate realistic and high-fidelity images with a small number of training samples, is a very meaningful yet challenging task for the widespread applications of image generation.

With limited training samples, the trained generation model suffers from discriminator over-fitting (Zhao et al. 2020; Karras et al. 2020a) which leads to degraded generation. Specifically, over-fitting discriminator produces very small discriminator loss as illustrated in the upper graph in Fig. 1. The very small discriminator loss then leads to very large generator loss and gradients which accumulate during training and lead to training divergence and degraded generation (Pascanu, Mikolov, and Bengio 2012, 2013) as illustrated in the lower graph in Fig. 1. The over-fitting issue has attracted increasing interest recently, and the prevalent approach addresses the issue through massive data augmentation. The idea is to massively augment the limited training samples to expand the data distributions as much as possible. Though prior studies (Karras et al. 2020a; Zhao et al. 2020)

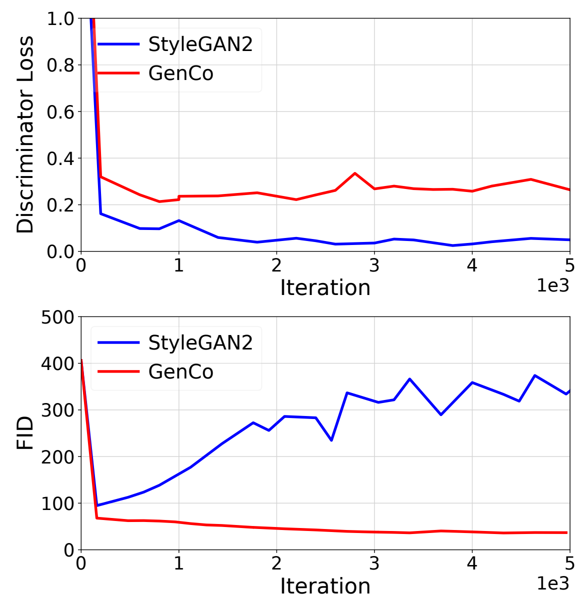

Figure 1: The proposed GenCo improves data-limited image generation clearly (on 100-shot Obama): With limited training samples, discriminator in most GANs such as StyleGAN2 tends to become over-fitting and produces very small discriminator loss as shown in the upper graph. The very small discriminator loss further leads to very large generator loss as well as gradients which cause training to diverge and generation to deteriorate as shown in the lower graph. The proposed GenCo mitigates the discriminator over-fitting effectively with more stable training and better generation.

*indicates equal contribution.
†corresponding author.

demonstrate the effectiveness of this approach, they address the problem at the input end only without considering much about network training.

We tackle the over-fitting issue from a very different perspective. Specifically, we introduce the idea of co-training into the data-limited image generation task, aiming to learn limited data from multiple distinctive yet complementary views. Co-training was originally proposed to boost the inspection performance when only limited data is available (Blum and Mitchell 1998). It alleviates the data constraint effectively by employing multiple classifiers that learn from different views and capture complementary information about the limited data. In recent years, co-training has been adopted in different deep learning tasks such as semi-supervised image recognition (Qiao et al. 2018), unsupervised domain adaptation (Saito et al. 2018; Luo et al. 2019), etc., where the amount of training data becomes more critical as compared with traditional learning tasks without using deep neural networks.

Specifically, we design GenCo, a Generative Co-training network that adapts the idea of co-training into data-limited image generation for tackling its inherent over-fitting issue. GenCo trains the generator with multiple discriminators that mitigate the over-fitting issue by learning from multiple distinct yet complementary views of the limited data. We design two instances of GenCo that enable the discriminators to learn from distinctive and comprehensive views. The first is Weight-Discrepancy Co-training (WeCo) which co-trains multiple distinctive discriminators by diversifying their parameters with a weight discrepancy loss. The second is Data-Discrepancy Co-training (DaCo) that co-trains distinctive discriminators by feeding them with different views of the input images. The proposed GenCo mitigates the discriminator over-fitting issue and improves data-limited image generation effectively as illustrated in Fig. 1, more details to be discussed in the Experiments section.

The contribution of this work can be summarized in three aspects. *First*, we propose to tackle the data-limited image generation challenge from a co-training perspective. To this end, we design GenCo, a Generative Co-training network that mitigates the discriminator over-fitting issue effectively by training the generator with multiple distinctive discriminators. *Second*, we design two instances of GenCo that are complementary to each other, namely, WeCo that introduces weight discrepancy loss to diversify multiple discriminators and DaCo that learns distinctive discriminators by employing different views of input images. *Third*, extensive experiments show that GenCo achieves superior generation quality and it is also complementary with the state-of-the-art augmentation and regularization approaches with consistent performance gains.

Related Works

Generative Adversarial Networks (GANs): The pioneer GANs (Goodfellow et al. 2014) greatly changes the paradigm of automated image generation. Leveraging this idea, quite a few GANs have been developed for realistic image generation in the past few years. They strive to improve the generation realism and fidelity from different aspects by introducing task-specific training objectives (Arjovsky, Chintala, and Bottou 2017), sophisticated network architectures (Miyato and Koyama 2018; Wu et al. 2020; Cui et al. 2021), and different training strategies (Liu et al. 2020), etc. On the other hand, most existing GANs still require a large number of training images for capturing the data distributions comprehensively. When only a limited number of training images are available, they often suffer from clear discriminator over-fitting and degraded generation.

We target data-limited image generation, which strives to learn robust generation models from limited training images yet without sacrificing much generation quality.

Data-Limited Image Generation: Data-limited image generation has attracted increasing interest for mitigating the laborious image collection process. The earlier studies (Webster et al. 2019; Gulrajani, Raffel, and Metz 2020) suggest that one of the main obstacles of training GANs with limited training data is the discriminator over-fitting. The recent studies strive to address the issue through massive data augmentation. For example, Zhao et al. (2020) introduces differentiable augmentation to stabilize the network training which leads to a clear improvement in generation. Karras et al. (2020a) presents an adaptive augmentation mechanism that prevents leaking of augmentation to generated images.

In this paper, we tackle the discriminator over-fitting issue from a different perspective and propose Generative Co-training that employs the idea of co-training to view the limited data from multiple complementary views.

Co-training: Co-training aims to learn multiple complementary information from different views for training more generalizable models. The idea traces back to a few pioneer studies (Blum and Mitchell 1998; Sun and Jin 2011; Yu et al. 2011) that propose co-training to tackle the data insufficiency problem while training classification models. With the recent advance of deep neural networks and demands for larger amounts of training data, the idea of co-training has attracted increasing interest in various deep network training tasks. For example, Qiao et al. (2018) and Saito et al. (2018) co-train networks with multiple classifiers while Huang et al. (2021a) conducts co-training with multiple tasks.

We introduce co-training into the data-limited image generation task for mitigating its inherent over-fitting issue during adversarial training. To the best of our knowledge, this is the first work that explores the discriminative co-training idea for the generative image generation task.

Method

This section describes the detailed methodology of the proposed GenCo. As illustrated in Fig. 2, we co-train multiple distinctive discriminators to mitigate the over-fitting issue. In addition, we design two instances of GenCo, including a Weight-Discrepancy Co-training (WeCo) that trains multiple distinctive discriminators by diversifying their parameters and a Data-Discrepancy Co-training (DaCo) that trains multiple distinctive discriminators by feeding them with different views of training images. We focus on two discriminators in WeCo and DaCo and will discuss the extension with more discriminators in Experiments. The ensuing subsections will describe the problem definition of data-limited

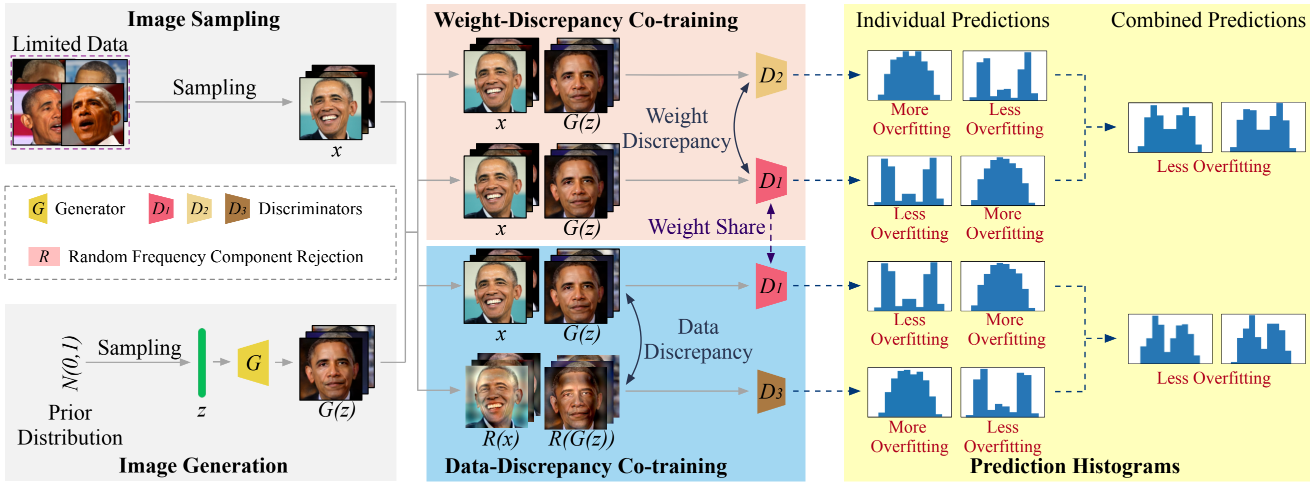

Figure 2: The architecture of the proposed GenCo: GenCo consists of four modules on Image Sampling, Image Generation, Weight-Discrepancy Co-training (WeCo) and Data-Discrepancy Co-training (DaCo). Image Sampling samples images x from limited training data and Image Generation generates images $G(z)$ with a generator G. x and $G(z)$ are fed to WeCo to co-train discriminators D_1 and D_2 which are differentiated by a weight discrepancy loss. They are also fed to DaCo to co-train discriminators D_1 and D_3, where a different view of data ($i.e.$, $R(x)$ and $R(G(z))$) is fed to D_3. The box on the right shows six groups of prediction histograms. The left four groups are produced by D_1 (2 with shared weights), D_2 and D_3, and the right two groups are the combined prediction histograms by WeCo and DaCo, respectively. The horizontal axis of these histograms shows the discriminator score in [-4, 4] and the vertical axis shows the numbers of occurrence. The four distinctive yet complementary discriminators capture different information of the training images, and the fusion of them with more comprehensive information mitigates the discriminator overfitting issue effectively.

image generation, the network architecture of GenCo, details of the proposed WeCo and DaCo, and the overall training objective, respectively.

Problem Definition

The GAN models are the cornerstone techniques for image generation tasks. Each GAN consists of a discriminator D and a generator G. The general loss function for discriminator and generator is defined as:

$$\mathcal{L}_d(D; x, G(z)) = \mathbb{E}[\log(D(x))] + \mathbb{E}[\log(1 - D(G(z))] \tag{1}$$

$$\mathcal{L}_g(D; G(z)) = \mathbb{E}[\log(1 - D(G(z))] \tag{2}$$

where $\mathcal{L}_d$ and $\mathcal{L}_g$ denote the discriminator and generator losses, respectively. x denotes a training sample and z is sampled from a prior distribution.

With limited training data X_L, discriminator in GANs tends to become over-fitting, leading to sub-optimal image generation. Concretely, the over-fitting discriminator produces high prediction scores and very small discriminator loss $\mathcal{L}_d$. The very small discriminator loss leads to very large generator loss $\mathcal{L}_g$ as well as gradients which accumulate during training and further cause training divergence and degraded generation. The following subsections describe how the proposed GenCo mitigates the discriminator over-fitting.

Overview of Network Architecture

GenCo consists of four major modules as demonstrated in Fig. 2: Image Sampling, Image Generation, Weight-Discrepancy Co-training (WeCo) and Data-Discrepancy Co-training (DaCo). Image Sampling samples images x from the limited dataset X_L and Image Generation generates fake samples $G(z)$ from a prior distribution with generator G. x and $G(z)$ are then passed to WeCo to co-train discriminators D_1 and D_2 that are differentiated by a weight discrepancy loss. Meanwhile, x and $G(z)$ are also fed to DaCo to co-train discriminators D_1 and D_3 that are differentiated by distinctive views of the inputs.

Weight-Discrepancy Co-training

The proposed WeCo aims to learn two distinctive discriminators D_1 and D_2 by diversifying their parameters. We achieve diverse parameters by defining a weight discrepancy loss $\mathcal{L}_{wd}$ that minimizes the cosine distance between the weights of D_1 and D_2:

$$\mathcal{L}_{wd}(D_1, D_2) = \frac{\overrightarrow{W_{D_1}}\overrightarrow{W_{D_2}}}{|\overrightarrow{W_{D_1}}||\overrightarrow{W_{D_2}}|} \tag{3}$$

where $\overrightarrow{W_{D_1}}$ and $\overrightarrow{W_{D_2}}$ are the weights of D_1 and D_2. The loss of D_1 and D_2 can thus be formulated by:

$$\mathcal{L}_{D_1} = \mathcal{L}_d(D_1; x, G(z)) \tag{4}$$

$$\mathcal{L}_{D_2} = \mathcal{L}_d(D_2; x, G(z)) + \mathcal{L}_{wd}(D_1, D_2) \tag{5}$$

Methods	Massive Augmentation	Pre-training w/ 70K images	100-shot Obama	100-shot Grumpy Cat	100-shot Panda	AFHQ Cat	AFHQ Dog
Scale/shift	No	Yes	50.72	34.20	21.38	54.83	83.04
MineGAN	No	Yes	50.63	34.54	14.84	54.45	93.03
TransferGAN	No	Yes	48.73	34.06	23.20	52.61	82.38
TransferGAN + DA	Yes	Yes	39.85	29.77	17.12	49.10	65.57
FreezeD	No	Yes	41.87	31.22	17.95	47.70	70.46
DA	Yes	No	46.87	27.08	12.06	42.44	58.85
ADA	Yes	No	45.69	26.62	12.90	40.77	56.83
LeCam-GAN	Yes	No	33.16	24.93	10.16	34.18	54.88
DA + **GenCo**	Yes	No	**32.21**	**17.79**	**9.49**	**30.89**	**49.63**
StyleGAN2	No	No	80.20	48.90	34.27	71.71	130.19
LeCam-GAN	No	No	38.58	41.38	19.88	60.26	112.39
GenCo	No	No	**36.35**	**33.57**	**15.50**	**54.78**	**94.47**

Table 1: Comparison with the state-of-the-arts over 100-shot and AFHQ: Training with 100 (Obama, Grumpy Cat and Panda), 160 (AFHQ Cat), and 389 (AFHQ Dog) samples, GenCo performs the best consistently. It achieves comparable results as transfer learning methods (Rows 1-5) pre-trained with 70K images. We report FIDs (↓) averaged over three runs.

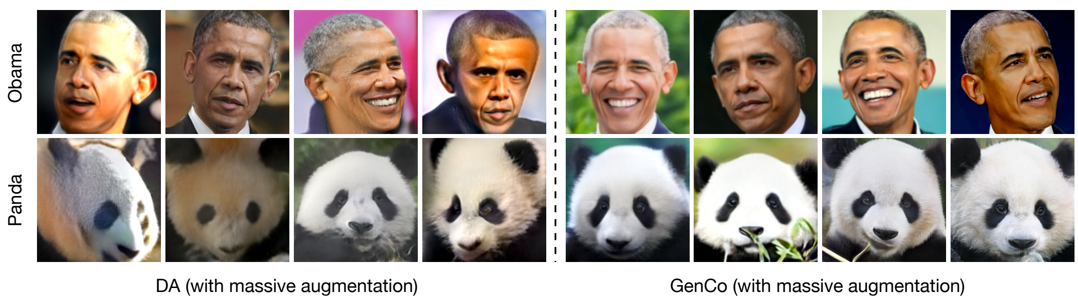

Figure 3: Qualitative results over 100-shot datasets ($e.g.$, Obama and Panda): The generation by GenCo is clearly more realistic than that by DA (Zhao et al. 2020), the state-of-the-art data-limited generation method.

where L_d is the general discriminator loss as in Eq.1. $\mathcal{L}_{wd}$ is the weight discrepancy loss as defined in Eq.3. We apply $\mathcal{L}_{wd}$ on only one discriminator for simplicity because applying it on two discriminators does not make much difference.

The overall WeCo loss $\mathcal{L}^{WeCo}_{D_1,D_2}$ can thus be defined by:

$$\mathcal{L}^{WeCo}_{D_1,D_2} = \mathcal{L}_{D_1} + \mathcal{L}_{D_2} \tag{6}$$

Data-Discrepancy Co-training

DaCo co-trains two distinctive discriminators D_1 and D_3 that take different views of the input images. Specifically, D_1 is fed with the original images while D_3 takes partial frequency components (FCs) of the input images (generated by Random Frequency Component Filter (R)) as input.

The component R consists of three processes including R_t, R_r, and $R_{t^{-1}}$. Specifically, R_t first converts the images x to frequency space. R_r then filters some FCs randomly with the other FCs passed(Yang and Soatto 2020; Huang et al. 2021b,c). Finally, $R_{t^{-1}}$ converts the passed FCs back to RGB images to form the new inputs of D_3. Detailed definitions of $R_t, R_r, R_{t^{-1}}$ are available in the supplementary material. Note the percentage of the filtered FCs is controlled by a hyper-parameter P which is empirically set at 0.2.

The loss functions of D_1 and D_3 can thus be defined by:

$$\mathcal{L}_{D_1} = \mathcal{L}_d(D_1; x, G(z)) \tag{7}$$

$$\mathcal{L}_{D_3} = \mathcal{L}_d(D_3; R(x), R(G(z))) \tag{8}$$

where the loss of D_1 is the same as the loss of D_1 (Eq. 4) in WeCo (they share weights). The loss of D_3 is close to that of D_1 and the differences are largely due to the different inputs by the Random Frequency Component Filter (R).

The overall DaCo loss $\mathcal{L}^{DaCo}_{D_1,D_3}$ can thus be defined by:

$$\mathcal{L}^{DaCo}_{D_1,D_3} = \mathcal{L}_{D_1} + \mathcal{L}_{D_3} \tag{9}$$

Overall Training Objective

The generator G learns with information from all three discriminators. Its loss $\mathcal{L}^{total}_G$ can be formulated by:

$$\mathcal{L}^{total}_G = \mathcal{L}_g(D_1; G(z)) + \mathcal{L}_g(D_2; G(z)) + \mathcal{L}_g(D_3; R(G(z))) \tag{10}$$

The overall training objective of the proposed GenCo can thus be formulated by,

$$\underset{G}{min} \underset{D_1,D_2,D_3}{max} \mathcal{L}^{total}_G + \mathcal{L}^{WeCo}_{D_1,D_2} + \mathcal{L}^{DaCo}_{D_1,D_3} \tag{11}$$

Methods	CIFAR-10			CIFAR-100		
	100% data	20% data	10% data	100% data	20% data	10% data
Non-saturated GAN	9.83±0.06	18.59±0.15	41.99±0.18	13.87±0.08	32.64 ±0.19	70.5±0.38
LS-GAN	9.07±0.01	21.60±0.11	41.68±0.18	12.43±0.11	27.09±0.09	54.69±0.12
RAHinge GAN	11.31±0.04	23.90±0.22	48.13±0.33	14.61±0.21	28.79±0.17	52.72±0.18
StyleGAN2 + DA	9.89±0.04	12.15±0.09	14.50±0.08	15.22±0.01	16.65±0.06	20.75±0.09
StyleGAN2 + DA **+ GenCo**	**6.57**±0.01	**8.02**±0.05	**10.61**±0.06	**8.27**±0.01	**10.14**±0.04	**13.48**±0.06
BigGAN + DA	8.75±0.03	14.53±0.10	23.34±0.09	11.99±0.02	22.55±0.06	35.39±0.08
BigGAN + DA **+ GenCo**	**7.98**±0.02	**12.61**±0.05	**18.10**±0.13	**10.92**±0.02	**18.44**±0.04	**25.22**±0.06
StyleGAN2	11.07±0.03	23.08±0.11	36.02±0.15	16.54±0.04	32.30 ±0.11	45.87±0.15
GenCo	**8.87**±0.02	**15.65**±0.08	**27.16**±0.12	**9.50**±0.03	**18.77**±0.07	**33.88**±0.09
BigGAN	9.07±0.06	21.86±0.29	48.08±0.10	13.60±0.07	32.99±0.24	66.71±0.01
GenCo	**8.83**±0.04	**16.57**±0.08	**28.08**±0.11	**11.90**±0.02	**26.15**±0.08	**40.98**±0.09

Table 2: Comparing GenCo with the state-of-the-arts over CIFAR: GenCo mitigates the discriminator over-fitting issue and outperforms the state-of-the-arts consistently. We report FID (↓) scores averaged over three runs.

Methods	FFHQ				LSUN-Cat			
	30K	10K	5K	1K	30K	10K	5K	1K
StyleGAN2	12.30	30.79	49.66	100.13	15.92	50.63	96.44	186.88
GenCo	**8.27**	**15.66**	**27.96**	**65.31**	**12.25**	**20.15**	**40.79**	**140.08**

Table 3: Quantitative results on the FFHQ and LSUN-Cat datasets : We report FID (↓) over three runs.

Why is GenCo effective? In data-limited image generation, one major issue is that discriminator in GANs tends to suffer from over-fitting by capturing simple structures and patterns only (Bau et al. 2019; Zhang et al. 2021). The proposed GenCo mitigates this issue by co-training two discriminators in WeCo and DaCo. With the co-training design, although one discriminator ($e.g.$, D_1) may overfit and focuses on learning simple structures and patterns, the other distinctive discriminator ($e.g.$, D_2 in WeCo and D_3 in DaCo) with different parameters or data inputs will be encouraged to learn different information like complex structures and patterns. The two discriminators thus complement each other to focus on different types of information, which helps mitigate the discriminator over-fitting issue effectively (as shown in Fig.2). From another view, the intrinsic cause of the discriminator over-fitting is the large generator loss that leads to training divergence. In GenCo, the overall over-fitting with two distinctive discriminators in either WeCo or DaCo is reduced which leads to smaller generator loss and further mitigates training divergence.

In addition, WeCo and DaCo in GenCo also complement each other to mitigate the overall over-fitting as they achieve co-training from different perspectives. Specifically, WeCo achieves co-training by diversifying the discriminator parameters, whereas DaCo achieves co-training by feeding two discriminators with different views of the inputs.

Experiments

In this section, we conduct extensive experiments to evaluate our proposed GenCo. We first briefly introduce the datasets and evaluation metrics used in our experiments. We then benchmark GenCo across these datasets and provide a visualization of GenCo. Moreover, we conduct extensive ablation studies and discussions to support our design choices.

Datasets and Evaluation Metrics

We conduct experiments over multiple public datasets: CIFAR (Krizhevsky et al. 2009), 100-shot (Zhao et al. 2020), AFHQ (Si and Zhu 2011), FFHQ (Karras, Laine, and Aila 2019) and LSUN-Cat (Yu et al. 2015). We follow Zhao et al. (2020) and perform evaluations with FID (Heusel et al. 2017) and IS (Salimans et al. 2016).

Experiments on 100-shot and AFHQ

Rows 6-9 of Table 1 compare GenCo with state-of-the-art methods in data-limited image generation ($i.e.$, DA (Zhao et al. 2020), ADA (Karras et al. 2020a) and LeCam-GAN (Tseng et al. 2021)) over 100-shot and AFHQ. GenCo performs the best consistently, demonstrating the effectiveness of GenCo in mitigating discriminator over-fitting.

Table 1 (Rows 10 and 12) compares GenCo with state-of-the-art GANs ($i.e.$, StyleGAN2 (Karras et al. 2020b)). It shows that GenCo improves the generation consistently by large margins. In addition, several studies explore transfer learning by pre-training the model with large datasets ($i.e.$, Scale/shift (Noguchi and Harada 2019), Mine-GAN (Wang et al. 2020), TransferGAN (Wang et al. 2018) and FreezeD (Mo, Cho, and Shin 2020)). The top part of Table 1 shows their FID scores (pre-trained with 70K FFHQ images). We can see that GenCo achieves comparable FIDs by using only 100 – 400 training samples instead. Fig. 3 qualitatively demonstrates that GenCo outperforms the state-of-the-art in data-limited generation, especially in terms of the generated shapes and textures.

Experiments on CIFAR-10 and CIFAR-100

Table 2 (Rows 1-3) compares GenCo with several state-of-the-art GANs ($i.e.$, Non-saturated GAN (Goodfellow et al. 2014), LS-GAN (Mao et al. 2017) and RAHinge GAN (Jolicoeur-Martineau 2018)) over CIFAR-10 and CIFAR-100. And rows 8-11 show the results of GenCo and two of its baselines ($i.e.$, StyleGAN2 and BigGAN (Brock,

Figure 4: Activation maps of discriminators in GenCo: GenCo mitigates discriminator over-fitting with three distinctive discriminators that capture complementary information. As illustrated, D_1, D_2, and D_3 attend to facial styles (color, brightness, etc.), facial details (wrinkles, face outline, etc.) and facial expressions (eyes, mouth, etc.), respectively.

Design Choice		Cifar-10		100shot
WeCo	DaCo	20% data	10% data	Obama
-	-	23.08	36.02	80.16
✓	-	20.16	32.55	55.34
-	✓	16.70	28.46	41.96
✓	✓	**15.65**	**27.16**	**36.28**

Table 4: Ablation study of GenCo: WeCo and DaCo in GenCo both mitigate discriminator over-fitting effectively with improved generation over the baseline. GenCo performs simply the best as WeCo and DaCo are complementary to each other. The FIDs (↓) are averaged over three runs.

Donahue, and Simonyan 2018)). We can see that GenCo performs the best consistently. The superior performance is largely attributed to the co-training idea in GenCo which mitigates the discriminator over-fitting effectively.

Rows 4-7 of Table 2 compare GenCo with DA (Zhao et al. 2020), the state-of-the-art in data-limited generation. It shows that GenCo outperforms DA consistently under the massive augmentation setup, demonstrating the effectiveness of GenCo in mitigating discriminator over-fitting.

Experiments on FFHQ and LSUN-Cat

Table 3 quantitatively compares GenCo with StyleGAN2 over FFHQ and LSUN-Cat. Following DA, we perform experiments on 30K, 10K, 5K and 1K training samples. As Table 3 shows, GenCo improves the baseline consistently.

Visualization of GenCo

GenCo mitigates the discriminator over-fitting effectively by co-training multiple distinctive discriminators (D_1 and D_2 in WeCo, D_1 and D_3 in DaCo) that learn from different views and capture complementary information. This can be observed from their activation maps (Selvaraju et al. 2017) in Fig. 4, which show that the three discriminators attend and capture different visual information. The fusion of them provides more comprehensive supervision signals which lead to less discriminator over-fitting and better generation.

Ablation study

The proposed GenCo consists of two major components, namely, WeCo and DaCo. We study the two components separately to examine their contributions. As Table 4 shows, including either WeCo or DaCo outperforms the baseline clearly, demonstrating the effectiveness of the proposed co-training which mitigates discriminator over-fitting by learning from multiple distinctive views. In addition, combining WeCo and DaCo performs clearly the best which verifies that the distinctive views in WeCo (by weight discrepancy) and DaCo (by input discrepancy) complement to each other.

Qualitative ablation studies in Fig. 5 show that the proposed WeCo and DaCo can produce clearly more realistic generation than baseline. In addition, GenCo produces the most realistic generation, which verifies that WeCo and DaCo complement each other.

Figure 5: Qualitative ablation study over 100-shot Obama: The generation by WeCo (Row 2) and DaCo (Row 3) alone is clearly more realistic than the generation by the baseline (Row 1). In addition, the generation by GenCo (Row 4) that combines WeCo and GenCo is most realistic.

Methods	Baseline		+GenCo	
	FID (↓)	IS (↑)	FID (↓)	IS (↑)
BigGAN	48.08	7.09	**28.08**	**8.01**
+ noise	47.06	7.12	**27.88**	**8.06**
+ CR	44.16	7.27	**27.03**	**8.12**
+ GP-0	42.22	7.38	**26.58**	**8.15**
+ LeCam-GAN	35.23	7.97	**25.89**	**8.23**

Table 5: Experiments on GenCo and regularization-based generation methods: GenCo and regularization-based methods are clearly complementary in data-limited generation. The FIDs (↓) and IS (↑) are averaged over three runs.

Discussion

In this subsection, we analyze our GenCo from several perspectives. All the experiments are based on the CIFAR-10 dataset with 10% data unless specified otherwise.

Complementary with regularization methods: Existing regularization methods introduce a regularization term to network parameters or training losses to improve training

Methods	Baseline	+GenCo
StyleGAN2	36.02	**27.16**
BigGAN	48.08	**28.08**
LS-GAN	41.68	**26.64**
RAHinge GAN	48.13	**36.47**
BigGAN + DA	23.34	**18.10**

Table 6: Experiments on the generalization of GenCo with different baselines (FIDs (↓) averaged over three runs).

Metrics	Weight-Discrepancy	Number of discriminators			
		1	2	3	4
FID (↓)	x	36.02	35.44	33.56	**32.70**
IS (↑)	x	7.34	7.46	7.55	**7.65**
FID (↓)	✓	36.02	32.55	29.16	**27.28**
IS (↑)	✓	7.34	7.68	7.80	**8.03**

Table 7: Experiments with different number of discriminators in WeCo (results averaged over three evaluation runs).

stability and mitigate the discriminator over-fitting in data-limited image generation. The proposed GenCo addresses the same issue from a very different co-training perspective instead, which can complement these regularization approaches ($i.e.$, noise (Sønderby et al. 2016), CR (Zhang et al. 2019), GP-0 (Mescheder et al. 2018) and LeCam-GAN (Tseng et al. 2021)) effectively. Table 5 reveals that existing regularization methods do improve the generation clearly. Meanwhile, incorporating GenCo into them further improves the generation consistently by large margins.

Generalization of GenCo: The proposed GenCo can work with various baselines. Table 6 shows that GenCo improves the generation consistently while working with different baselines. The superior generalization is largely attributed to the co-training design in GenCo, which is independent of the network architectures and training losses.

More than two discriminators: WeCo and DaCo mitigate discriminator over-fitting by co-training two distinctive discriminators. The idea can be easily extended to more than two discriminators. As WeCo and DaCo share similar co-training idea, we evaluate on WeCo only with two different setups: the first simply employs multiple discriminators while the second employs the same number of discriminators but includes weight discrepancy loss between each pair of discriminators. As Table 7 shows, the generation improves consistently with the increasing number of discriminators. This is intuitive as more views from multiple discriminators help mitigate the over-fitting. In addition, including the discrepancy loss improves the generation consistently as well, indicating the effectiveness of the proposed weight discrepancy loss which encourages discriminators to learn more diverse and complementary information.

Effectiveness of DaCo: DaCo performs light data augmentation as R produces a new input for each input image. To demonstrate that DaCo works due to our co-training design instead of the light augmentation, we compare DaCo and its variant that employs R for augmentation only without co-training. Table 8 shows that DaCo achieves clearly better generation than employing R for augmentation only. This is largely because DaCo employs two distinctive views of the inputs to co-train two different discriminators to mitigate their over-fitting whereas the light augmentation alone does not expand the data distribution much.

Metrics	Baseline	R as augmentation	DaCo
FID (↓)	36.02	34.86	**28.46**
IS (↑)	7.34	7.51	**7.88**

Table 8: Experiments on the random frequency component filter R in Daco (results averaged over three runs).

Metrics	Percentage of filtered frequency components				
	0.1	0.2	0.3	0.4	0.5
FID (↓)	29.07	**27.16**	28.04	28.36	29.18
IS (↑)	7.83	**8.05**	7.92	7.87	7.81

Table 9: Experiments on the amount of filtered frequency components in DaCo (results averaged over three runs).

Robustness of DaCo: We introduce a hyper-parameter P in DaCo to control the percentage of filtered frequency components (FCs). We perform experiments to study how different P affect the generation performance. As shown in Table 9, different P produce quite similar FID. We conjecture that the random filtering of different FCs in each input creates sufficient distinctive views which makes P not that sensitive to the overall generation performance.

Conclusion

This paper presents a novel Generative Co-training (GenCo) network that adapts the co-training idea into data-limited generation for tackling its inherent over-fitting issue. We propose two instances of GenCo, namely, Weight-Discrepancy Co-training (WeCo) and Data-Discrepancy Co-training (DaCo). WeCo co-trains multiple distinctive discriminators by diversifying their parameters with a weight discrepancy loss. DaCo achieves co-training by feeding two discriminators with different views of the inputs. We demonstrate that both instances can improve the generation performance and combining WeCo and DaCo achieves the best results. We also show that our GenCo complements state-of-the-art data-augmentation and regularization methods. In addition, GenCo is a general technique for limited-data GANs, which could be applied to various downstream tasks.

Acknowledgements

This work is conducted within the Delta-NTU Corporate Lab for Cyber-Physical Systems with funding support from Delta Electronics Inc. and the National Research Foundation (NRF) Singapore under the Corp Lab @ University Scheme (Project No.: DELTA-NTU CORP-SMA-RP15). This study is also partially supported under the RIE2020 Industry Alignment Fund – Industry Collaboration Projects (IAF-ICP) Funding Initiative, as well as cash and in-kind contribution from the industry partner(s).

References

Arjovsky, M.; Chintala, S.; and Bottou, L. 2017. Wasserstein generative adversarial networks. In *International conference on machine learning*, 214–223. PMLR.

Bau, D.; Zhu, J.-Y.; Wulff, J.; Peebles, W.; Strobelt, H.; Zhou, B.; and Torralba, A. 2019. Seeing what a gan cannot generate. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4502–4511.

Blum, A.; and Mitchell, T. 1998. Combining labeled and unlabeled data with co-training. In *Proceedings of the eleventh annual conference on Computational learning theory*, 92–100.

Brock, A.; Donahue, J.; and Simonyan, K. a. 2018. Large scale GAN training for high fidelity natural image synthesis. *arXiv preprint arXiv:1809.11096*.

Cui, K.; Zhang, G.; Zhan, F.; Huang, J.; and Lu, S. 2021. FBC-GAN: Diverse and Flexible Image Synthesis via Foreground-Background Composition. *arXiv preprint arXiv:2107.03166*.

Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative adversarial nets. *Advances in neural information processing systems*, 27.

Gulrajani, I.; Raffel, C.; and Metz, L. 2020. Towards GAN benchmarks which require generalization. *arXiv preprint arXiv:2001.03653*.

Heusel, M.; Ramsauer, H.; Unterthiner, T.; Nessler, B.; and Hochreiter, S. 2017. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in neural information processing systems*, 6626–6637.

Hoffman, J.; Tzeng, E.; Park, T.; Zhu, J.-Y.; Isola, P.; Saenko, K.; Efros, A.; and Darrell, T. 2018. Cycada: Cycle-consistent adversarial domain adaptation. In *International conference on machine learning*, 1989–1998. PMLR.

Hsu, H.-K.; Yao, C.-H.; Tsai, Y.-H.; Hung, W.-C.; Tseng, H.-Y.; Singh, M.; and Yang, M.-H. 2020. Progressive domain adaptation for object detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 749–757.

Huang, J.; Guan, D.; Xiao, A.; and Lu, S. 2021a. Cross-view regularization for domain adaptive panoptic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10133–10144.

Huang, J.; Guan, D.; Xiao, A.; and Lu, S. 2021b. Fsdr: Frequency space domain randomization for domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 6891–6902.

Huang, J.; Guan, D.; Xiao, A.; and Lu, S. 2021c. Rda: Robust domain adaptation via fourier adversarial attacking. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 8988–8999.

Jolicoeur-Martineau, A. 2018. The relativistic discriminator: a key element missing from standard GAN. *arXiv preprint arXiv:1807.00734*.

Karras, T.; Aittala, M.; Hellsten, J.; Laine, S.; Lehtinen, J.; and Aila, T. 2020a. Training generative adversarial networks with limited data. *arXiv preprint arXiv:2006.06676*.

Karras, T.; Laine, S.; and Aila, T. 2019. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4401–4410.

Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; and Aila, T. 2020b. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8110–8119.

Liu, S.; Wang, T.; Bau, D.; Zhu, J.-Y.; and Torralba, A. 2020. Diverse image generation via self-conditioned gans. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14286–14295.

Luo, Y.; Zheng, L.; Guan, T.; Yu, J.; and Yang, Y. 2019. Taking a closer look at domain shift: Category-level adversaries for semantics consistent domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2507–2516.

Mao, X.; Li, Q.; Xie, H.; Lau, R. Y.; Wang, Z.; and Paul Smolley, S. 2017. Least squares generative adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, 2794–2802.

Mescheder, L.; Geiger, A.; Nowozin, S.; and ... 2018. Which training methods for GANs do actually converge? In *International conference on machine learning*, 3481–3490. PMLR.

Miyato, T.; and Koyama, M. 2018. cGANs with projection discriminator. *arXiv preprint arXiv:1802.05637*.

Mo, S.; Cho, M.; and Shin, J. 2020. Freeze Discriminator: A Simple Baseline for Fine-tuning GANs. *arXiv preprint arXiv:2002.10964*.

Noguchi, A.; and Harada, T. 2019. Image generation from small datasets via batch statistics adaptation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2750–2758.

Pascanu, R.; Mikolov, T.; and Bengio, Y. 2012. Understanding the exploding gradient problem. *CoRR, abs/1211.5063*, 2(417): 1.

Pascanu, R.; Mikolov, T.; and Bengio, Y. 2013. On the difficulty of training recurrent neural networks. In *International conference on machine learning*, 1310–1318. PMLR.

Qiao, S.; Shen, W.; Zhang, Z.; Wang, B.; and Yuille, A. 2018. Deep co-training for semi-supervised image recognition. In *Proceedings of the european conference on computer vision (eccv)*, 135–152.

Saito, K.; Watanabe, K.; Ushiku, Y.; and Harada, T. 2018. Maximum classifier discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 3723–3732.

Salimans, T.; Goodfellow, I.; Zaremba, W.; Cheung, V.; Radford, A.; and Chen, X. 2016. Improved techniques for training gans. In *Advances in neural information processing systems*, 2234–2242.

Selvaraju, R. R.; Cogswell, M.; Das, A.; Vedantam, R.; Parikh, D.; and Batra, D. 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*, 618–626.

Si, Z.; and Zhu, S.-C. 2011. Learning hybrid image templates (HIT) by information projection. *IEEE Transactions on pattern analysis and machine intelligence*, 34(7): 1354–1367.

Sønderby, C. K.; Caballero, J.; Theis, L.; Shi, W.; and Huszár, F. 2016. Amortised map inference for image super-resolution. *arXiv preprint arXiv:1610.04490*.

Sun, S.; and Jin, F. 2011. Robust co-training. *International Journal of Pattern Recognition and Artificial Intelligence*, 25(07): 1113–1126.

Tseng, H.-Y.; Jiang, L.; Liu, C.; Yang, M.-H.; and Yang, W. 2021. Regularizing Generative Adversarial Networks under Limited Data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 7921–7931.

Wang, Y.; Gonzalez-Garcia, A.; Berga, D.; Herranz, L.; Khan, F. S.; and Weijer, J. v. d. 2020. Minegan: effective knowledge transfer from gans to target domains with few images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9332–9341.

Wang, Y.; Wu, C.; Herranz, L.; van de Weijer, J.; Gonzalez-Garcia, A.; and Raducanu, B. 2018. Transferring gans: generating images from limited data. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 218–234.

Webster, R.; Rabin, J.; Simon, L.; and Jurie, F. 2019. Detecting overfitting of deep generative networks via latent recovery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11273–11282.

Wu, R.; Zhang, G.; Lu, S.; and Chen, T. 2020. Cascade ef-gan: Progressive facial expression editing with local focuses. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5021–5030.

Yang, Y.; and Soatto, S. 2020. Fda: Fourier domain adaptation for semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4085–4095.

Yu, F.; Seff, A.; Zhang, Y.; Song, S.; Funkhouser, T.; and Xiao, J. 2015. Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop. *arXiv preprint arXiv:1506.03365*.

Yu, J.; Lin, Z.; Yang, J.; Shen, X.; Lu, X.; and Huang, T. S. 2019. Free-form image inpainting with gated convolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 4471–4480.

Yu, S.; Krishnapuram, B.; Rosales, R.; and Rao, R. B. 2011. Bayesian co-training. *The Journal of Machine Learning Research*, 12: 2649–2680.

Yu, Y.; Zhan, F.; Lu, S.; Pan, J.; Ma, F.; Xie, X.; and Miao, C. 2021a. WaveFill: A Wavelet-based Generation Network for Image Inpainting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 14114–14123.

Yu, Y.; Zhan, F.; Wu, R.; Pan, J.; Cui, K.; Lu, S.; Ma, F.; Xie, X.; and Miao, C. 2021b. Diverse image inpainting with bidirectional and autoregressive transformers. *arXiv preprint arXiv:2104.12335*.

Zhan, F.; Yu, Y.; Cui, K.; Zhang, G.; Lu, S.; Pan, J.; Zhang, C.; Ma, F.; Xie, X.; and Miao, C. 2021. Unbalanced Feature Transport for Exemplar-based Image Translation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 15028–15038.

Zhan, F.; Zhu, H.; and Lu, S. 2019. Spatial fusion gan for image synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 3653–3662.

Zhang, G.; Cui, K.; Hung, T.-Y.; and Lu, S. 2021. Defect-GAN: High-fidelity defect synthesis for automated defect inspection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2524–2534.

Zhang, H.; Zhang, Z.; Odena, A.; and Lee, H. 2019. Consistency regularization for generative adversarial networks. *arXiv preprint arXiv:1910.12027*.

Zhao, S.; Liu, Z.; Lin, J.; Zhu, J.-Y.; and Han, S. 2020. Differentiable augmentation for data-efficient gan training. *arXiv preprint arXiv:2006.10738*.

Zhu, J.-Y.; Park, T.; Isola, P.; and Efros, A. A. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, 2223–2232.