InfoMax-GAN: Mutual Information Maximization for Improved Adversarial Image Generation

Kwot Sin Lee[1,2] Ngoc-Trung Tran[1] Ngai-Man Cheung[1]

[1]Singapore University of Technology and Design (SUTD)
[2]Wolfson College, University of Cambridge

Overview

We investigate the use of an InfoMax objective in GANs, in order to mitigate two key issues of GANs: catastrophic forgetting of the discriminator [1] and mode collapse of the generator.

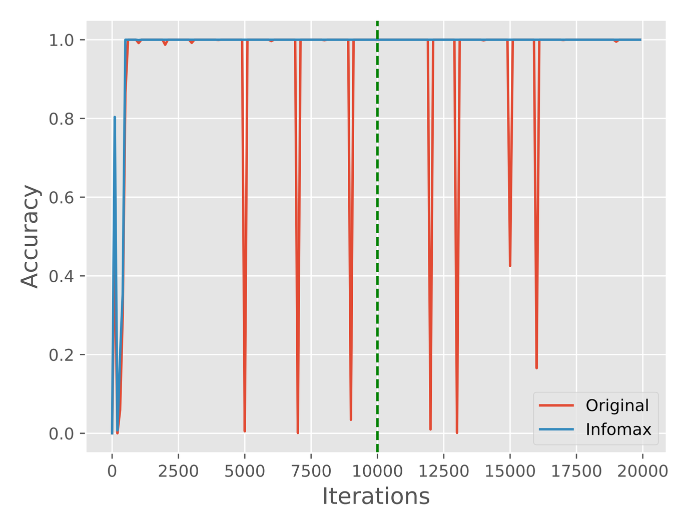
To do so, we maximize the mutual information between the local and global features of a discriminator encoder, which can be shown to be a lower bound of the InfoMax objective [4]. We follow the approach in Deep InfoMax [2] for obtaining the local/global features, and maximize the mutual information using the InfoNCE task [3].

Addressing GAN Issues

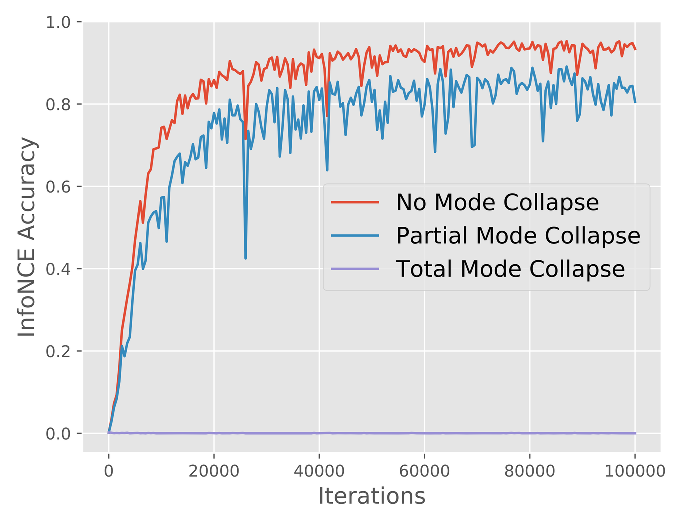
Catastrophic forgetting. Similar to [1], we train a classifier on the one-vs-all CIFAR-10 classification task, where the underlying class distribution changes every 1K iteration. From (a) we show that using the InfoMax objective prevents overfitting of the classifier to any one class distribution, allowing it to remember all prior classes it was trained on. In a GAN setting, this helps stabilize the non-stationary training environment where the generated image distribution always changes.

Mode collapse. We simulate 3 different kinds of generators using CIFAR-10 test data: (i) a perfect generator with no mode collapse that can generate all classes of images; (ii) a partially mode collapsed generator that can only generate one class of images; and (iii) a totally mode collapsed generator that can only generate one image. From (b), we observe the following:

- The perfect generator produces higher InfoNCE task accuracy than a partially mode collapsed one, due to a greater diversity of images.
- The totally mode collapse generator has near 0 accuracy: each positive sample for InfoNCE task now has multiple negative samples identical to it, thus making the task very difficult to solve.
- Thus, to solve InfoNCE task well, the generator is encouraged to produce images covering more modes.

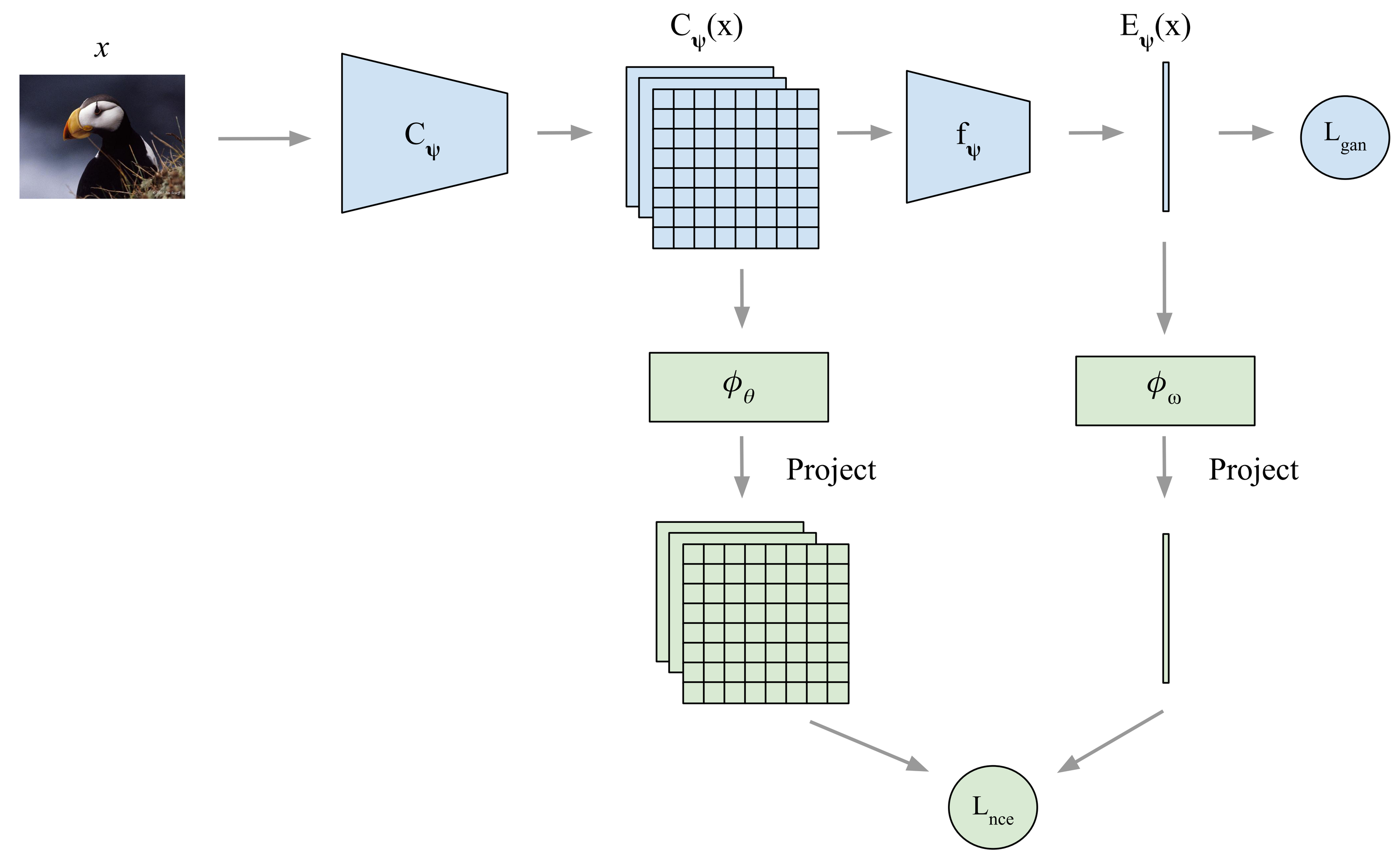


(a) Catastrophic forgetting.



(b) Mode collapse.

This Work



1. Send an input real/fake image x from the training data or the generator.
2. Image x passes through a discriminator encoder $E_\psi = f_\psi \circ C_\psi$, producing local feature map $C_\psi(x)$ and global feature vector $E_\psi(x)$.
3. Networks ψ_θ and ψ_ω project local and global features to a higher dimension to compute the InfoNCE loss for the InfoMax objective.
4. Discriminator is trained on this auxiliary loss, with no modifications made to the original GAN architecture.

Training Objectives:

$$L_D = L_{gan}(D, \hat{G}) + \alpha L_{nce}(X_r) \quad (1)$$
$$L_G = L_{gan}(\hat{D}, G) + \beta L_{nce}(X_g) \quad (2)$$

$$\begin{aligned} L_{nce}(X) &= -\mathbb{E}_{x \in X} \mathbb{E}_{i \in \mathcal{A}} \left[\log p(C_\psi^{(i)}(x), E_\psi(x) \mid X) \right] \\ &= -\mathbb{E}_{x \in X} \mathbb{E}_{i \in \mathcal{A}} \left[\log \frac{\exp(g_{\theta,\omega}(C_\psi^{(i)}(x), E_\psi(x)))}{\sum_{x' \in X} \exp(g_{\theta,\omega}(C_\psi^{(i)}(x'), E_\psi(x)))} \right] \end{aligned} \quad (3)$$

- L_{gan} is the hinge loss for GANs.
- L_{nce} is the InfoNCE loss for the discriminator, and α and β are hyperparameters.
- C_ψ and E_ψ are local and global feature encoders part of the same discriminator architecture, and $g_{\theta,\omega}$ contains networks to project the features.

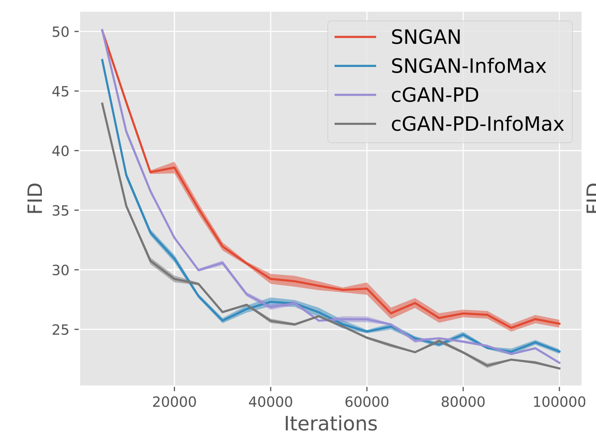
Experimental Setup

- **Metric:** Mean Fréchet Inception Distance (FID) computed using 50K real images and 10K fake images, over 3 different random seeds.
- **Datasets:** CIFAR-10 (32×32), ImageNet (32×32), STL-10 (48×48).
- **Models:** Spectral Normalization GAN (SNGAN) and Conditional GAN with Projection Discriminator (cGAN-PD).

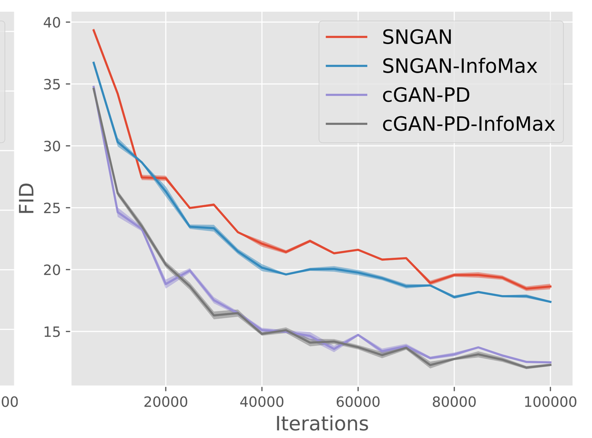
Results

Datasets	SNGAN		cGAN-PD	
	Original	InfoMax	Original	InfoMax
CIFAR-10	18.45 ± 0.19	$\mathbf{17.39 \pm 0.08}$	12.50 ± 0.07	$\mathbf{12.08 \pm 0.13}$
STL-10	39.94 ± 0.06	$\mathbf{38.85 \pm 0.22}$	-	-
ImageNet	25.13 ± 0.32	$\mathbf{23.12 \pm 0.27}$	22.19 ± 0.08	$\mathbf{21.72 \pm 0.06}$

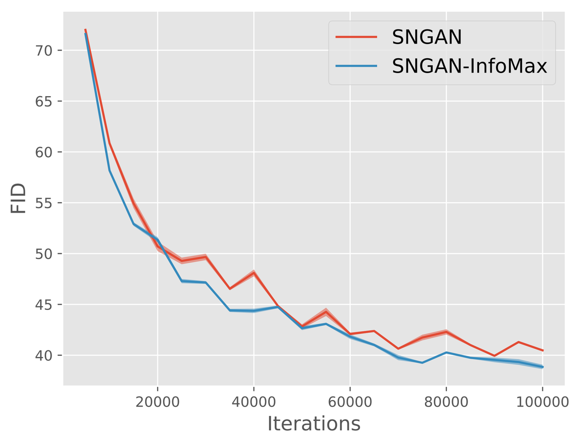
Table: Mean FID scores with standard deviation of all models across different datasets. For cGAN-PD, we test it on only CIFAR-10 and ImageNet, where labels are available.



(c) ImageNet



(d) CIFAR-10



(e) STL-10

Figure: FID curves for all models and datasets. The InfoMax objective allows for a faster and consistent improvement in FID.

Key References

[1] Ting Chen, Xiaohua Zhai, Marvin Ritter, Mario Lucic, and Neil Houlsby. Self-supervised generative adversarial networks. *arXiv preprint arXiv:1811.11212*, 2018.

[2] R Devon Hjelm, Alex Fedorov, Samuel Lavoie-Marchildon, Karan Grewal, Phil Bachman, Adam Trischler, and Yoshua Bengio. Learning deep representations by mutual information estimation and maximization. *arXiv preprint arXiv:1808.06670*, 2018.

[3] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

[4] Michael Tschannen, Josip Djolonga, Paul K Rubenstein, Sylvain Gelly, and Mario Lucic. On mutual information maximization for representation learning. *arXiv preprint arXiv:1907.13625*, 2019.