CVPR SEATTLE WASHINGTON JUNE 16-18 2020

Freeze the Discriminator: a Simple Baseline for Fine-Tuning GANs

Sangwoo Mo[1] Minsu Cho[2] Jinwoo Shin[1]

[1] Korea Advanced Institute of Science and Technology (KAIST)

[2] Pohang University of Science and Technology (POSTECH)

KAIST

POSTECH

Motivation

- Modern GANs show remarkable success in image generation[1,2,3]…

- …but they require lots of training data and computational resources
  - BigGAN is trained on 1M of images for 120 GPU days

[1] Brock et al. Large Scale GAN Training for High Fidelity Natural Image Synthesis. ICLR 2019.
[2] Karras et al. A Style-Based Generator Architecture for Generative Adversarial Networks. CVPR 2019.
[3] Karras et al. Analyzing and Improving the Image Quality of StyleGAN. arXiv 2020.

Prior Work

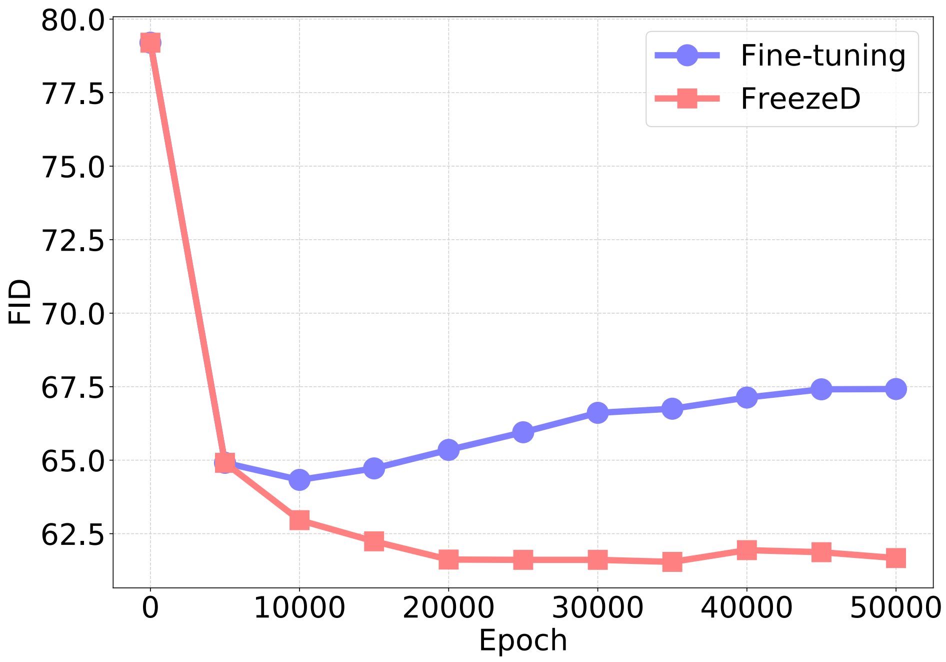
- Several works propose a transfer learning approach…
- …but they are often prone to overfitting (w/ limited data)
  - Fine-tuning[4]
- …or limited to learning small distribution shifts
  - Fine-tuning only BN/IN parameters[5]
  - Fine-tuning with supervised loss[5]
  - Modify the prior distribution $p(z)$[6]

[4] Wang et al. Transferring GANs: generating images from limited data. ECCV 2018.
[5] Noguchi & Harada. Image Generation From Small Datasets via Batch Statistics Adaptation. ICCV 2019.
[6] Wang et al. MineGAN: effective knowledge transfer from GANs to target domains with few images. arXiv 2020.

FreezeD: a simple but strong baseline

- We find that simply **freezing** the **lower layers** of the **discriminator** while fine-tuning GANs surprisingly work well

- FreezeD stably converges to the better optima than fine-tuning



Experiments: unconditional image generation

- FreezeD outperforms fine-tuning on StyleGAN[2] architecture

(a) Original (FFHQ) (b) Cat (c) Dog

Table 1: FID scores under Animal Face dataset. Left and right values indicate the best and final FID scores.

	Bear	Cat	Chicken	Cow	Deer	Dog	Duck	Eagle	Elephant	Human
Fine-tuning	82.82/84.38	71.76/73.47	88.10/88.83	87.07/87.46	82.11/84.04	64.28/67.42	92.54/92.54	85.52/86.88	84.10/84.33	76.62/76.72
FreezeD	**78.77/78.77**	**69.64/69.97**	**86.20/86.53**	**84.32/84.39**	**78.67/79.73**	**61.46/61.67**	**88.82/89.14**	**82.15/82.62**	**80.00/80.24**	**73.51/73.89**
	Lion	**Monkey**	**Mouse**	**Panda**	**Pigeon**	**Pig**	**Rabbit**	**Sheep**	**Tiger**	**Wolf**
Fine-tuning	76.86/78.36	86.70/87.30	84.95/85.61	74.29/76.07	81.24/81.36	85.31/86.08	89.11/89.82	86.98/87.89	73.21/75.06	79.97/81.37
FreezeD	**73.49/73.59**	**82.31/82.61**	**81.72/82.30**	**72.19/72.62**	**77.79/78.07**	**83.22/83.31**	**85.65/85.65**	**84.33/84.55**	**71.26/71.54**	**76.47/76.47**

[2] Karras et al. A Style-Based Generator Architecture for Generative Adversarial Networks. CVPR 2019.

Experiments: conditional image generation

- FreezeD outperforms fine-tuning on SNGAN-projection[7] architecture
  - FreezeD generates more class-consistent samples

(a) Flower (fine-tuning)

(b) Flower (FreezeD)

Table 4: FID scores under SNGAN-projection architecture. Left and right values indicate the best and final FID scores.

	Oxford Flower	CUB-200-2011	Caltech-256
Fine-tuning	27.05/**32.51**	32.29/32.60	62.20/63.37
FreezeD	**24.80**/52.92	**26.37/27.63**	**60.53/60.53**

[7] Miyato & Koyama. cGANs with Projection Discriminator. ICLR 2018.

Experiments: comparison & ablation study

- FreezeD outperforms various prior methods

Table 3: Comparison of various methods under 'Cat' and 'Dog' classes in the Animal Face dataset. Left and right values indicate the best and final FID scores. † indicates the model is trained by GLO loss, otherwise by GAN loss.

	Fine-tuning	Fine-tuning†	Scale/shift	Scale/shift†	MineGAN	MineGAN†	L2-SP (G)	L2-SP (D)	L2-SP (G,D)	FreezeD
Cat	71.76/73.47	78.21/78.32	71.99/73.42	80.63/80.63	82.67/82.67	82.68/82.95	71.77/73.78	71.54/72.67	71.70/73.47	**69.64/69.97**
Dog	64.28/67.42	75.19/75.45	64.12/67.79	79.08/79.91	79.05/79.23	79.11/79.20	64.18/67.14	64.28/66.68	64.25/66.06	**61.46/61.67**

- Freezing until intermediate layers performs the best
  - The optimal layer depends on the gap of source and target distribution

Table 6: Ablation study on freezing layers of D on SNGAN-projection architecture under Oxford Flower, CUB-200-2011, Caltech-256 datasets. Layer i indicates that the first i layers of the discriminator are frozen.

	Fine-tuning	Layer 1	Layer 2	Layer 3	Layer 4	Layer 5
Oxford Flower	27.05/**32.51**	27.65/42.14	25.85/42.31	**24.80**/52.92	25.41/87.60	25.35/104.07
CUB-200-2011	32.29/32.60	28.80/31.80	**26.37/27.63**	28.48/28.48	26.87/29.29	29.92/34.08
Calteth-256	62.20/63.37	**60.53/60.53**	61.59/61.94	61.29/61.95	61.92/62.88	62.90/62.90

Conclusion

- A simple baseline FreezeD outperforms the prior methods
- Lower layers of the discriminator learn some general features
  - It could be applied the universal detector of generated images
- Investigating advanced methods would be an interesting direction
  - Feature distillation seems to be a promising direction

Thank you for your attention 😀