

---

# Differentiable Augmentation for Data-Efficient GAN Training Supplementary Material

---

Shengyu Zhao      Zhijian Liu      Ji Lin      Jun-Yan Zhu      Song Han  
IIIS, Tsinghua University and MIT      MIT      MIT      Adobe and CMU      MIT

## Appendix A Hyperparameters and Training Details

### A.1 ImageNet Experiments

The Compare GAN codebase\* suffices to replicate BigGAN’s FID on ImageNet dataset at  $128 \times 128$  resolution but has some small differences to the original paper [2]. First, the codebase uses a learning rate of  $10^{-4}$  for  $G$  and  $5 \times 10^{-4}$  for  $D$ . Second, it processes the raw images into  $128 \times 128$  resolution with random scaling and random cropping. Since we find that random cropping leads to a worse IS, we process the raw images with random scaling and center cropping instead. We additionally augment the images with random horizontal flips, yielding the best re-implementation of BigGAN to our knowledge. With DiffAugment, we find that  $D$ ’s learning rate of  $5 \times 10^{-4}$  often makes  $D$ ’s loss stuck at a high level, so we reduce  $D$ ’s learning rate to  $4 \times 10^{-4}$  for the 100% data setting and  $2 \times 10^{-4}$  for the 10% and 20% data settings. However, we note that the baseline model does not benefit from this reduced learning rate: if we reduce  $D$ ’s learning rate from  $5 \times 10^{-4}$  to  $2 \times 10^{-4}$  under the 50% data setting, its performance degrades from an FID/IS of 9.64/89.9 to 10.79/75.7. All the models achieve the best FID within 200k iterations and deteriorate after that, taking up to 3 days on a TPU v2/v3 Pod with 128 cores.

See Figure 3 for a qualitative comparison between BigGAN and BigGAN + DiffAugment. Our method improves the image quality of the samples in both 25% and 100% data settings. The visual difference is more clear under the 25% data setting.

**Notes on CR-BigGAN [50].** CR-BigGAN [50] reports an FID of 6.66, which is slightly better than ours 6.80 (BigGAN + DiffAugment) with 100% data. However, the code and pre-trained models of CR-BigGAN [50] are not available, while its IS is not reported either. Our reimplemented CR-BigGAN only achieves an FID of 7.95 with an IS of 82.0, even worse than the baseline BigGAN. Nevertheless, our CIFAR experiments suggest the potential of applying DiffAugment on top of CR.

### A.2 FFHQ and LSUN-Cat Experiments

We use the official TensorFlow implementation of StyleGAN2<sup>†</sup> and the default network configuration at  $256 \times 256$  resolution with an  $R_1$  regularization  $\gamma$  of 1, but without the path length regularization and the lazy regularization since they do not improve FID [18]. The number of feature maps at shallow layers ( $64 \times 64$  resolution and above) is halved to match the architecture of ADA [16]. All the models in our experiments are augmented with random horizontal flips, trained on 8 GPUs with a maximum training length of 25,000k images.

See Figure 4-5 for qualitative comparisons between StyleGAN2 and StyleGAN2 + DiffAugment. Our method considerably improves the image quality with limited data available.

\*[https://github.com/google/compare\\_gan](https://github.com/google/compare_gan)

†<https://github.com/NVlabs/stylegan2>

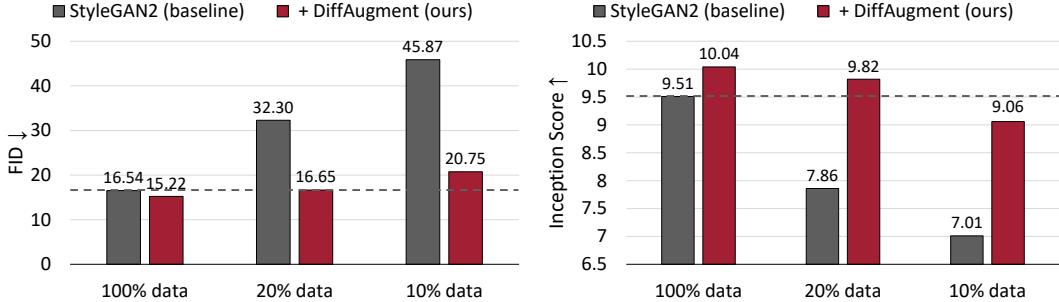


Figure 1: Unconditional generation results on **CIFAR-100**. We are able to roughly match StyleGAN2’s FID and outperform its IS using only 20% training data.

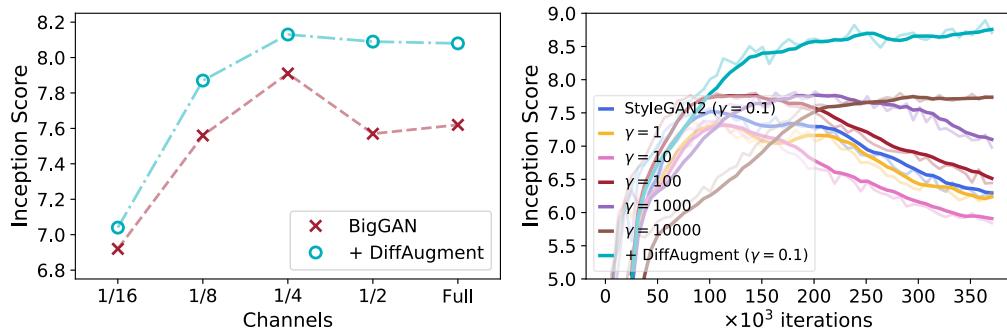


Figure 2: **Analysis of smaller models or stronger regularization** on CIFAR-10 with **10%** training data. *left*: Smaller models reduce overfitting for the BigGAN baseline, while our method outperforms it at all the model capacities. *right*: Over a wide sweep of the  $R_1$  regularization  $\gamma$  for the baseline StyleGAN2, its best IS (7.75) is still **12%** worse than ours (8.84).

Method	100% training data		20% training data		10% training data	
	IS	FID	IS	FID	IS	FID
BigGAN [2]	9.06	9.59	8.41	21.58	7.62	39.78
+ DiffAugment	<b>9.16</b>	<b>8.70</b>	<b>8.65</b>	<b>14.04</b>	<b>8.09</b>	<b>22.40</b>
CR-BigGAN [50]	<b>9.20</b>	9.06	8.43	20.62	7.66	37.45
+ DiffAugment	9.17	<b>8.49</b>	<b>8.61</b>	<b>12.84</b>	<b>8.49</b>	<b>18.70</b>
StyleGAN2 [18]	9.18	11.07	8.28	23.08	7.33	36.02
+ DiffAugment	<b>9.40</b>	<b>9.89</b>	<b>9.21</b>	<b>12.15</b>	<b>8.84</b>	<b>14.50</b>

Table 1: **CIFAR-10 results**. IS and FID are measured using 10k samples; the validation set is the reference distribution for FID calculation. We select the snapshot with the best FID for each method. Results are averaged over 5 evaluation runs; all standard deviations are less than 1% relatively.

Method	100% training data		20% training data		10% training data	
	IS	FID	IS	FID	IS	FID
BigGAN [2]	<b>10.92</b>	12.87	9.11	33.11	5.94	66.71
+ DiffAugment	10.66	<b>12.00</b>	<b>9.47</b>	<b>22.14</b>	<b>8.38</b>	<b>33.70</b>
CR-BigGAN [50]	<b>10.95</b>	11.26	8.44	36.91	7.91	47.16
+ DiffAugment	10.81	<b>11.25</b>	<b>9.12</b>	<b>20.28</b>	<b>8.70</b>	<b>26.90</b>
StyleGAN2 [18]	9.51	16.54	7.86	32.30	7.01	45.87
+ DiffAugment	<b>10.04</b>	<b>15.22</b>	<b>9.82</b>	<b>16.65</b>	<b>9.06</b>	<b>20.75</b>

Table 2: **CIFAR-100 results**. IS and FID are measured using 10k samples; the validation set is the reference distribution for FID calculation. We select the snapshot with the best FID for each method. Results are averaged over 5 evaluation runs; all standard deviations are less than 1% relatively.

### A.3 CIFAR-10 and CIFAR-100 Experiments

We replicate BigGAN and CR-BigGAN baselines on CIFAR using the PyTorch implementation<sup>‡</sup>. All hyperparameters are kept unchanged from the default CIFAR-10 configuration, including the batch size (50), the number of  $D$  steps (4) per  $G$  step, and a learning rate of  $2 \times 10^{-4}$  for both  $G$  and  $D$ . The hyperparameter  $\lambda$  of Consistency Regularization (CR) is set to 10 as recommended [50]. All the models are run on 2 GPUs with a maximum of 250k training iterations on CIFAR-10 and 500k iterations on CIFAR-100.

For StyleGAN2, we use the official TensorFlow implementation<sup>§</sup> but include some changes to make it work better on CIFAR. The number of channels is 128 at  $32 \times 32$  resolution and doubled at each coarser level with a maximum of 512 channels. We set the half-life of the exponential moving average of the generator’s weights to 1,000k instead of 10k images since it stabilizes the FID curve and leads to consistently better performance. We set  $\gamma = 0.1$  instead of 10 for the  $R_1$  regularization, which significantly improves the baseline’s performance under the 100% data setting on CIFAR. The path length regularization and the lazy regularization are also disabled. The baseline model can already achieve the best FID and IS to our knowledge for unconditional generation on the CIFAR datasets. All StyleGAN2 models are trained on 4 GPUs with the default batch size (32) and a maximum training length of 25,000k images.

We apply DiffAugment to BigGAN, CR-BigGAN, and StyleGAN2 without changes to the baseline settings. There are several things to note when applying DiffAugment in conjunction with gradient penalties [12] or CR [50]. The  $R_1$  regularization penalizes the gradients of  $D(\mathbf{x})$  w.r.t. the input  $\mathbf{x}$ . With DiffAugment, the gradients of  $D(T(\mathbf{x}))$  can be calculated w.r.t. either  $\mathbf{x}$  or  $T(\mathbf{x})$ . We choose to penalize the gradients of  $D(T(\mathbf{x}))$  w.r.t.  $T(\mathbf{x})$  for the CIFAR, FFHQ, and LSUN experiments since it slightly outperforms the other choice in practice; for the low-shot generation experiments, we penalize the gradients of  $D(T(\mathbf{x}))$  w.r.t.  $\mathbf{x}$  instead from which we observe better diversity of the generated images. As CR has already used image translation to calculate the consistency loss, we only apply Cutout DiffAugment on top of CR under the 100% data setting. For the 10% and 20% data settings, we exploit stronger regularization by directly applying CR between  $\mathbf{x}$  and  $T(\mathbf{x})$ , i.e., before and after the Translation + Cutout DiffAugment.

We match the top performance for unconditional generation on CIFAR-100 as well as CIFAR-10 using only 20% data (see Figure 1). See Figure 2 for the analysis of smaller models or stronger regularization in terms of IS. See Table 1 and Table 2 for quantitative results.

### A.4 Low-Shot Generation Experiments

We compare our method to transfer learning algorithms using the FreezeD’s codebase<sup>¶</sup> (for TransferGAN [45], Scale/shift [31], and FreezeD [30]) and the newly released MineGAN [44] code<sup>||</sup>. All the models are transferred from a pre-trained StyleGAN model from the FFHQ dataset [17] at  $256 \times 256$  resolution. FreezeD reports the best performance when freezing the first 4 layers of  $D$  [30]; when applying DiffAugment to FreezeD, we only freeze the first 2 layers of  $D$ . All other hyperparameters are kept unchanged from the default settings. All the models are trained on 1 GPU with a maximum of 10k training iterations on our 100-shot datasets and 20k iterations on the AnimalFace [37] datasets.

When training the StyleGAN2 model from scratch, we use their default network configuration at  $256 \times 256$  resolution with an  $R_1$  regularization  $\gamma$  of 10 but without the path length regularization and the lazy regularization. We use a smaller batch size of 16, which improves the performance of both the StyleGAN2 baseline and ours, compared to the default batch size of 32. All the models are trained on 4 GPUs with a maximum training length of 300k images on our 100-shot datasets and 500k images on the AnimalFace datasets.

See Figure 6 for the additional interpolation results, Figure 7 and Figure 8 for the nearest neighbor tests of our method without pre-training both in pixel space and in the LPIPIIS feature space [51]. See Figures 9-13 for qualitative comparisons.

<sup>‡</sup><https://github.com/ajbrock/BigGAN-PyTorch>

<sup>§</sup><https://github.com/NVlabs/stylegan2>

<sup>¶</sup><https://github.com/sangwoomo/FreezeD>

<sup>||</sup><https://github.com/yaxingwang/MineGAN>

## Appendix B Evaluation Metrics

We measure FID and IS using the official Inception v3 model in TensorFlow for all the methods and datasets. Note that some papers using PyTorch implementations, including FreezeD [30], report different numbers from the official TensorFlow implementation of FID and IS. On ImageNet, CIFAR-10, and CIFAR-100, we inherit the setting from the Compare GAN codebase that the number of samples of generated images equals the number of real images in the validation set, and the validation set is used as the reference distribution for FID calculation. For the low-shot generation experiments, we sample 5k generated images and we use the training set as the reference distribution. For the FFHQ and LSUN experiments, we use the same evaluation setting as ADA [16].

## Appendix C 100-Shot Generation Benchmark

We collect the 100-shot datasets from the Internet. We then manually filter and crop each image as a pre-processing step. The full datasets are available [here](#).



Figure 3: Qualitative comparison on **ImageNet**  $128 \times 128$  without the truncation trick [2].

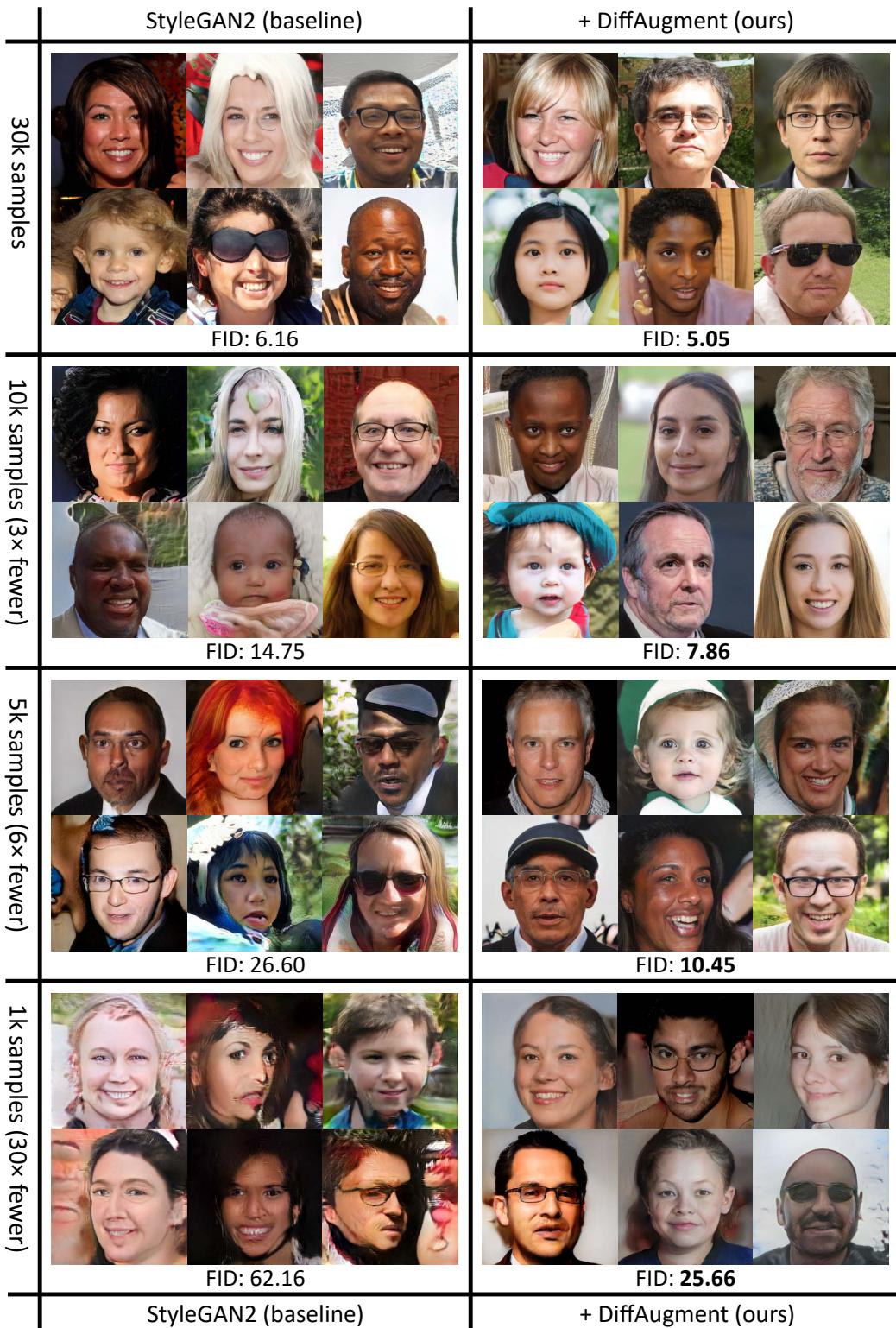


Figure 4: Qualitative comparison on FFHQ at 256×256 resolution with 1k, 5k, 10k, and 30k training images. Our method consistently outperforms the StyleGAN2 baselines [18] under different data percentage settings.



Figure 5: Qualitative comparison on **LSUN-cat** at 256×256 resolution with 1k, 5k, 10k, and 30k training images. Our method consistently outperforms the StyleGAN2 baselines [18] under different data percentage settings.

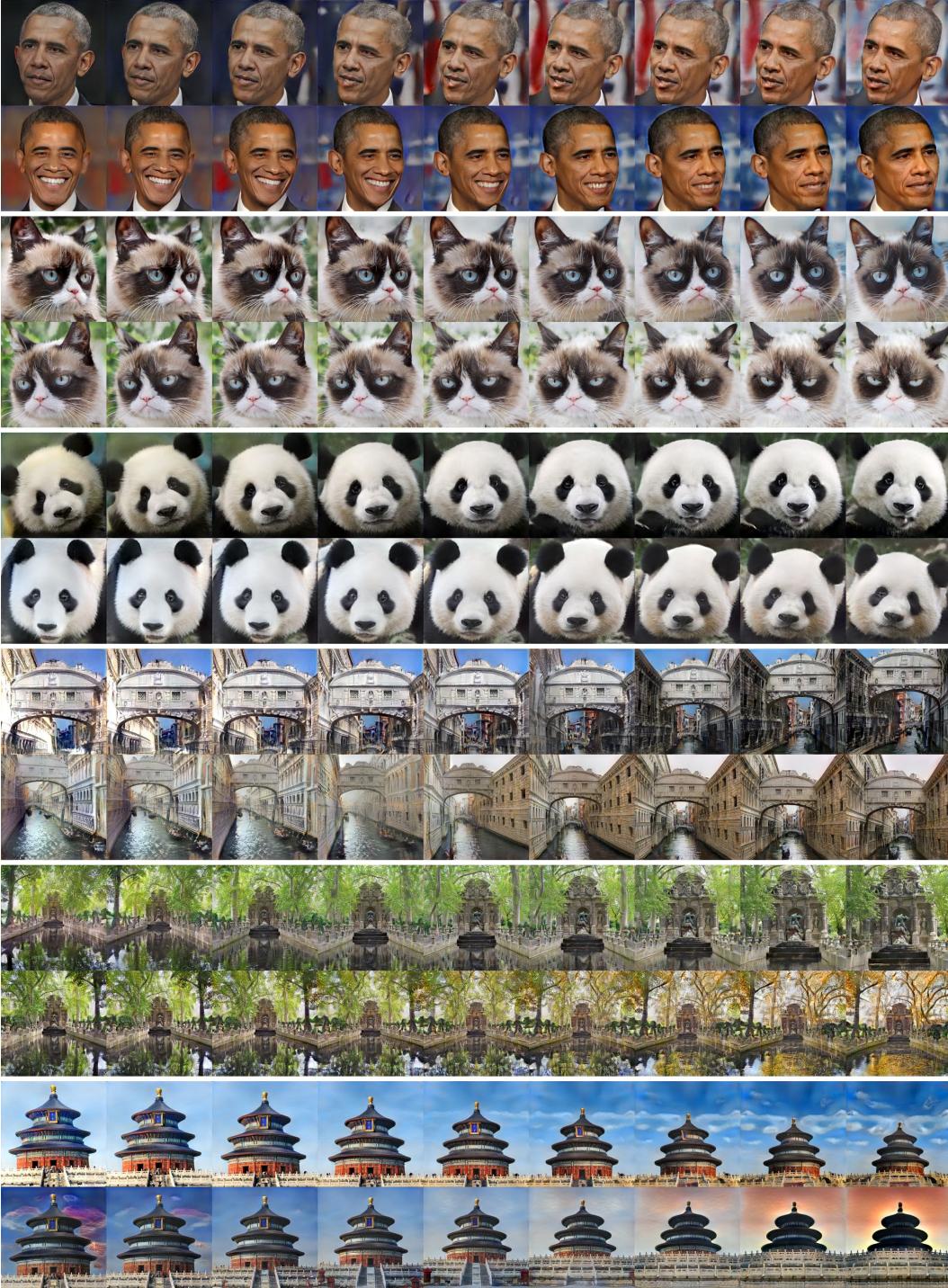


Figure 6: **Style space interpolation** of our method on the 100-shot Obama, grumpy cat, panda, the Bridge of Sighs, the Medici Fountain, and the Temple of Heaven datasets without pre-training. The smooth interpolation results suggest little overfitting of our method even given small datasets.



Figure 7: **Nearest neighbors in pixel space** measured by the pixel-wise  $L_1$  distance. Each query (on the left of the dashed lines) is a generated image of our method without pre-training (StyleGAN2 + DiffAugment) on the 100-shot or AnimalFace generation datasets. Each nearest neighbor (on the right of the dashed lines) is an original image queried from the training set with horizontal flips. The generated images are different from the training set, indicating that our model does not simply memorize the training images or overfit even given small datasets.



Figure 8: **Nearest neighbors in feature space** measured by the Learned Perceptual Image Patch Similarity (LPIPS) [51]. Each query (on the left of the dashed lines) is a generated image of our method without pre-training (StyleGAN2 + DiffAugment) on the 100-shot or AnimalFace generation datasets. Each nearest neighbor (on the right of the dashed lines) is an original image queried from the training set with horizontal flips. The generated images are different from the training set, indicating that our model does not simply memorize the training images or overfit even given small datasets.

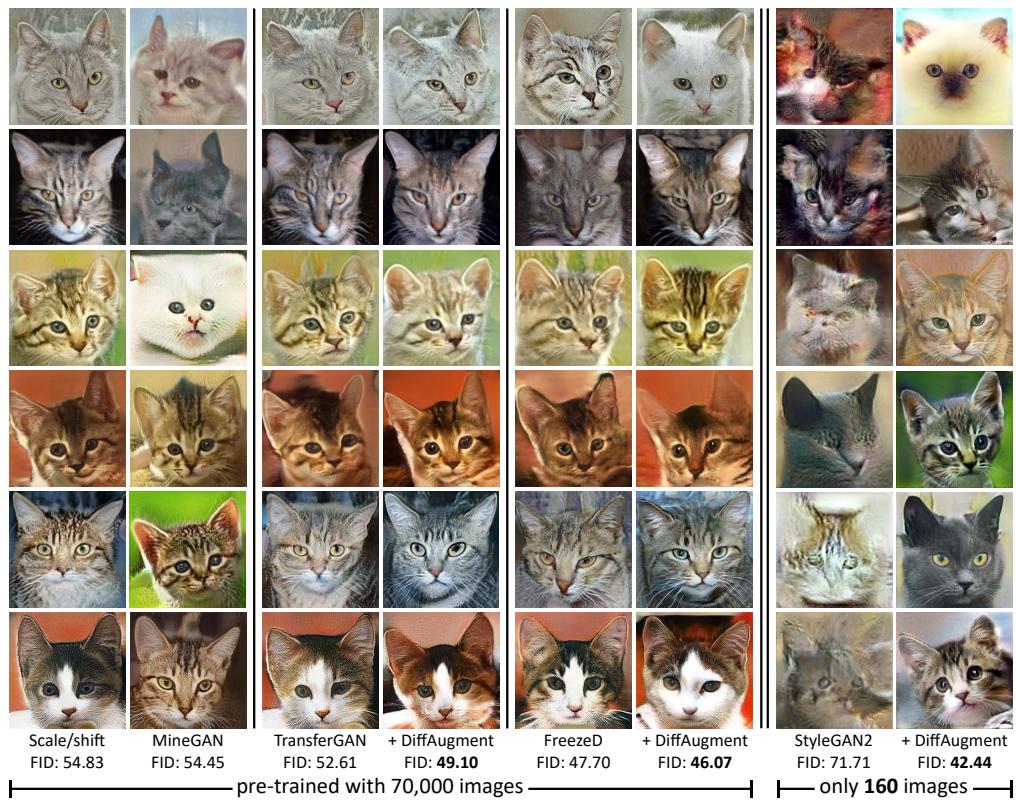


Figure 9: Qualitative comparison on the **AnimalFace-cat** [37] dataset.

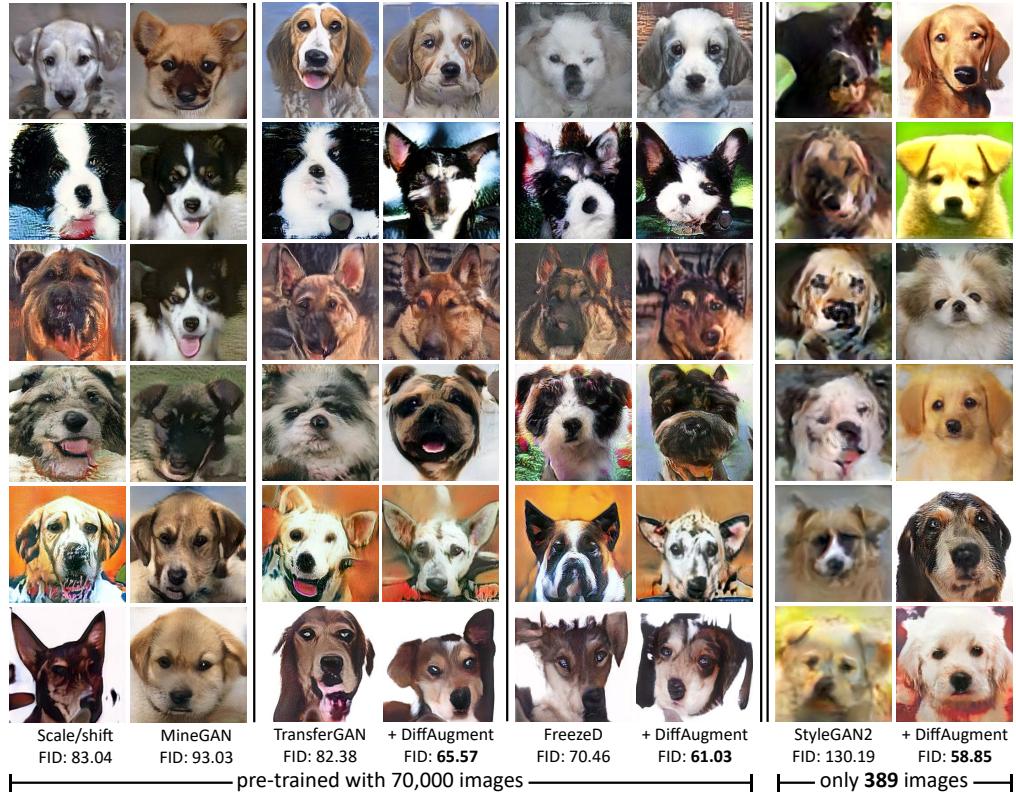


Figure 10: Qualitative comparison on the **AnimalFace-dog** [37] dataset.

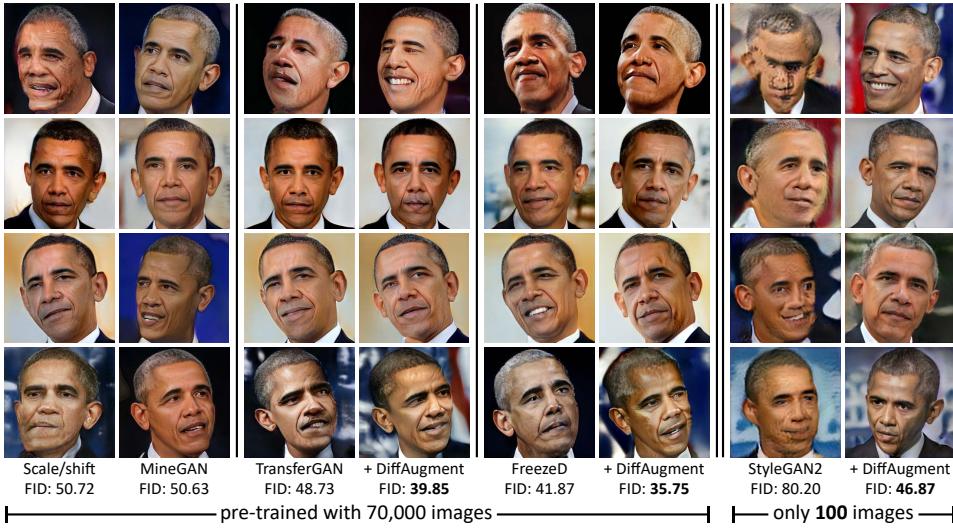


Figure 11: Qualitative comparison on the **100-shot Obama** dataset.

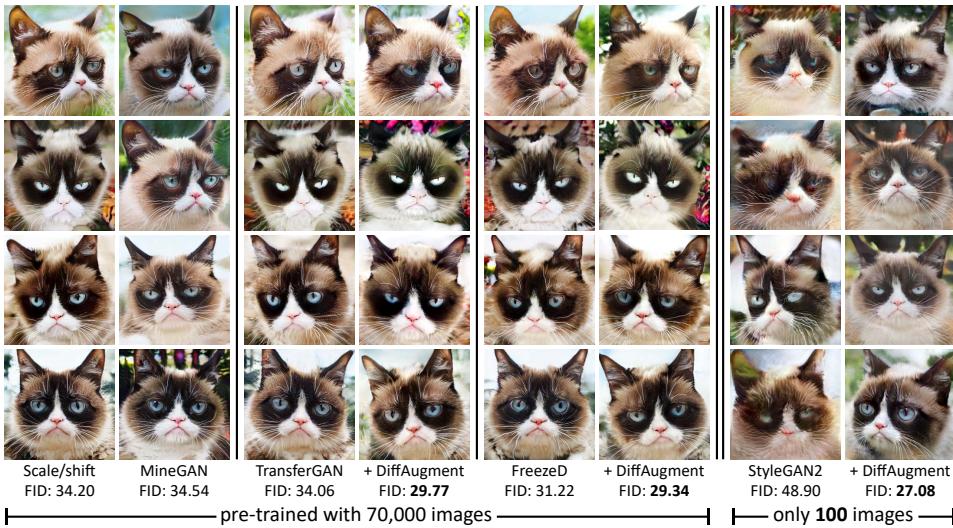


Figure 12: Qualitative comparison on the **100-shot grumpy cat** dataset.

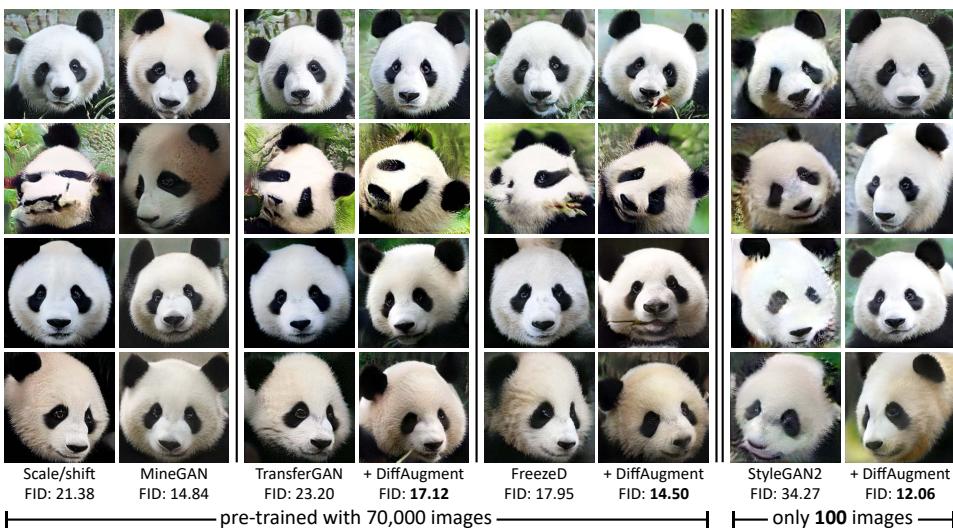


Figure 13: Qualitative comparison on the **100-shot panda** dataset.

## References

- [1] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein Generative Adversarial Networks. In *International Conference on Machine Learning (ICML)*, 2017.
- [2] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large Scale GAN Training for High Fidelity Natural Image Synthesis. In *International Conference on Learning Representations (ICLR)*, 2018. [1](#), [2](#), [5](#)
- [3] Ting Chen, Xiaohua Zhai, Marvin Ritter, Mario Lucic, and Neil Houlsby. Self-supervised gans via auxiliary rotation loss. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [4] Ekin D Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V Le. AutoAugment: Learning Augmentation Policies from Data. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [5] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. RandAugment: Practical Automated Data Augmentation with a Reduced Search Space. *arXiv*, 2019.
- [6] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2009.
- [7] Emily L Denton, Soumith Chintala, Rob Fergus, et al. Deep generative image models using a laplacian pyramid of adversarial networks. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2015.
- [8] Terrance DeVries and Graham W Taylor. Improved regularization of convolutional neural networks with cutout. *arXiv*, 2017.
- [9] Li Fei-Fei, Rob Fergus, and Pietro Perona. One-shot learning of object categories. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 28(4):594–611, 2006.
- [10] Shiry Ginosar, Kate Rakelly, Sarah Sachs, Brian Yin, and Alexei A Efros. A century of portraits: A visual historical record of american high school yearbooks. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2015.
- [11] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Nets. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2014.
- [12] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron C Courville. Improved training of wasserstein gans. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2017. [3](#)
- [13] 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 *Conference on Neural Information Processing Systems (NeurIPS)*, 2017.
- [14] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *arXiv*, 2020.
- [15] Tero Karras, Timo Aila, Samuli Laine, and Jaakko Lehtinen. Progressive growing of gans for improved quality, stability, and variation. In *International Conference on Learning Representations (ICLR)*, 2018.
- [16] Tero Karras, Miika Aittala, Janne Hellsten, Samuli Laine, Jaakko Lehtinen, and Timo Aila. Training generative adversarial networks with limited data. *arXiv*, 2020. [1](#), [4](#)
- [17] Tero Karras, Samuli Laine, and Timo Aila. A Style-Based Generator Architecture for Generative Adversarial Networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. [3](#)
- [18] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and Improving the Image Quality of StyleGAN. *arXiv*, 2019. [1](#), [2](#), [6](#), [7](#)
- [19] Alex Krizhevsky and Geoffrey Hinton. Learning multiple layers of features from tiny images. Technical report, Citeseer, 2009.
- [20] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet Classification with Deep Convolutional Neural Networks. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2012.
- [21] Brenden M Lake, Ruslan Salakhutdinov, and Joshua B Tenenbaum. Human-level concept learning through probabilistic program induction. *Science*, 350(6266):1332–1338, 2015.
- [22] Muyang Li, Ji Lin, Yaoyao Ding, Zhijian Liu, Jun-Yan Zhu, and Song Han. GAN Compression: Efficient Architectures for Interactive Conditional GANs. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
- [23] Sungbin Lim, Ildoo Kim, Taesup Kim, Chiheon Kim, and Sungwoong Kim. Fast AutoAugment. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2019.
- [24] Steven Liu, Tongzhou Wang, David Bau, Jun-Yan Zhu, and Antonio Torralba. Diverse image generation via self-conditioned gans. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.
- [25] Mario Lučić, Michael Tschannen, Marvin Ritter, Xiaohua Zhai, Olivier Bachem, and Sylvain Gelly. High-Fidelity Image Generation With Fewer Labels. In *International Conference on Machine Learning (ICML)*, 2019.

- [26] Xudong Mao, Qing Li, Haoran Xie, Raymond YK Lau, Zhen Wang, and Stephen Paul Smolley. Least squares generative adversarial networks. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [27] Lars Mescheder, Andreas Geiger, and Sebastian Nowozin. Which training methods for gans do actually converge? In *International Conference on Machine Learning (ICML)*, 2018.
- [28] Takeru Miyato, Toshiki Kataoka, Masanori Koyama, and Yuichi Yoshida. Spectral normalization for generative adversarial networks. In *International Conference on Learning Representations (ICLR)*, 2018.
- [29] Takeru Miyato and Masanori Koyama. cgans with projection discriminator. In *International Conference on Learning Representations (ICLR)*, 2018.
- [30] Sangwoo Mo, Minsu Cho, and Jinwoo Shin. Freeze Discriminator: A Simple Baseline for Fine-tuning GANs. *arXiv*, 2020. [3](#), [4](#)
- [31] Atsuhiro Noguchi and Tatsuya Harada. Image generation from small datasets via batch statistics adaptation. In *IEEE International Conference on Computer Vision (ICCV)*, 2019. [3](#)
- [32] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, and Jun-Yan Zhu. Semantic image synthesis with spatially-adaptive normalization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [33] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. In *International Conference on Learning Representations (ICLR)*, 2016.
- [34] Kevin Roth, Aurelien Lucchi, Sebastian Nowozin, and Thomas Hofmann. Stabilizing training of generative adversarial networks through regularization. In *Conference on Neural Information Processing Systems (NeurIPS)*, pages 2018–2028, 2017.
- [35] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training GANs. In *Conference on Neural Information Processing Systems (NeurIPS)*, 2016.
- [36] Han Shu, Yunhe Wang, Xu Jia, Kai Han, Hanting Chen, Chunjing Xu, Qi Tian, and Chang Xu. Co-Evolutionary Compression for Unpaired Image Translation. In *IEEE International Conference on Computer Vision (ICCV)*, 2019.
- [37] Zhangzhang Si and Song-Chun Zhu. Learning Hybrid Image Templates (HIT) by Information Projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2011. [3](#), [11](#)
- [38] Patrice Y Simard, David Steinkraus, and John C Platt. Best practices for convolutional neural networks applied to visual document analysis. In *Proceedings of International Conference on Document Analysis and Recognition*, 2003.
- [39] Casper Kaae Sønderby, Jose Caballero, Lucas Theis, Wenzhe Shi, and Ferenc Huszár. Amortised map inference for image super-resolution. *arXiv*, 2016.
- [40] Kenneth O Stanley and Risto Miikkulainen. Evolving neural networks through augmenting topologies. *Evolutionary computation*, 10(2):99–127, 2002.
- [41] Ngoc-Trung Tran, Viet-Hung Tran, Ngoc-Bao Nguyen, Trung-Kien Nguyen, and Ngai-Man Cheung. Towards good practices for data augmentation in gan training. *arXiv*, 2020.
- [42] Li Wan, Matthew Zeiler, Sixin Zhang, Yann Le Cun, and Rob Fergus. Regularization of neural networks using dropconnect. In *International Conference on Machine Learning (ICML)*, 2013.
- [43] Xiaolong Wang, Abhinav Shrivastava, and Abhinav Gupta. A-fast-renn: Hard positive generation via adversary for object detection. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [44] Yaxing Wang, Abel Gonzalez-Garcia, David Berga, Luis Herranz, Fahad Shahbaz Khan, and Joost van de Weijer. Minegan: effective knowledge transfer from gans to target domains with few images. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. [3](#)
- [45] Yaxing Wang, Chenshen Wu, Luis Herranz, Joost van de Weijer, Abel Gonzalez-Garcia, and Bogdan Raducanu. Transferring gans: generating images from limited data. In *European Conference on Computer Vision (ECCV)*, 2018. [3](#)
- [46] 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*, 2015.
- [47] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond Empirical Risk Minimization. In *International Conference on Learning Representations (ICLR)*, 2018.
- [48] Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena. Self-Attention Generative Adversarial Networks. In *International Conference on Machine Learning (ICML)*, 2019.
- [49] Han Zhang, Tao Xu, Hongsheng Li, Shaoting Zhang, Xiaogang Wang, Xiaolei Huang, and Dimitris N Metaxas. Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks. In *IEEE International Conference on Computer Vision (ICCV)*, 2017.
- [50] Han Zhang, Zizhao Zhang, Augustus Odena, and Honglak Lee. Consistency regularization for generative adversarial networks. In *International Conference on Learning Representations (ICLR)*, 2020. [1](#), [2](#), [3](#)
- [51] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. [3](#), [10](#)

- [52] Zhengli Zhao, Zizhao Zhang, Ting Chen, Sameer Singh, and Han Zhang. Image augmentations for gan training. *arXiv*, 2020.
- [53] Brady Zhou and Philipp Krähenbühl. Don't let your discriminator be fooled. In *International Conference on Learning Representations (ICLR)*, 2019.
- [54] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *IEEE International Conference on Computer Vision (ICCV)*, 2017.
- [55] Barret Zoph and Quoc V Le. Neural architecture search with reinforcement learning. In *International Conference on Learning Representations (ICLR)*, 2017.