Evaluating the Reproducibility of Training GAN with Limited Data

INTRO DEEP LEARNING Semester Project - Final Report

Deep(FAMS): Mohammad Alyetama, Ali Al-Ramini, Fangyi Li

Introduction

Background

- Generative Adversarial Networks (GANs) is a generative modeling approach using deep learning techniques, to automatically discover and learn the input patterns to generate an output.

- **Issues with GAN models**
  GANs are limited to smaller sized datasets
  - Overfitting
  - Augmented distribution

- Applying a wide range of augmentation using an **adaptive discriminator augmentation** (ADA) mechanism based on Karras et. al (2020).

Introduction

Objectives & Approach

- **Evaluation of ADA mechanism with small datasets:**
  - *Are we able to reproduce good results as paper reported?*
  - We used StyleGAN2-ADA with some of the same datasets as used in the paper.
  - *Are we able to produce good results with other datasets?*
  - We evaluated replicability of StyleGAN2-ADA using other small datasets.
  - *Any advantage of ADA mechanism?*
  - We trained StyleGAN2 models and compared results with StyleGAN2-ADA.
  - *Any effect on training results when the quantity of the data is limited?*
  - We artificially subset a relatively large dataset, FFHQ, into different sizes then make comparison with and without ADA.

Experimental Setup

DATASETS

Dataset	No. of images
METFACES	1,336
Animal-Faces-High-Quality (AFHQ)	WILD: 4,738 DOG: 4,739
Cars196	16,185
OxfordFlowers102	6,149
Flickr-Faces-High-Quality (FFHQ)	52,000

Experimental Setup

Preprocessing

- **Resizing**
  - We resize our dataset to half the original size that was used in the paper, that is, a resized resolution of (256 X 256) for all datasets except for the METFACES dataset, which we resized to (512 X 512).
- **Data Splitting**
  - We use data splitting on FFHQ (Flickr-Faces-High-Quality) to measure the performance of StyleGAN2-ADA over portions of data.

Experimental Setup

Architecture

- **Architecture**
  - The StyleGAN2-ADA mechanism is implemented on top of the StyleGAN2 official TensorFlow implementation with most of the details unchanged, including network architectures.

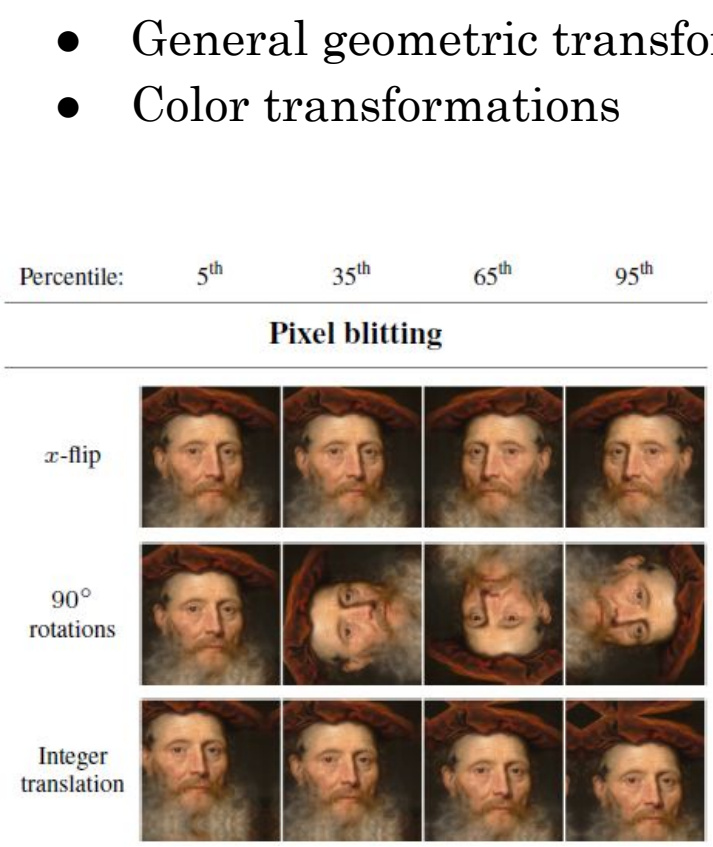
- **Data Augmentation**
  - One of the major improvements in StyleGAN2-ADA is dynamically changing the amount of image augmentation during training.

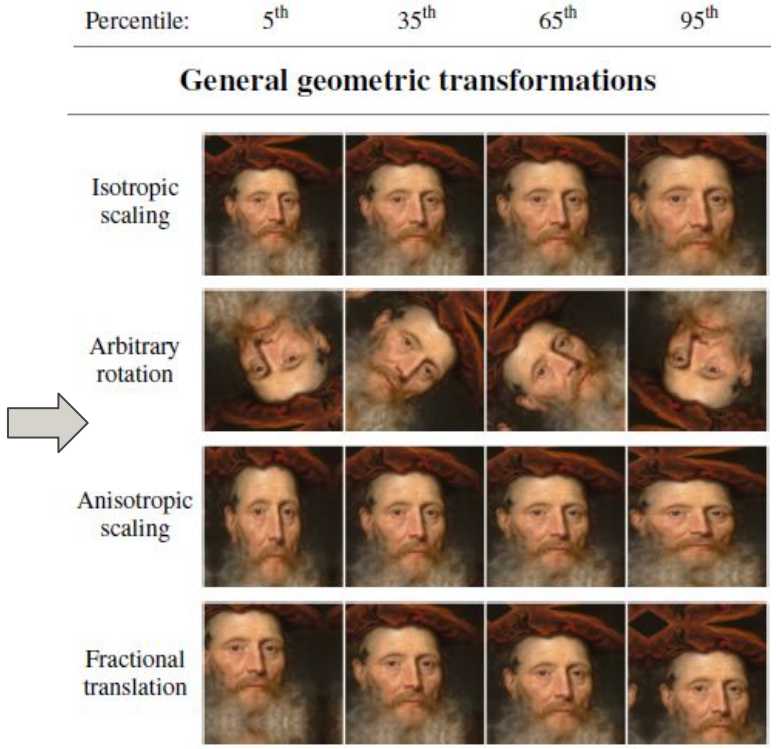
  - These augmentations consist of 18 transformations: geometric (7), color (5), filtering (4), and corruption (2).

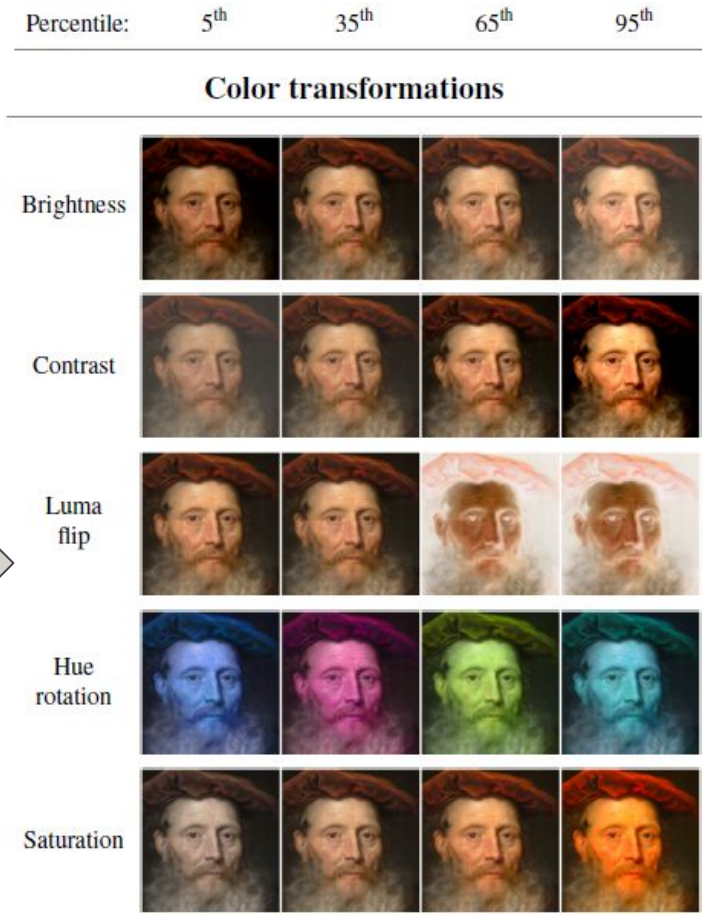
Experimental Setup

Data Augmentation Pipeline

1. **Geometric and color transformations**
   - Pixel Blitting
   - General geometric transformations
   - Color transformations







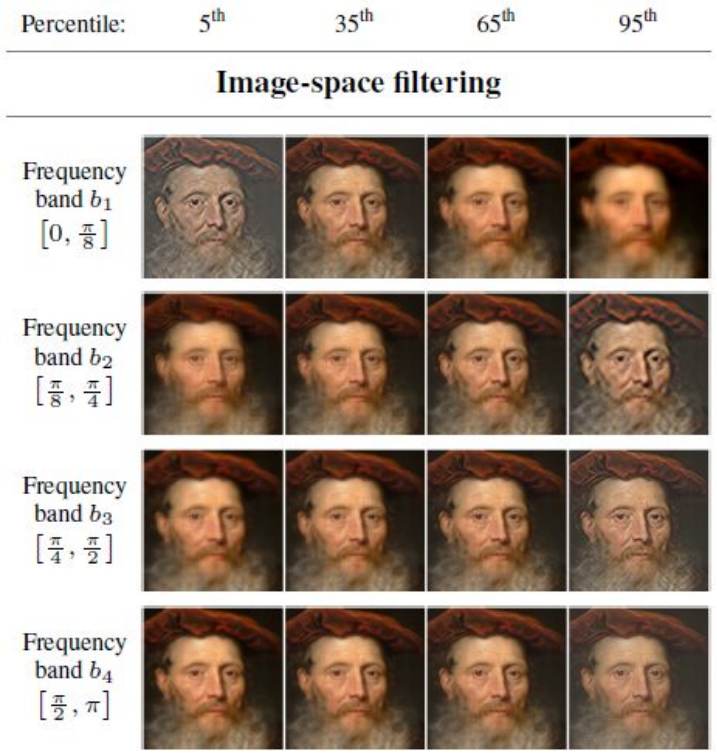
T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila, "Training generative adversarial networks with limited data," arXiv, 2020.

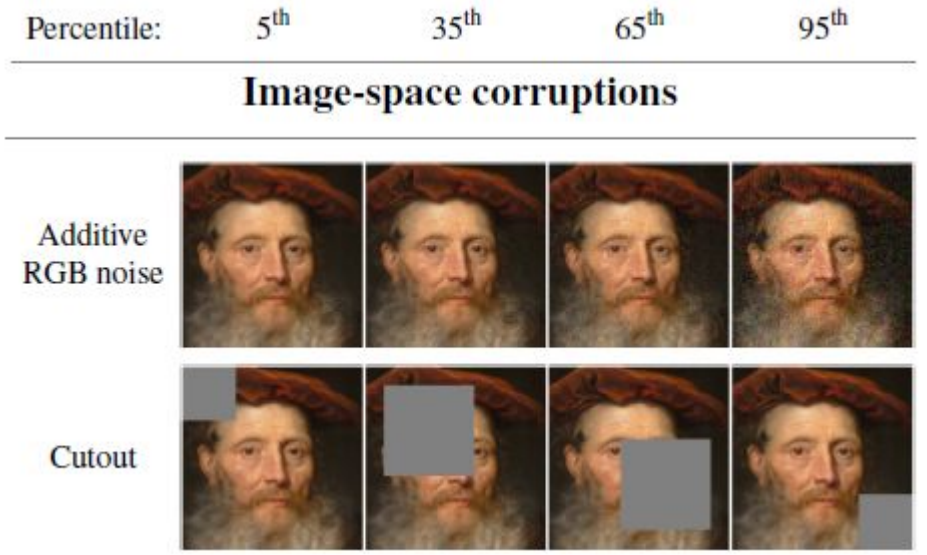
Experimental Setup

Data Augmentation Pipeline

2. Image-space filtering and corruptions

- Image-space filtering
- Image-space corruptions





T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila, "Training generative adversarial networks with limited data," arXiv, 2020.

Experimental Setup

Hyperparameters & Evaluation

- **Evaluation**

We used FID metric to evaluate our results, and compare it to the FID scores obtained from the original paper.

Parameter	METFACES	AFHQ-WILD	AFHQ-DOG	Cars196	Oxford Flowers102	FFHQ
Resolution	512*512	256*256				
No. of GPUs (Tesla V100)	2					
Base feature maps	16,384	8,192	5,000	16,185	8,192	52,000
Training time (hrs)	92.05	132.64	93.22	146.46	120.84	71.45
Minibatch size	16	32				
Minibatch stddev	4	4	16	16	4	16
Learning rate	0.0025					
R1 regularization	3.2768	0.4096				
Dataset x-flips; Mixed-precision; Mapping net depth; Style mixing reg.; Path length reg.; Resnet D	Yes	Yes	Yes	Yes	Yes	Yes

Results

Table 1. Final FID scores of datasets trained from scratch with StyleGAN2-ada

Dataset	Training time (in days)	Reported FID	Our FID
AFHQ-WILD	3	3.05 (original paper)	1.94
AFHQ-DOG	3	7.40 (original paper)	8.68
metfaces	3	15.34 (original paper)	18.26
cars196	5	16.03 (FineGAN)	8.07
102flowers	5	19.60 (MSG-StyleGAN)	6.85

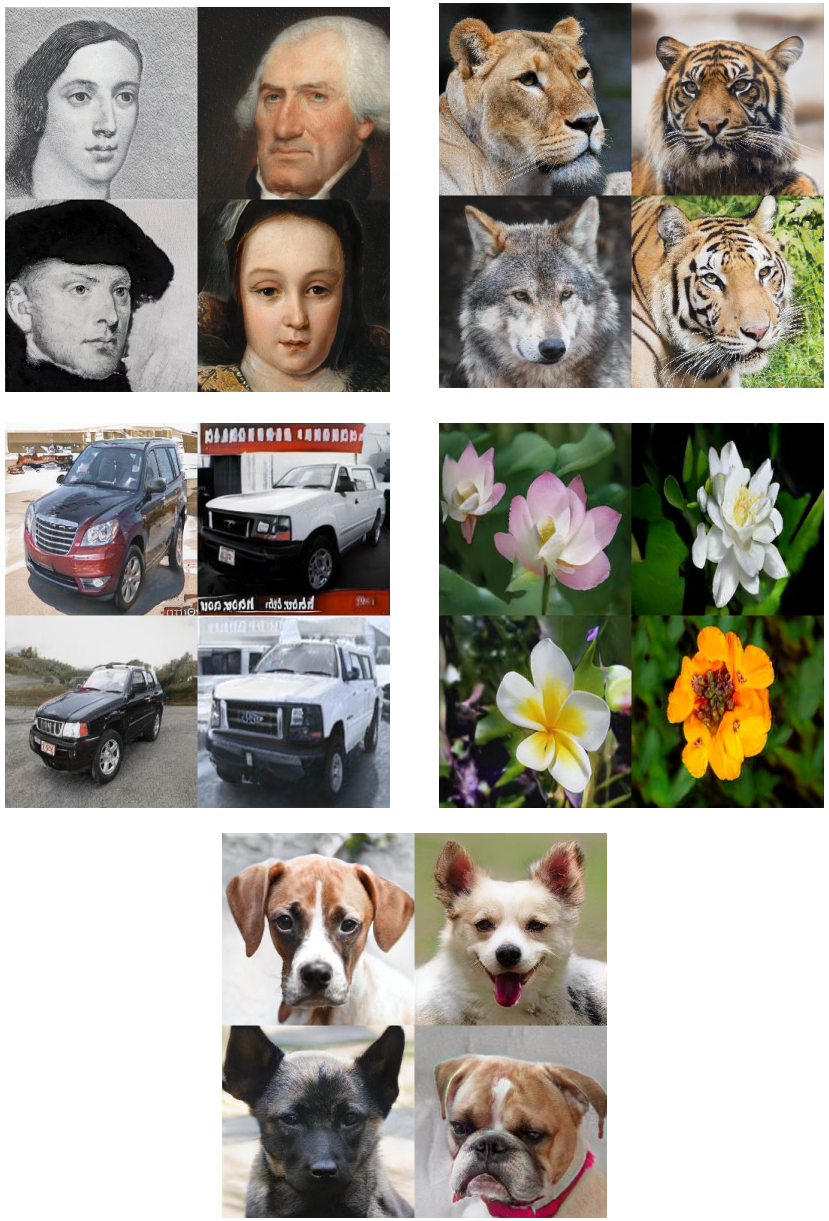


Figure 1. Example snapshots from various datasets trained with StyleGAN2-ada.

Cont. Result

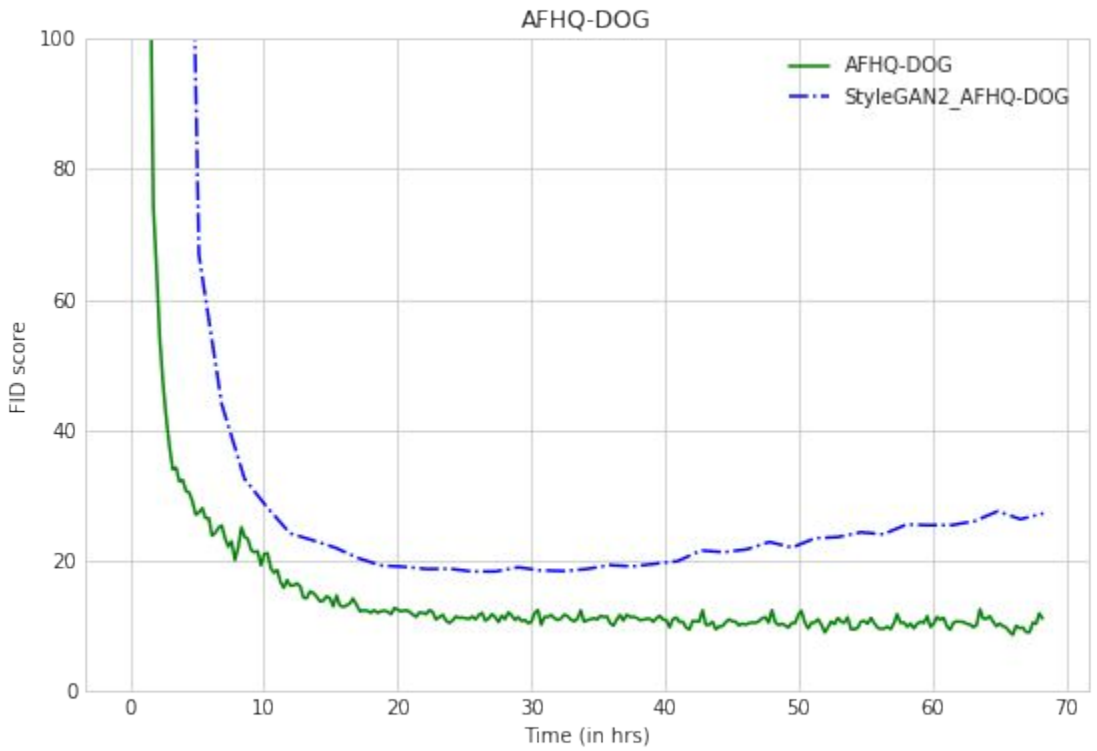


Figure 2. FID score as a function of wallclock time. Each curve corresponds to training the dataset from scratch with two NVIDIA Tesla V100 GPUs.

Comparison with baseline StyleGAN2

Original baseline StyleGAN2

StyleGAN2-ADA

Figure 3. Uncurated 256×256 results generated for AFHQ-DOG (4739 images) with and without ADA. Both generators were trained from scratch for ~3 days.

Cont. Results

Model performance with limited data

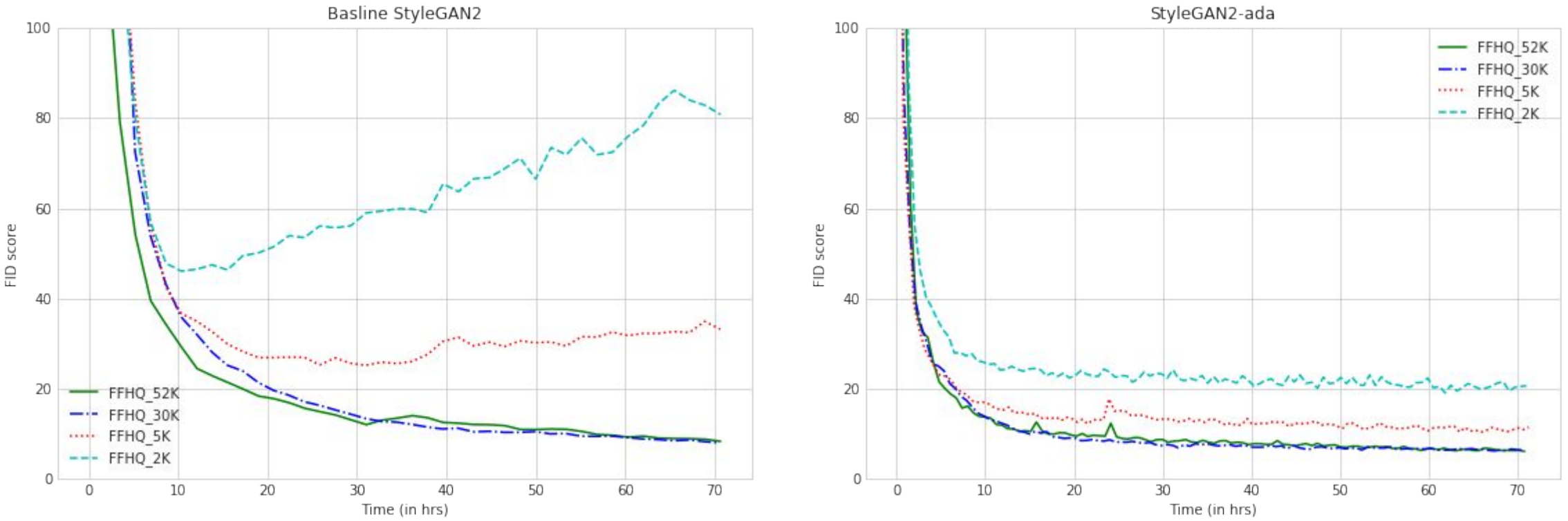


Figure 4. FID score as a function of wallclock time. Each curve corresponds to training the dataset from scratch with two NVIDIA Tesla V100 GPUs. Different lines color/shape correspond to different subsets of FFHQ.

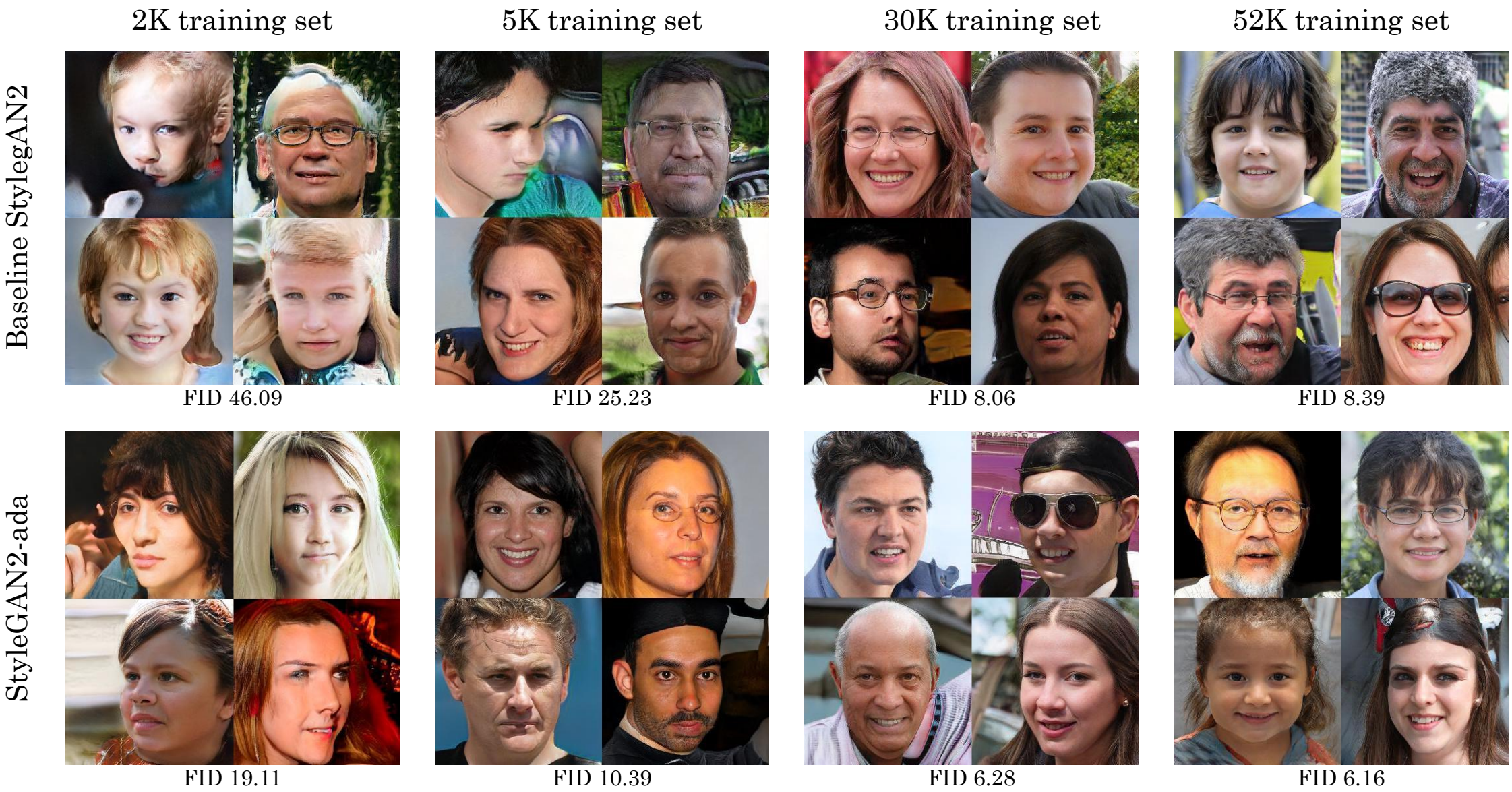
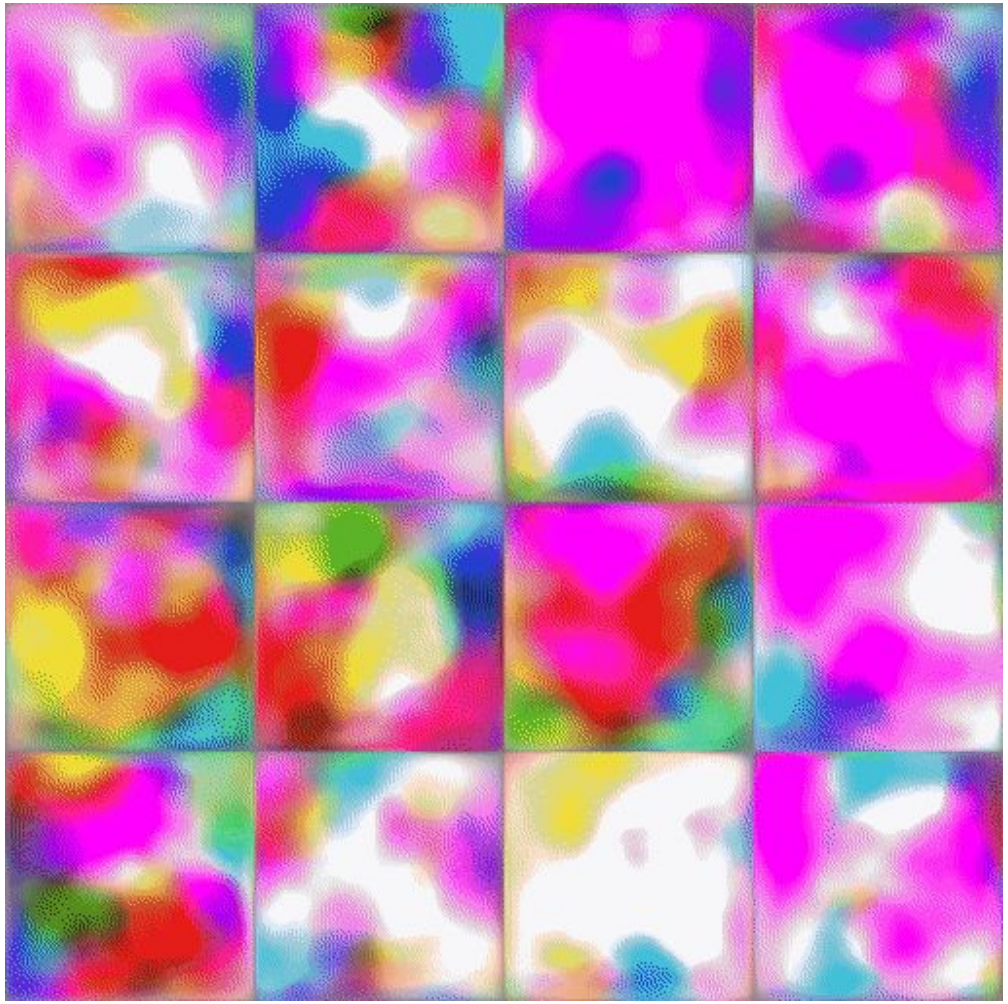


Figure 5. Images generated for different subsets of FFHQ at 256×256 resolution. We show the best snapshot of the **best** training run for each case, selected according to FID.

Discussion

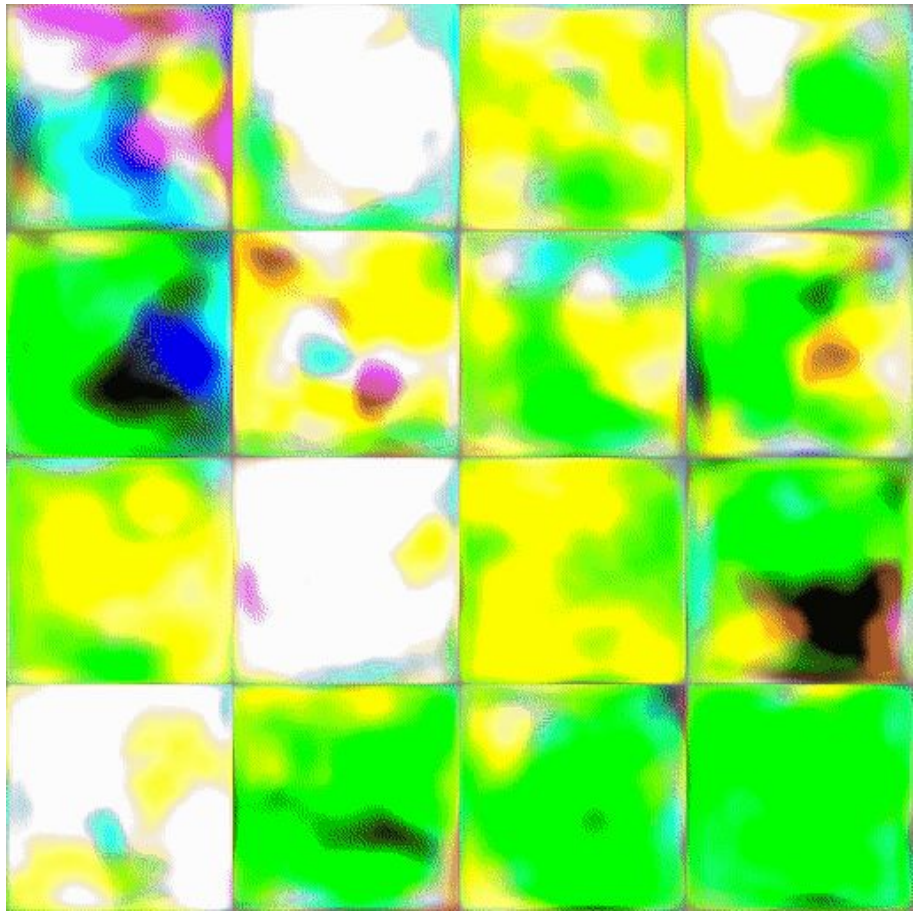
- StyleGAN2-ada performs a lot better than baseline StyleGAN2, especially when the quantity of the data is limited.
- Models that did not employ augmentation, degrade noticeably as the size of the training set decreases, generally yielding poor image quality and diversity with fewer than 30k training images.
- With ADA, the degradation is much more graceful, and the results remain reasonable even with a 5k training set.



*Animated image generated with the snapshots of the StyleGAN2-ada model trained from scratch on FFHQ.

Conclusion

- The ADA mechanism yields better results in a shorter time compared to baseline StyleGAN2.
- The adaptive discriminator augmentation reliably stabilizes training.
- Good results are possible using only a few thousand training images.
- The experiments that we attempted to reproduce from the paper were reproducible.
    - Downsides: poor forward compatibility.



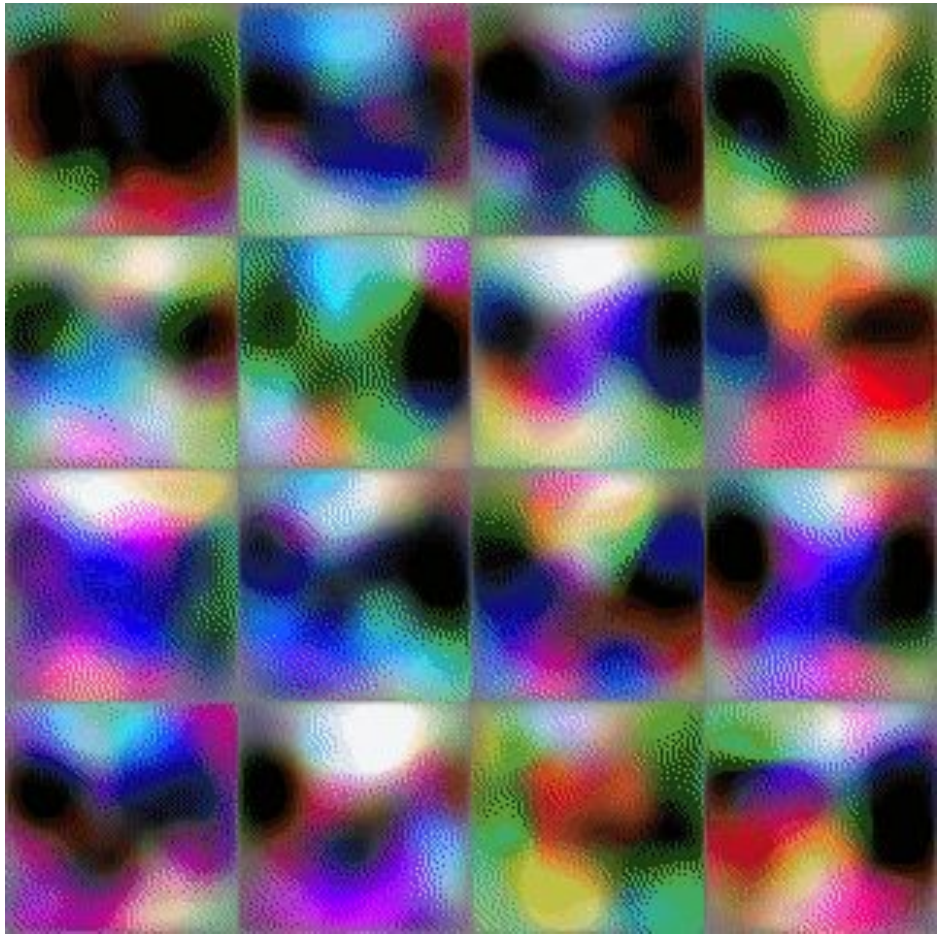
*Animated image generated with the snapshots of the StyleGAN2-ada model trained from scratch on AFHQ-WILD.

All source code to reproduce the reported results and figures are available on our GitHub repository and is ready for use with minimal to no adjustments.

https://github.com/Deep-FAMS/StyleGAN2-ada-Reproducibility

Thanks for Listening!

Any questions?



*Animated image is generated with the snapshots of the StyleGAN2-ada model trained from scratch on AFHQ-DOG.