Evaluating the Reproducibility of Training GAN with Limited Data

INTRO DEEP LEARNING Semester Project - Progress 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.

Experimental Setup

DATASETS and Architecture

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

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

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).

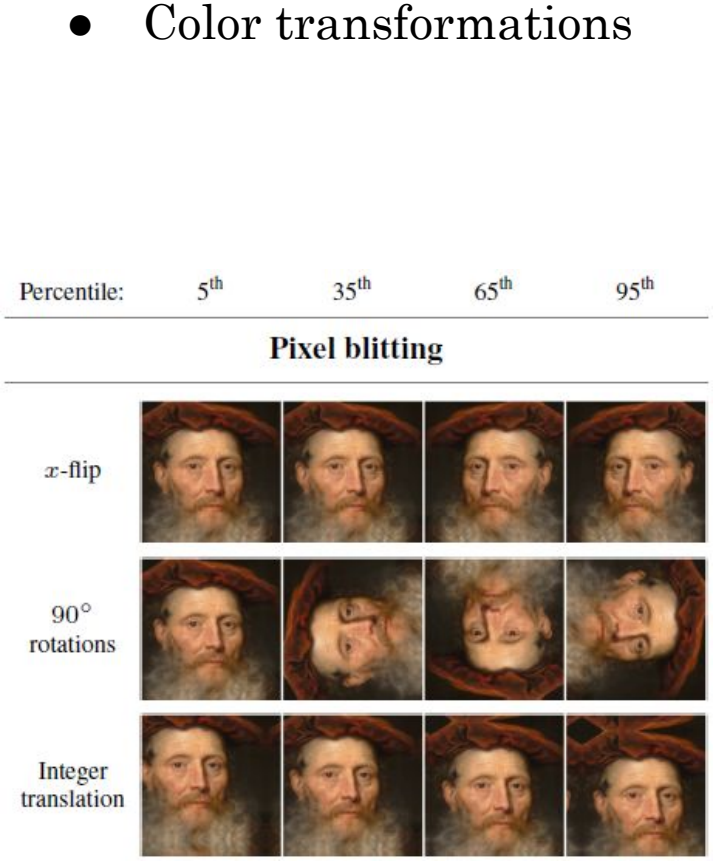
- **Data Augmentation**
  - The ADA mechanism mitigates issues related to the overfitting of the discriminator and undesirable distortions associated with StyleGAN2.

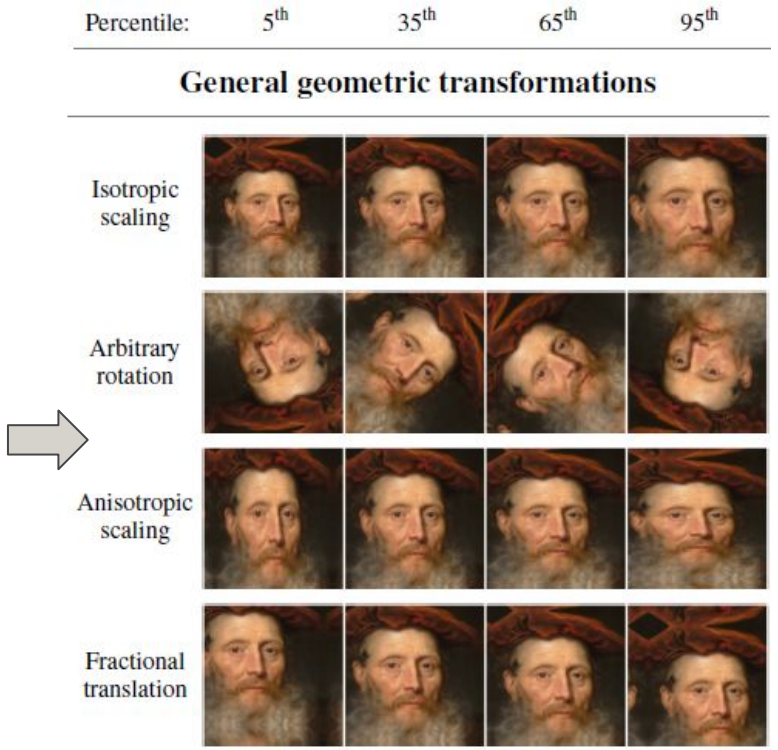
  - These augmentations consist of 18 transformations: geometric (7), color (5), filtering (4), and corruption (2).

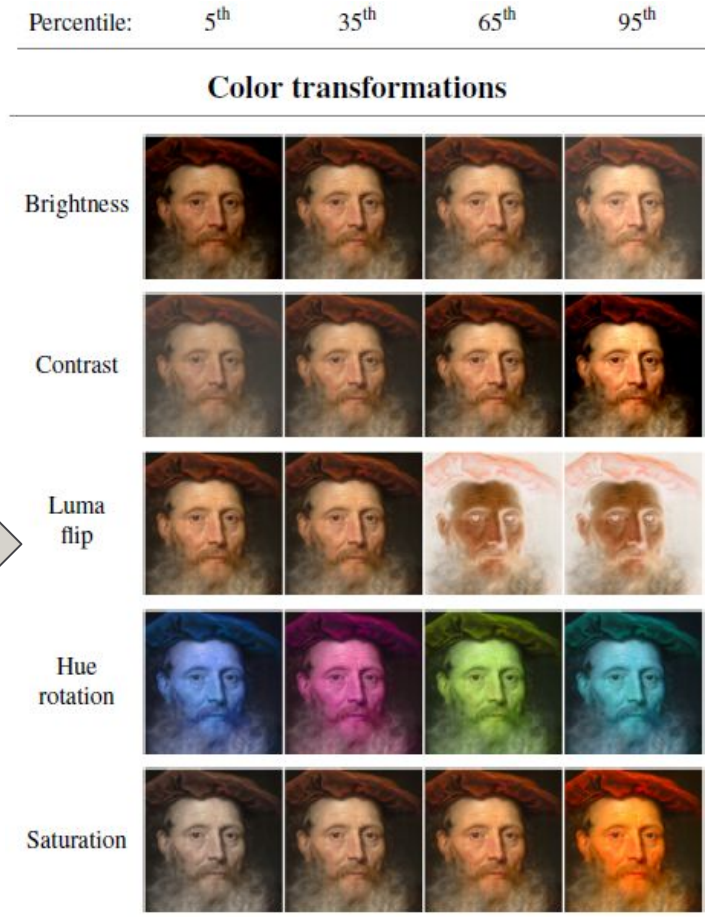
Experimental Setup

Data Augmentation Pipeline

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





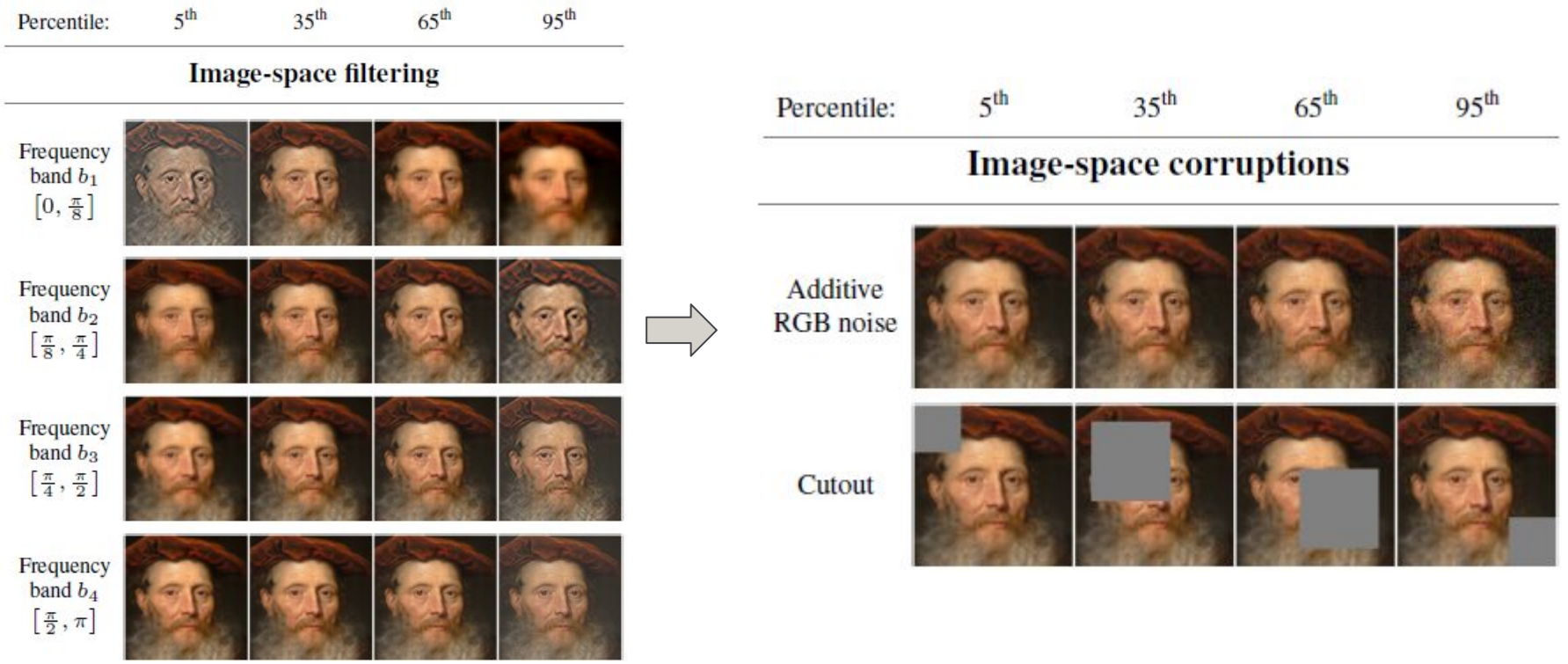


Experimental Setup

Data Augmentation Pipeline

2. Image-space filtering and corruptions

- Image-space filtering
- Image-space corruptions



Experimental Setup

Hyperparameters & Evaluation

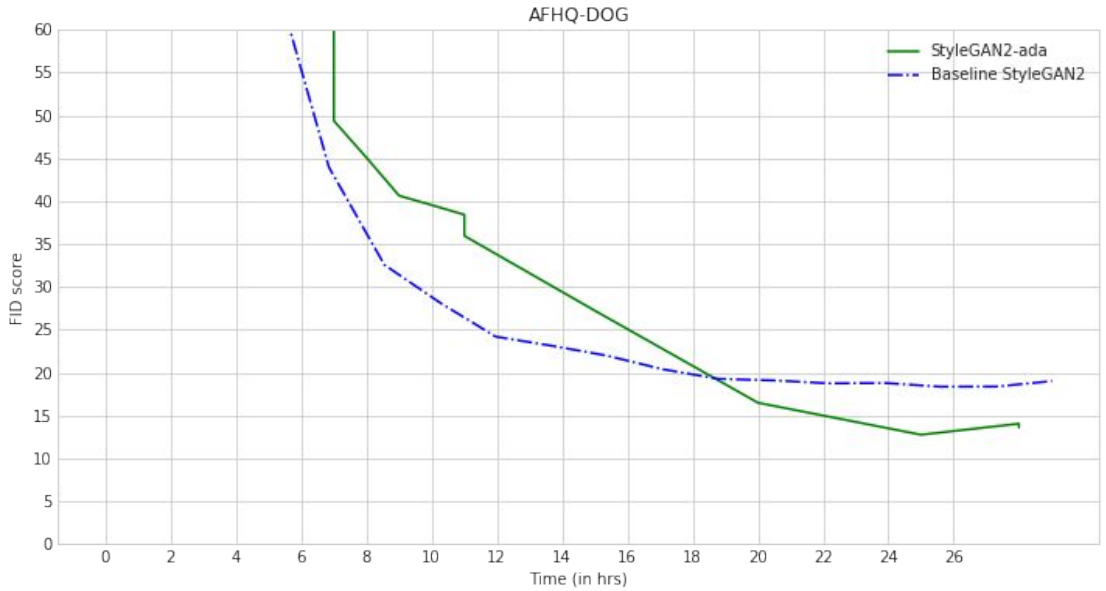
- **Evaluation**

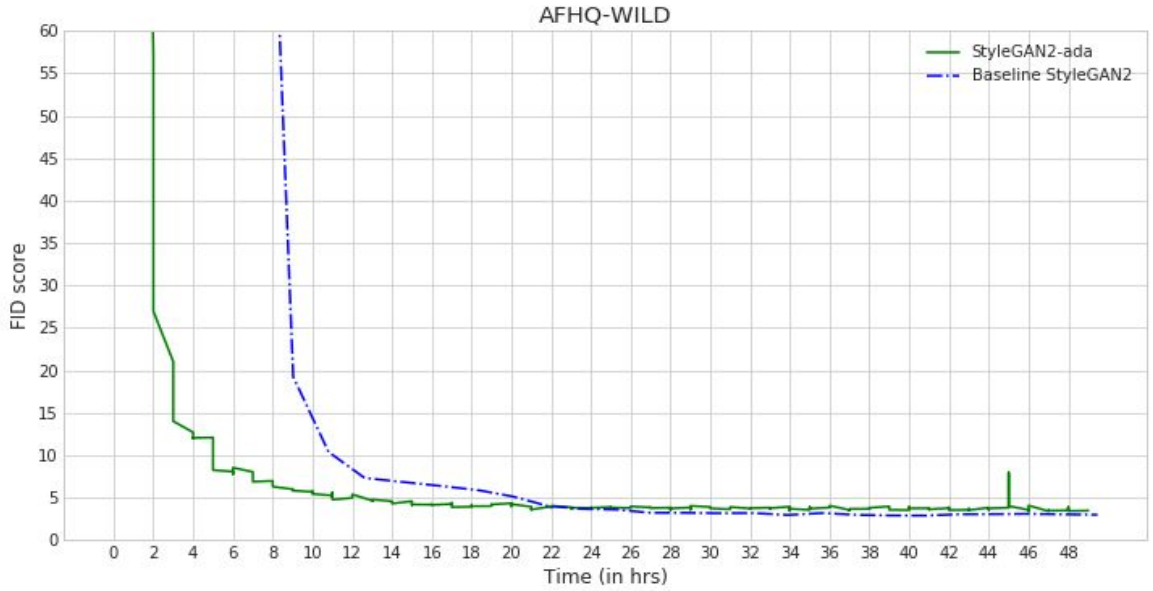
We used FID metric to evaluate our results, and compare it to the FID scores obtained from the original paper.

Parameter	METFACES	AFHQ-WILD	AFHQ-DOG	Cars196	Oxford Flowers102
Resolution	512*512	256*256			
No. of GPUs (Tesla V100)	2				
Base feature maps	16,384	8,192	5,000	16,185	8,192
Training time (hrs)	92.05	132.64	93.22	146.46	120.84
Minibatch size	16	32			
Minibatch stddev	4	4	16	16	4
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

Current Result

Comparison with baseline StyleGAN2





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

Current Result

Comparison with baseline StyleGAN2

Real images from the training set

Original baseline StyleGAN2

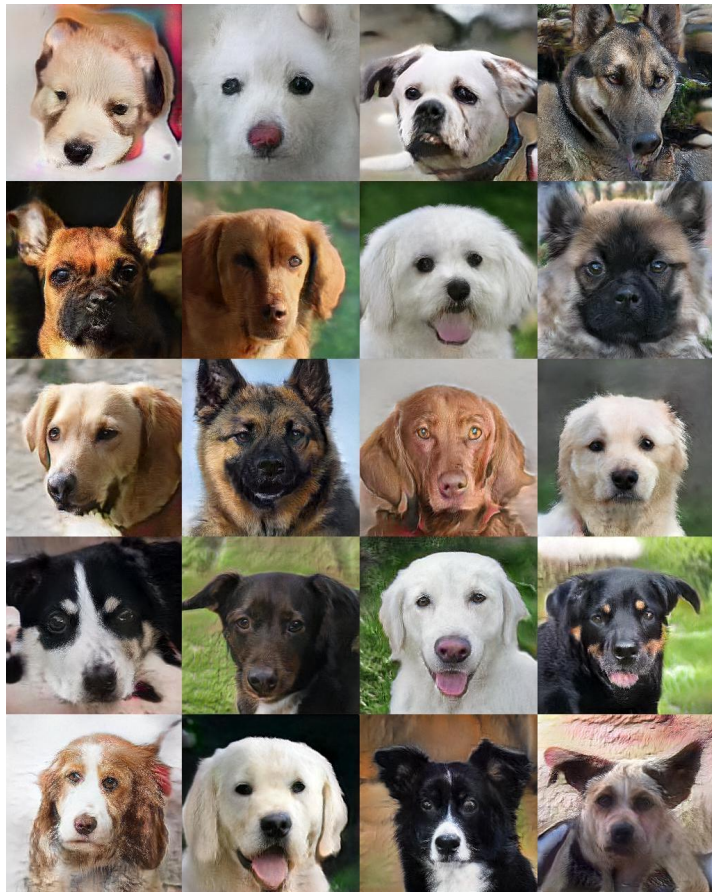
StyleGAN2-ADA

Uncurated 256×256 results generated for **AFHQ-WILD** (4738 images) with and without ADA, along with real images from the training set. Both generators were trained from scratch for 50 hours.

Current Result

Comparison with baseline StyleGAN2

Real images from the training set



Original baseline StyleGAN2

StyleGAN2-ADA

Uncurated 256×256 results generated for **AFHQ-DOG** (4739 images) with and without ADA, along with real images from the training set. Both generators were trained from scratch for 28 hours.

Current Result

Fréchet inception distance (FID) Scores

Dataset	Training Time (in hrs)	Current FID (ADA-StyleGAN2)	Reported FID
METFACES	102	20.89	18.22 (Original paper)
AFHQ-WILD	132	2.23	3.05 (Original paper)
AFHQ-DOG	93	10.83	7.40 (Original paper)
Cars196	146	9.21	16.03 (FineGAN)
OxfordFlowers 102	120	7.53	19.60 (MSG-StyleGAN)

Discussion

Model performance:

- ➔ ~1.8x faster training.
- ➔ Reaches convergence in a relatively short time.
- ➔ Reasonable out-of-the-box results for different dataset resolutions and few GPUs.
- ➔ The ADA mechanism improve results

Next step:

- ➔ Train subsets of the datasets with StyleGAN2-ada and compare performance.

Thanks for Listening!

Any questions?