FEEC UNICAMP

Presentation 2 IA376N 2s2025

Analyzing and Improving the Image Quality of StyleGAN

Henrique Parede de Souza - 260497

Overview

- Authors
- Classic GANs
- StyleGAN
- StyleGAN2
- Results
- Discussion & Conclusion

Analyzing and Improving the Image Quality of StyleGAN

Tero Karras NVIDIA
Samuli Laine NVIDIA
Miika Aittala NVIDIA
Janne Hellsten NVIDIA
Jaakko Lehtinen NVIDIA and Aalto University
Timo Aila NVIDIA

Abstract

The style-based GAN architecture (StyleGAN) yields state-of-the-art results in data-driven unconditional generative image modeling. We expose and analyze several of its characteristic artifacts, and propose changes in both model architecture and training methods to address them. In particular, we redesign the generator normalization, revisit progressive growing, and regularize the generator to encourage good conditioning in the mapping from latent codes to images. In addition to improving image quality, this path length regularizer yields the additional benefit that the generator becomes significantly easier to invert. This makes it possible to reliably attribute a generated image to a particular network. We furthermore visualize how well the generator utilizes its output resolution, and identify a capacity problem, motivating us to train larger models for additional quality improvements. Overall, our improved model redefines the state of the art in unconditional image modeling, both in terms of existing distribution quality metrics as well as perceived image quality.

1. Introduction

The resolution and quality of images produced by generative methods, especially generative adversarial networks (GAN) [13], are improving rapidly [20, 26, 4]. The current state-of-the-art method for high-resolution image synthesis is StyleGAN [21], which has been shown to work reliably on a variety of datasets. Our work focuses on fixing its characteristic artifacts and improving the result quality further.

The distinguishing feature of StyleGAN [21] is its unconventional generator architecture. Instead of feeding the input latent code $\mathbf{z} \in \mathcal{Z}$ only to the beginning of a the network, the *mapping network* f first transforms it to an intermediate latent code $\mathbf{w} \in \mathcal{W}$. Affine transforms then produce *styles* that control the layers of the *synthesis network* g via adaptive instance normalization (AdaIN) [18, 8, 11, 7]. Additionally, stochastic variation is facilitated by providing additional random noise maps to the synthesis network. It has been demonstrated [21, 33] that this design allows the intermediate latent space $\mathcal{W}$ to be much less entangled than the input latent space $\mathcal{Z}$. In this paper, we focus all analysis solely on $\mathcal{W}$, as it is the relevant latent space from the synthesis network's point of view.

Many observers have noticed characteristic artifacts in images generated by StyleGAN [3]. We identify two causes for these artifacts, and describe changes in architecture and training methods that eliminate them. First, we investigate the origin of common blob-like artifacts, and find that the generator creates them to circumvent a design flaw in its architecture. In Section 2, we redesign the normalization used in the generator, which removes the artifacts. Second, we analyze artifacts related to progressive growing [20] that has been highly successful in stabilizing high-resolution GAN training. We propose an alternative design that achieves the same goal — training starts by focusing on low-resolution images and then progressively shifts focus to higher and higher resolutions — without changing the network topology during training. This new design also allows us to reason about the effective resolution of the generated images, which turns out to be lower than expected, motivating a capacity increase (Section 4).

Quantitative analysis of the quality of images produced using generative methods continues to be a challenging topic. Fréchet inception distance (FID) [17] measures differences in the density of two distributions in the high-dimensional feature space of an InceptionV3 classifier [34]. Precision and Recall (P&R) [31, 22] provide additional visibility by explicitly quantifying the percentage of generated images that are similar to training data and the percentage of training data that can be generated, respectively. We use these metrics to quantify the improvements.

Both FID and P&R are based on classifier networks that have recently been shown to focus on textures rather than shapes [10], and consequently, the metrics do not accurately capture all aspects of image quality. We observe that the perceptual path length (PPL) metric [21], originally introduced as a method for estimating the quality of latent space

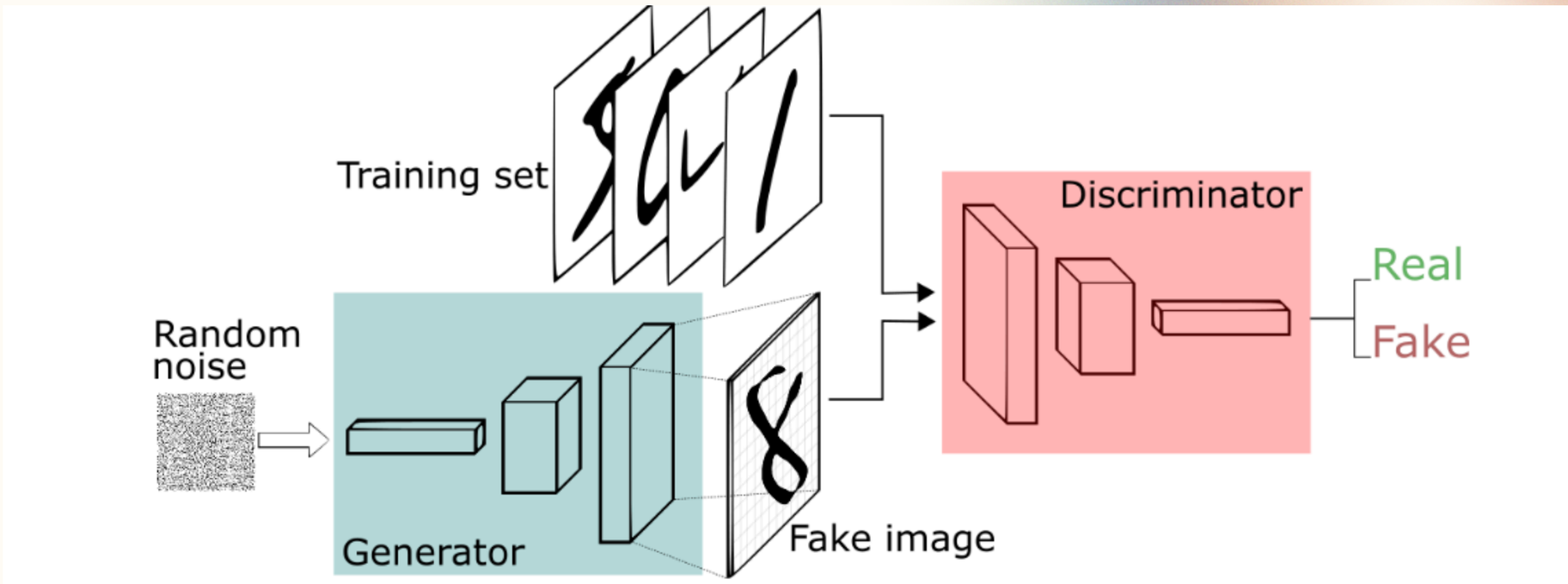
Authors

- Karras, Laine, and Aila (1st, 2nd, 6th) authored the first StyleGAN paper (2018).
- Senior researchers at NVIDIA Research, specializing in generative AI and graphics.
- Their work started the StyleGAN family (1 to 3), a benchmark in realistic image synthesis.

Main areas of study: Computer Graphics, Generative AI, Image Synthesis, Face Synthesis, 3D Synthesis Models

Classic GAN

Generative Adversarial Network



1) Train the discriminator to **distinguish real and fake images**.

2) Train the generator to create images that are able to **fool the discriminator**.

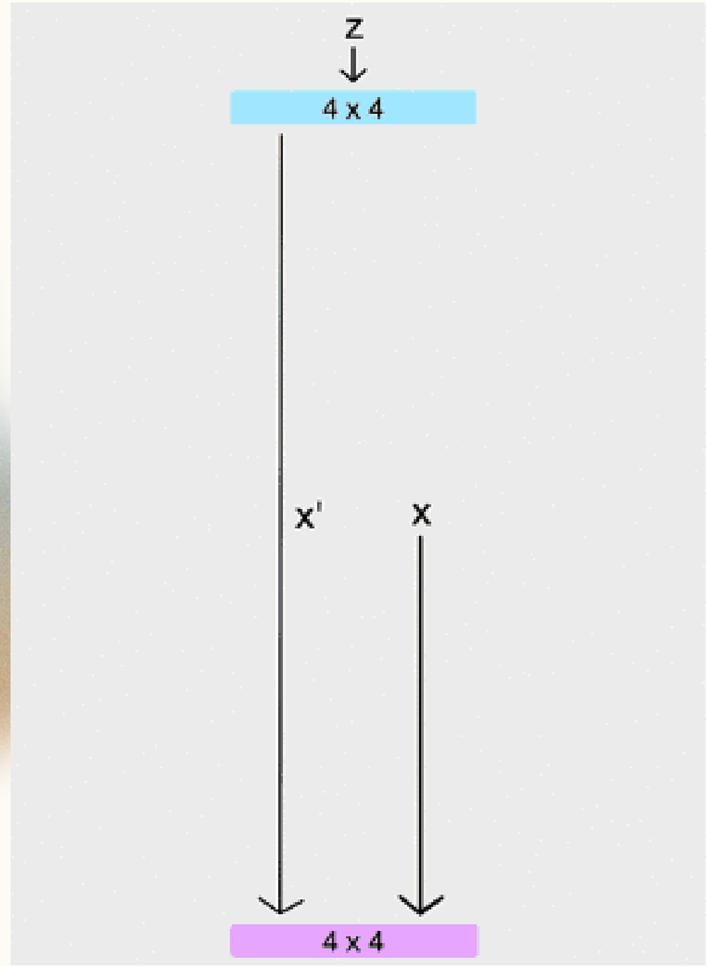
ProGAN

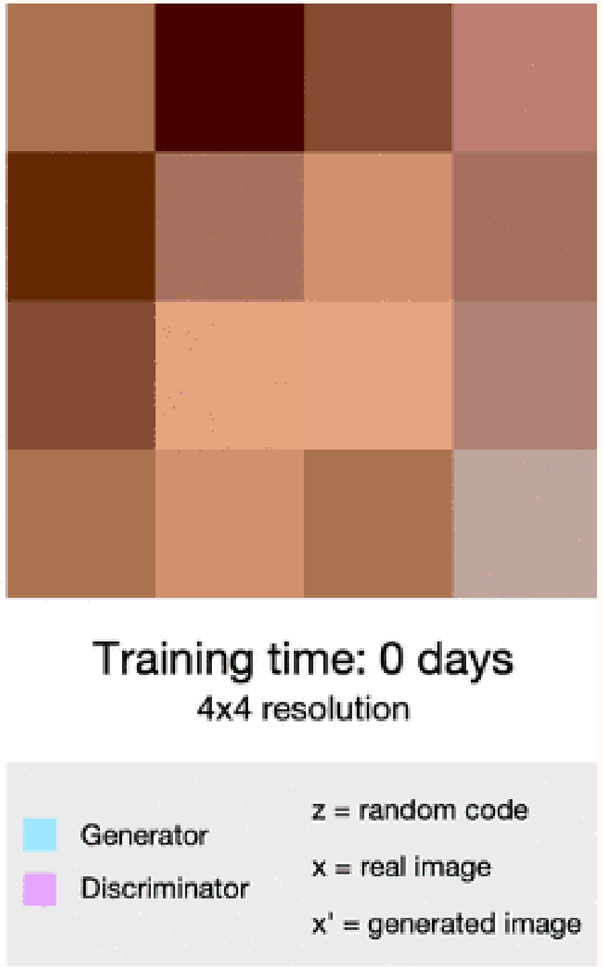
Progressive Growing
Generative Adversarial Networks

Main Limitation

There is no control over either the generation process or the image's **style**.

Karras et al. 2017



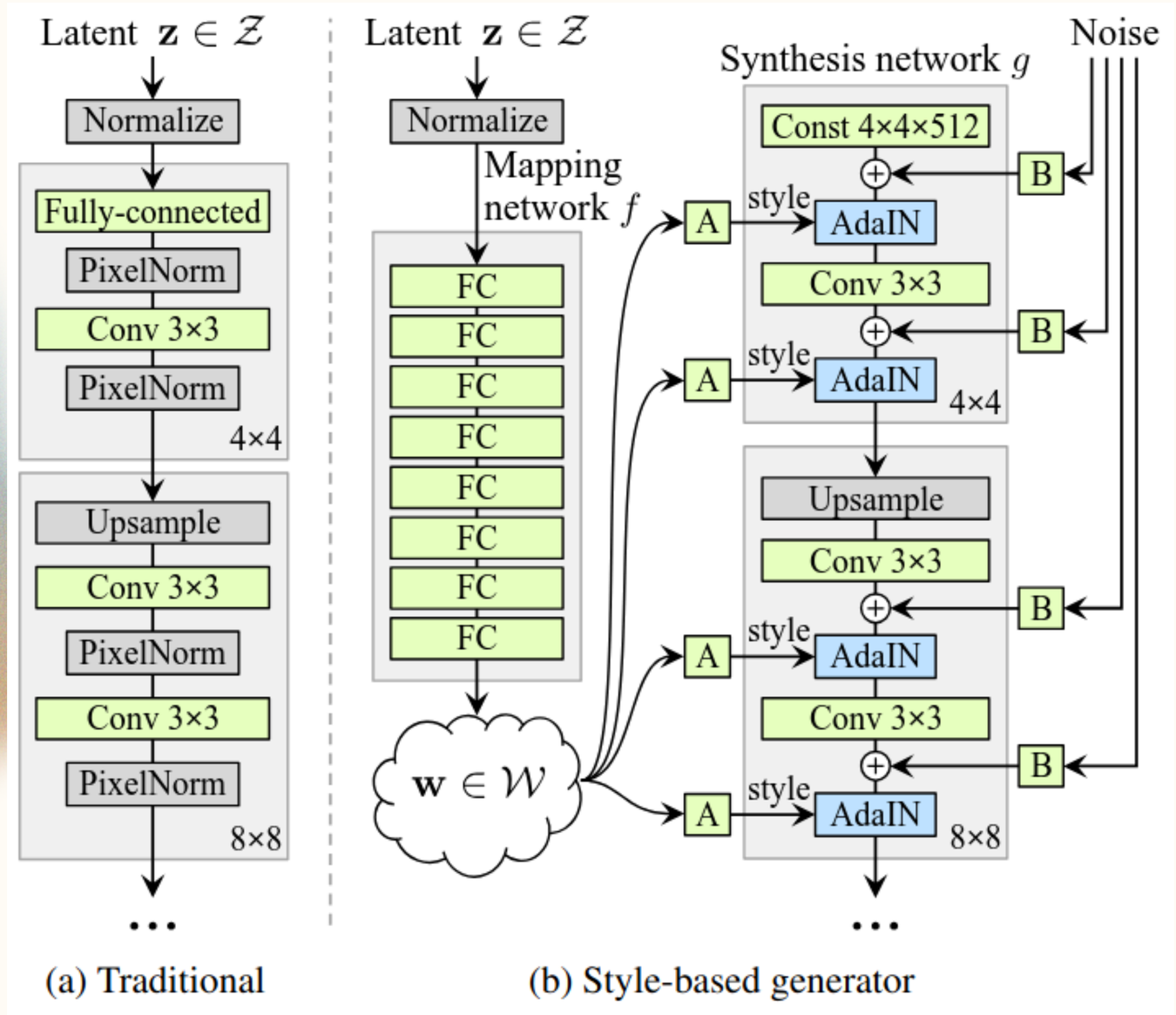


Fonte

StyleGAN

Generator Architecture

- Mapping network of 8 fully connected layers.
- Learned transformations **A** that specialize w to styles $y = (y_s,\ y_b)$.



Karras et al. 2018

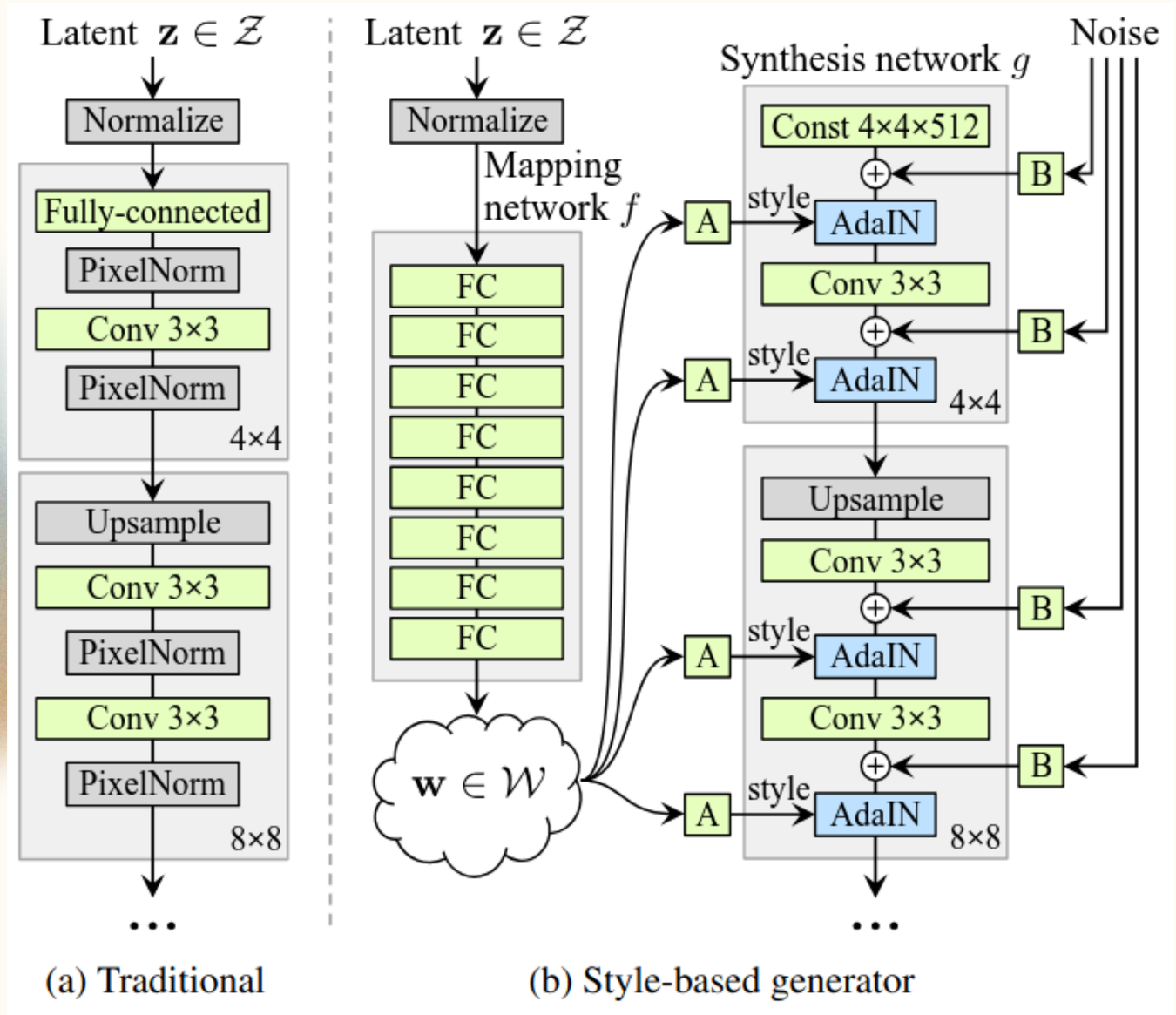
StyleGAN

Generator Architecture

- Add Gaussian noise for each channel.
- Usage of Adaptive Instance Normalization (**AdaIN**).

$$\mathbf{y}_{s,i}\frac{\mathbf{x}_i-\mu(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)}+\mathbf{y}_{b,i}$$

Karras et al. 2018



Style Mixing

Generator thinks of an image as a collection of **styles**, where each style controls the effects at a particular scale:

- Coarse styles → pose, hair, face shape
- Middle styles → facial features, eyes
- Fine styles → color scheme

Karras et al. 2018

Limitations

Characteristics Artifacts

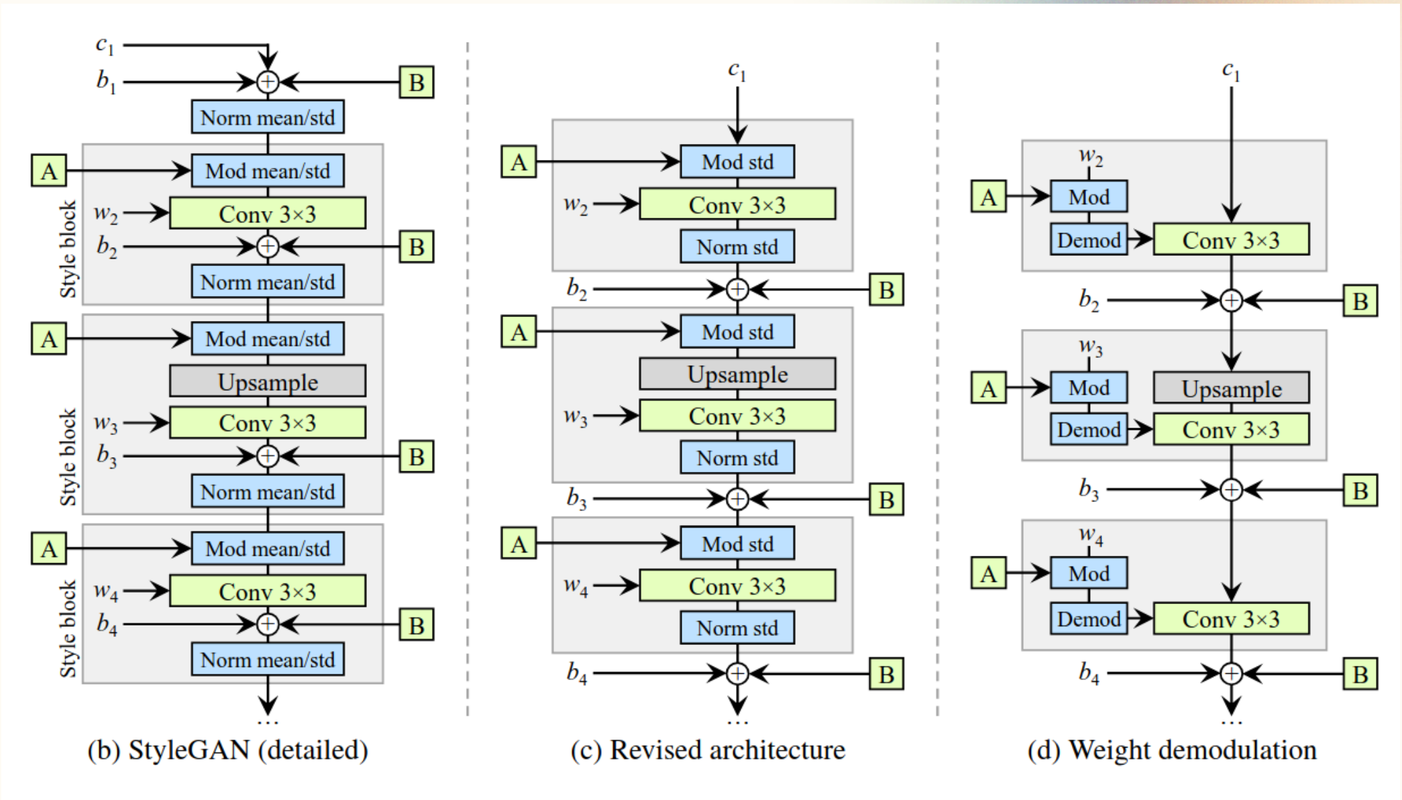
Although it was the state-of-the-art unconditional generative image model at the time, it still produced characteristic artifacts like blobs and phase issues.

StyleGAN 2

The paper analyzes the limitations and proposes 5 key improvements:

- **Redesign of generator normalization:** AdaIN causes droplets artifacts, proposes weight demodulation instead.
- **Improved training dynamics:** Path length regularization for smoother latent-to-image mapping & lazy regularization.
- **Revisiting progressive growing:** It caused phase/artifact issues. Proposed skip connections and residual discriminator instead.
- **Model capacity and resolution usage:** StyleGAN underutilized its highest resolution layers. Solved with larger models.
- **Image inversion & attribution:** Path length regularization made inversion easier. Generated images can be more reliably attributed to their source model.

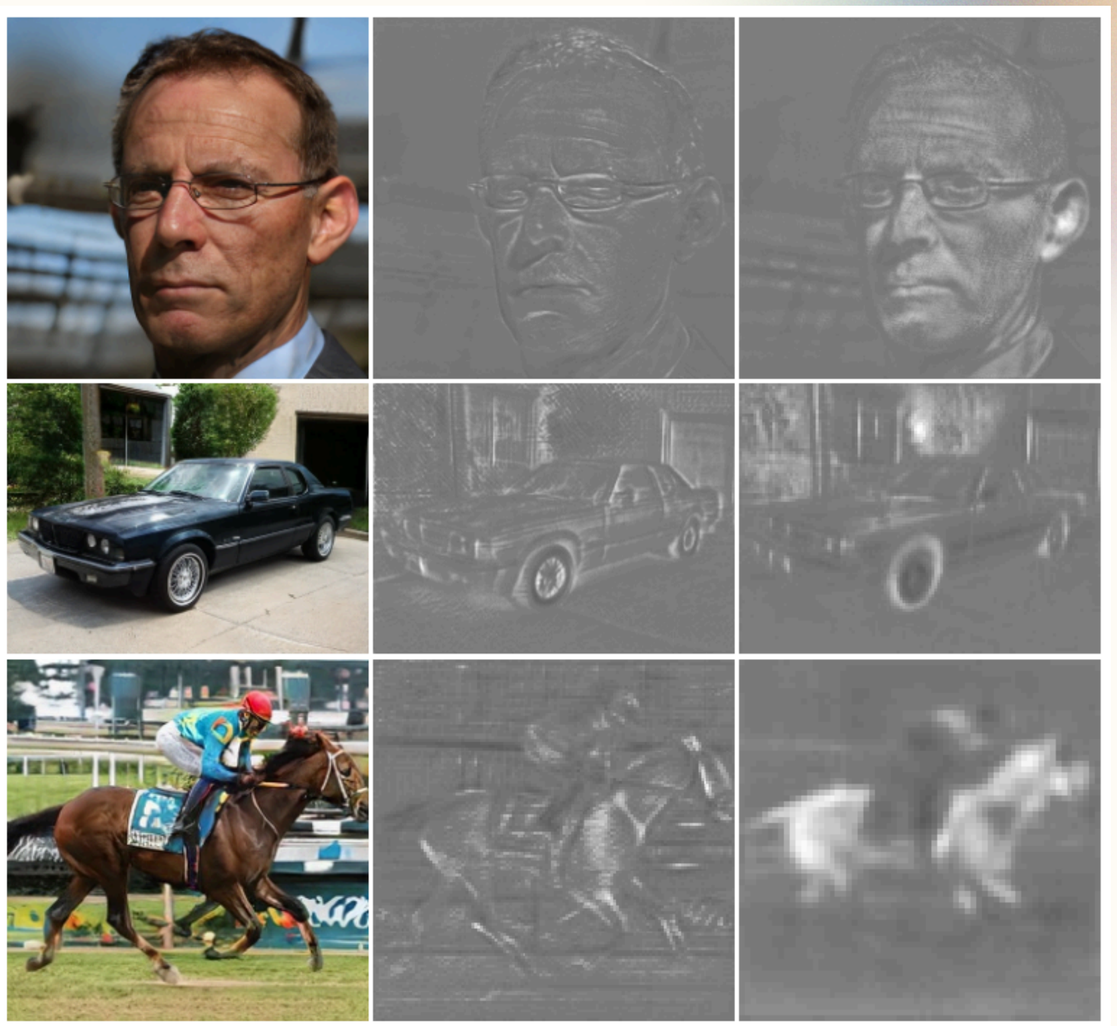
AdaIN revisited



1) Removal of the **mean** from the AdaIN and **decoupling the noise** input B from the style block.

2) Changes AdaIN to a **Modulation & Demodulation** architecture.

AdaIN revisited



Generator Smoothness

Our objective is to achieve a **smoother latent space**, higher perceptual quality, and more stable training.

Lazy Regularization

- Apply regularization less often (e.g., every 16 steps).
- Cuts training cost while keeping stability.
- Allows training larger, higher-quality models.

Path Length Regularization

- Encourages a well-conditioned mapping from latent space to image.
- Ensures smooth, consistent changes during interpolation.
- Leads to lower Perceptual Path Length (PPL) → sharper, more coherent images.
- Makes the generator easier to invert and improves editing control.

Progressive Grown

Original StyleGAN

- Trains from low to high resolutions in stages.
- Stabilizes training but introduces artifacts:
  - “Phase” issues (misaligned elements).
  - Overemphasis on high-frequency details in intermediate layers.

StyleGAN2 Solution

- Remove progressive growing.
- Use skip connections (generator) + residual design (discriminator).
- Still shifts focus from coarse, with fine details, but without artifacts.
- Produces more consistent, high-quality images across resolutions.

Resolution Usage

StyleGAN underutilized the highest resolution layers

- Generated images often looked like sharpened 512×512 outputs rather than true 1024×1024.
- Highest-resolution layers contributed less than expected.

Solution

- Usage of larger networks to increase the model's capacity.
- Leads to stronger high-resolution detail usage.

Resolution Usage

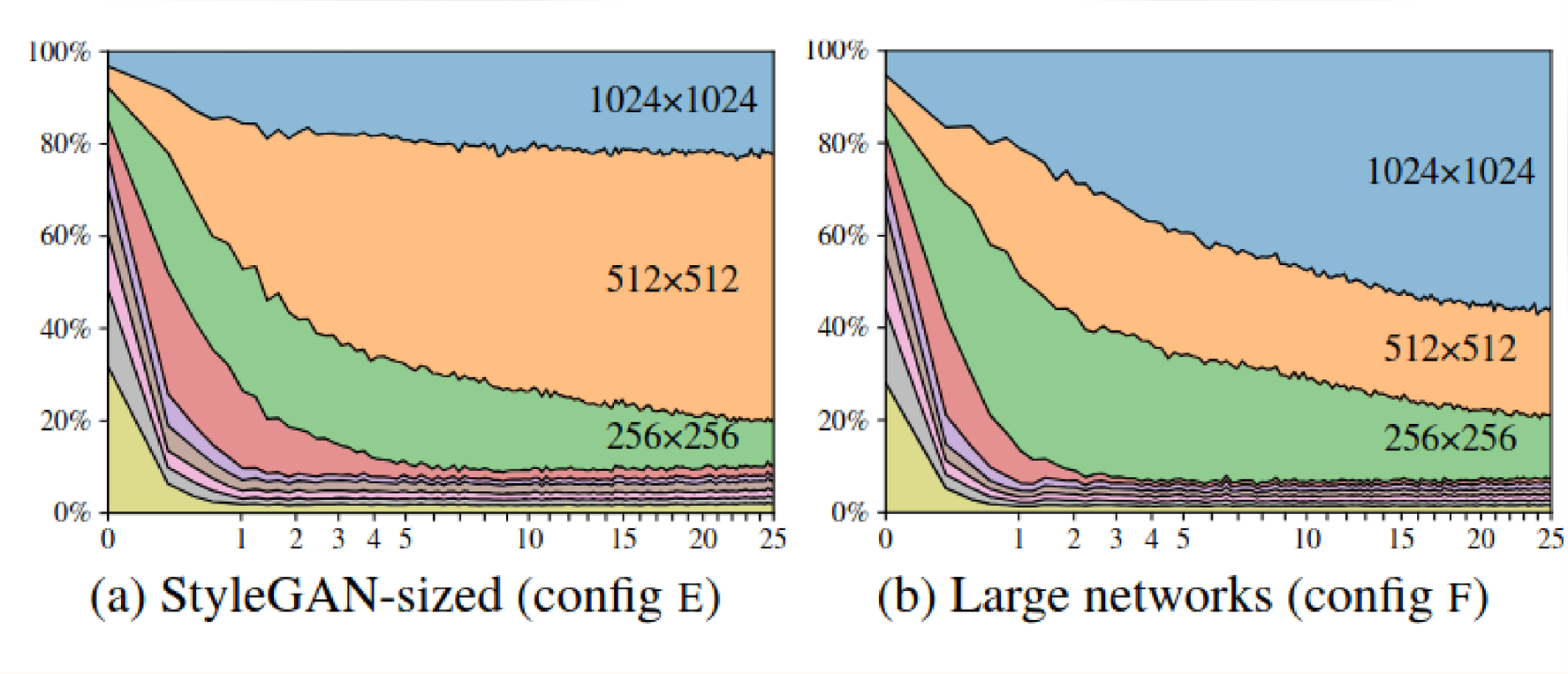


Image Inversion & Atributtion

Find latent code **w** that reproduces a given image.

- StyleGAN2 + path length regularization result in a more accurate inversion.
- Enables editing and manipulation directly in latent space.
- Generated images can be projected back almost perfectly.
- Supports source attribution: determining if an image came from a specific generator.

Results

	Configuration	FFHQ, 1024×1024				LSUN Car, 512×384			
		FID ↓	Path length ↓	Precision ↑	Recall ↑	FID ↓	Path length ↓	Precision ↑	Recall ↑
A	Baseline StyleGAN [21]	4.40	212.1	**0.721**	0.399	3.27	1484.5	**0.701**	0.435
B	+ Weight demodulation	4.39	175.4	0.702	0.425	3.04	862.4	0.685	0.488
C	+ Lazy regularization	4.38	158.0	0.719	0.427	2.83	981.6	0.688	0.493
D	+ Path length regularization	4.34	**122.5**	0.715	0.418	3.43	651.2	0.697	0.452
E	+ No growing, new G & D arch.	3.31	124.5	0.705	0.449	3.19	471.2	0.690	0.454
F	+ Large networks (StyleGAN2)	**2.84**	145.0	0.689	**0.492**	**2.32**	**415.5**	0.678	**0.514**
	Config A with large networks	3.98	199.2	0.716	0.422	–	–	–	–

Conclusion

StyleGAN2 addresses major problems of StyleGAN.

Better Training Architecture

- Lazy regularization.
- Larger model capacity.
- Skip connections and residual design.

Impact

- New state of the art model.
- Open Source.
- Enables attribution, improving control and transparency.

Summary

The StyleGAN family is a highly promising architecture for state of the art image synthesis.

Difficult points

Understand statistical metrics

Diagrams interpretation

Need to understand base article

AI Perspective

Main contributions

- **Fixes artifacts in StyleGAN**
  - Replaces instance normalization with **weight demodulation** → removes “droplet” artifacts.
  - Revisits progressive growing → eliminates “phase” issues with skip connections & residual design.
- **Improves image quality & stability**
  - Introduces **path length regularization** → smoother latent space, better interpolations.
  - Uses **lazy regularization** → same quality with reduced compute.
- **Enhances resolution usage**
  - Identifies underuse of high-res layers.
  - Increases **model capacity** → sharper, more detailed images.
- **Better inversion & attribution**
  - Generated images can be projected back to latent codes more reliably.
  - Improves editing and enables forensic detection of GAN-generated images.

GPT5

AI Perspective

Limitations & Future work

Limitations / Open Questions

- Still **data-hungry** → requires large, diverse datasets.
- Training remains **computationally expensive** despite optimizations.
- Quality metrics (FID, PPL, Precision/Recall) don't fully capture **human perception**.
- Inversion still imperfect for **real images** (gap between real and generated domains).

Future Research Directions

- **Smarter regularization**: use feature-space metrics (not pixel L2) for path length regularization.
- **Data efficiency**: develop methods that learn from smaller or less-curated datasets.
- **Better evaluation**: design metrics aligned with human perception (shapes + semantics).
- **Controlled generation**: integrate StyleGAN2 with multimodal inputs (e.g., text + image editing).

GPT5

Thank you!