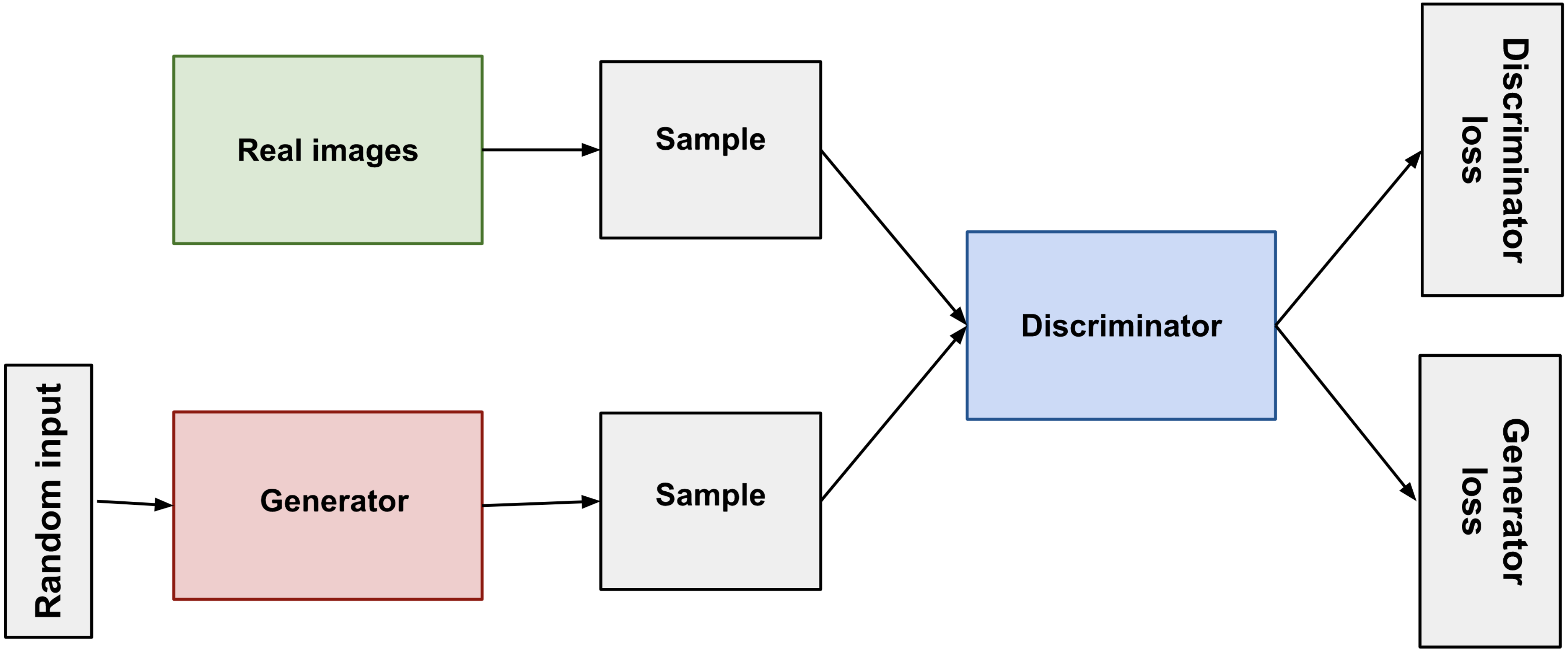
StyleGAN

Sergey Kim

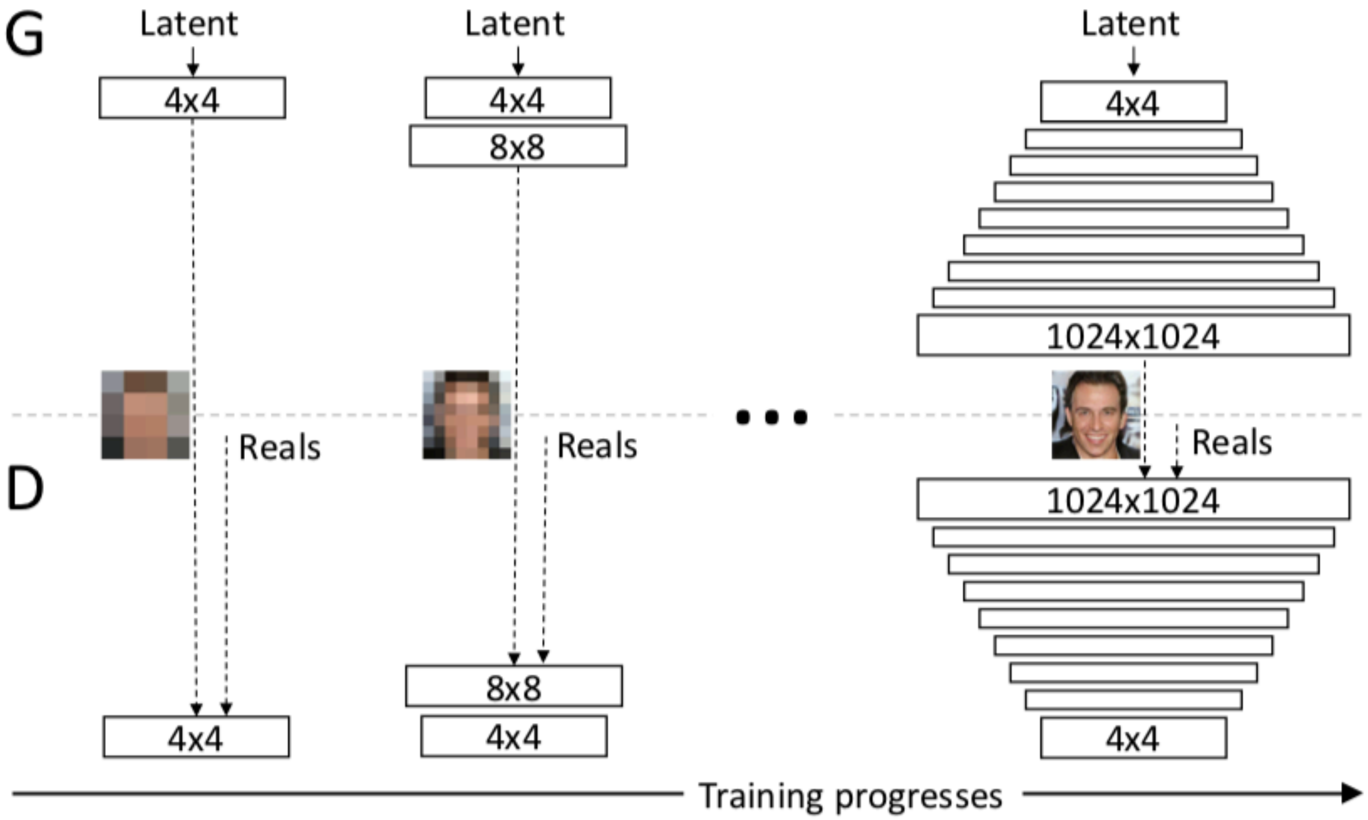
In the previous series

- GAN
- ProGAN

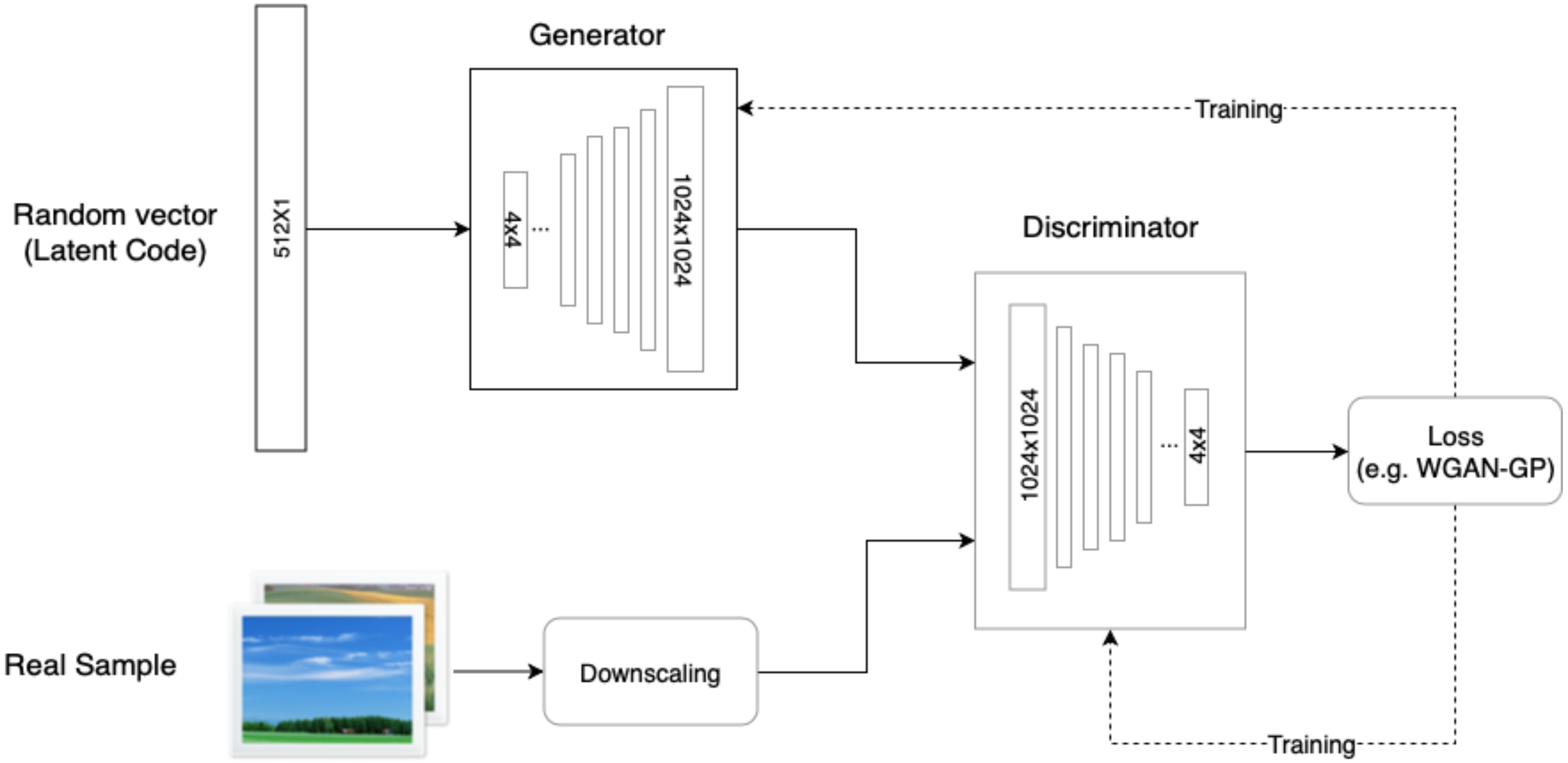
GAN – Generative Adversarial Network



ProGAN


G
Latent
4x4
Latent
4x4
8x8
Latent
4x4
1024x1024
Reals
Reals
Reals
D
1024x1024
8x8
4x4
4x4
4x4
Training progresses

ProGAN


Random vector
(Latent Code)
512X1
Generator
4x4
...
1024x1024
Training
Discriminator
1024x1024
...
4x4
Loss
(e.g. WGAN-GP)
Real Sample
Downscaling
Training

Evolution

Oct 2017 – ProGAN (https://arxiv.org/abs/1710.10196)

Dec 2018 – StyleGAN (https://arxiv.org/abs/1812.04948)

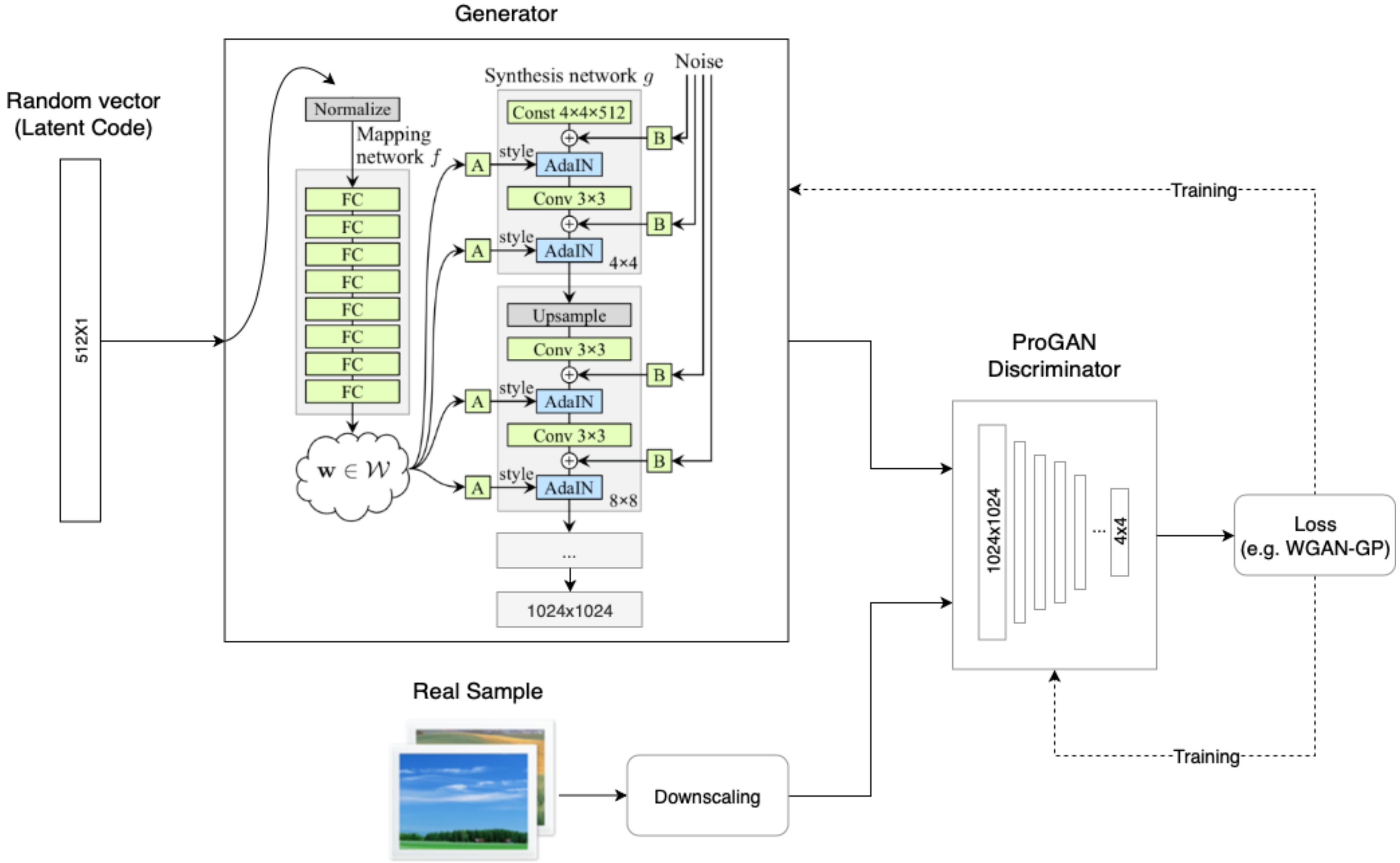
Dec 2019 – StyleGAN2 (https://arxiv.org/abs/1912.04958)

NVIDIA

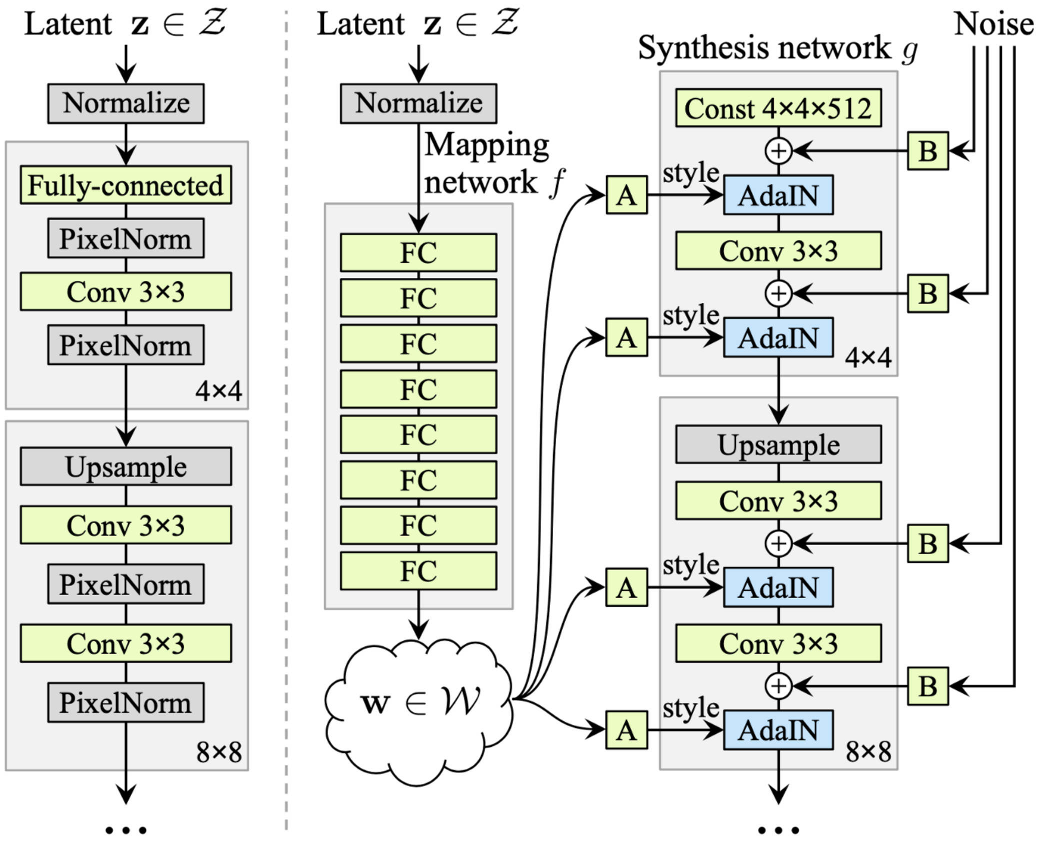
StyleGAN

- Architecture
- Regularization
- Metrics
- Results

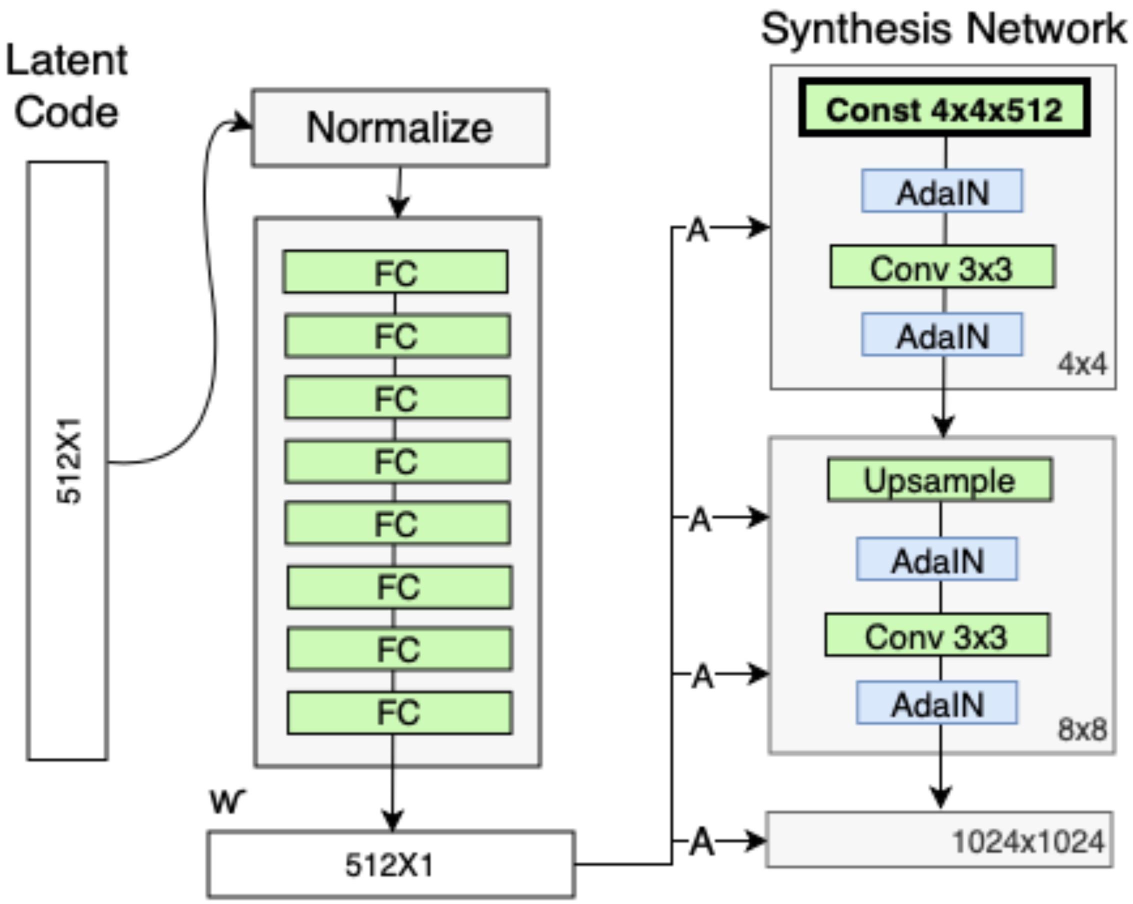
StyleGAN. Architecture



StyleGAN. Generator



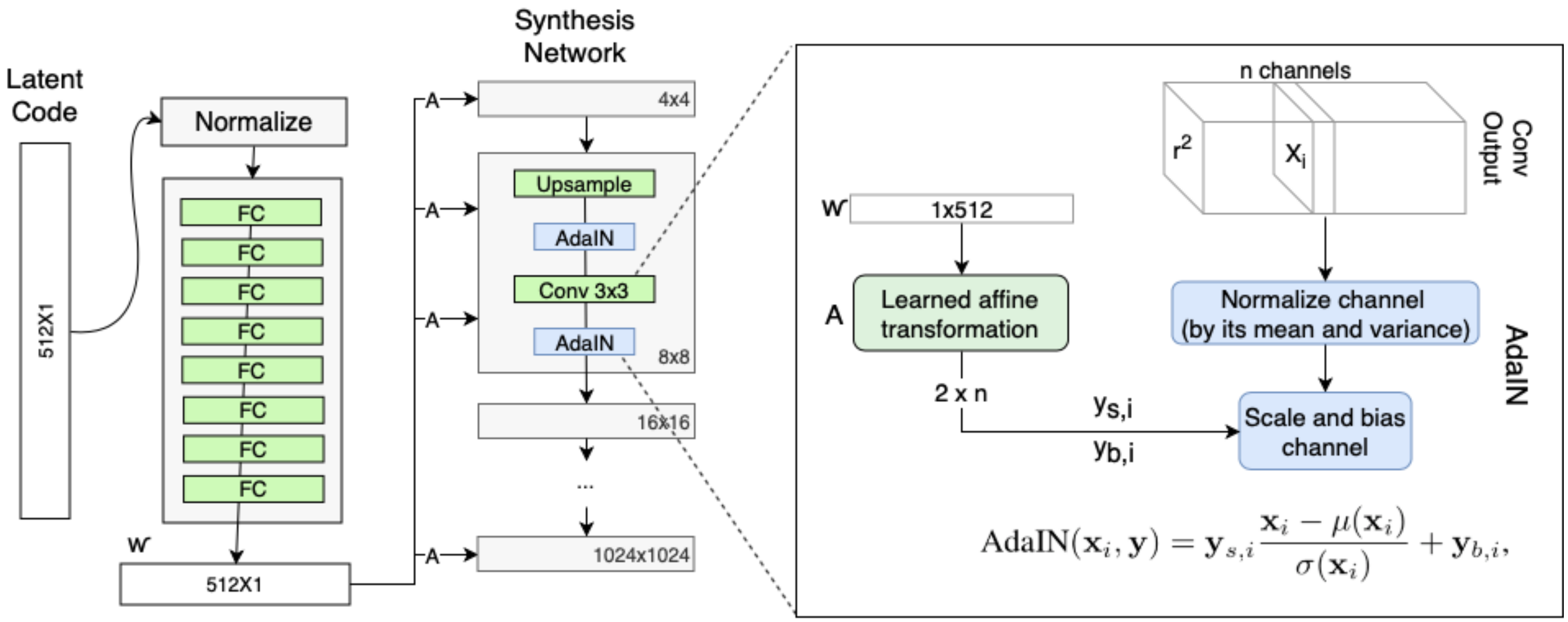
(a) Traditional (b) Style-based generator



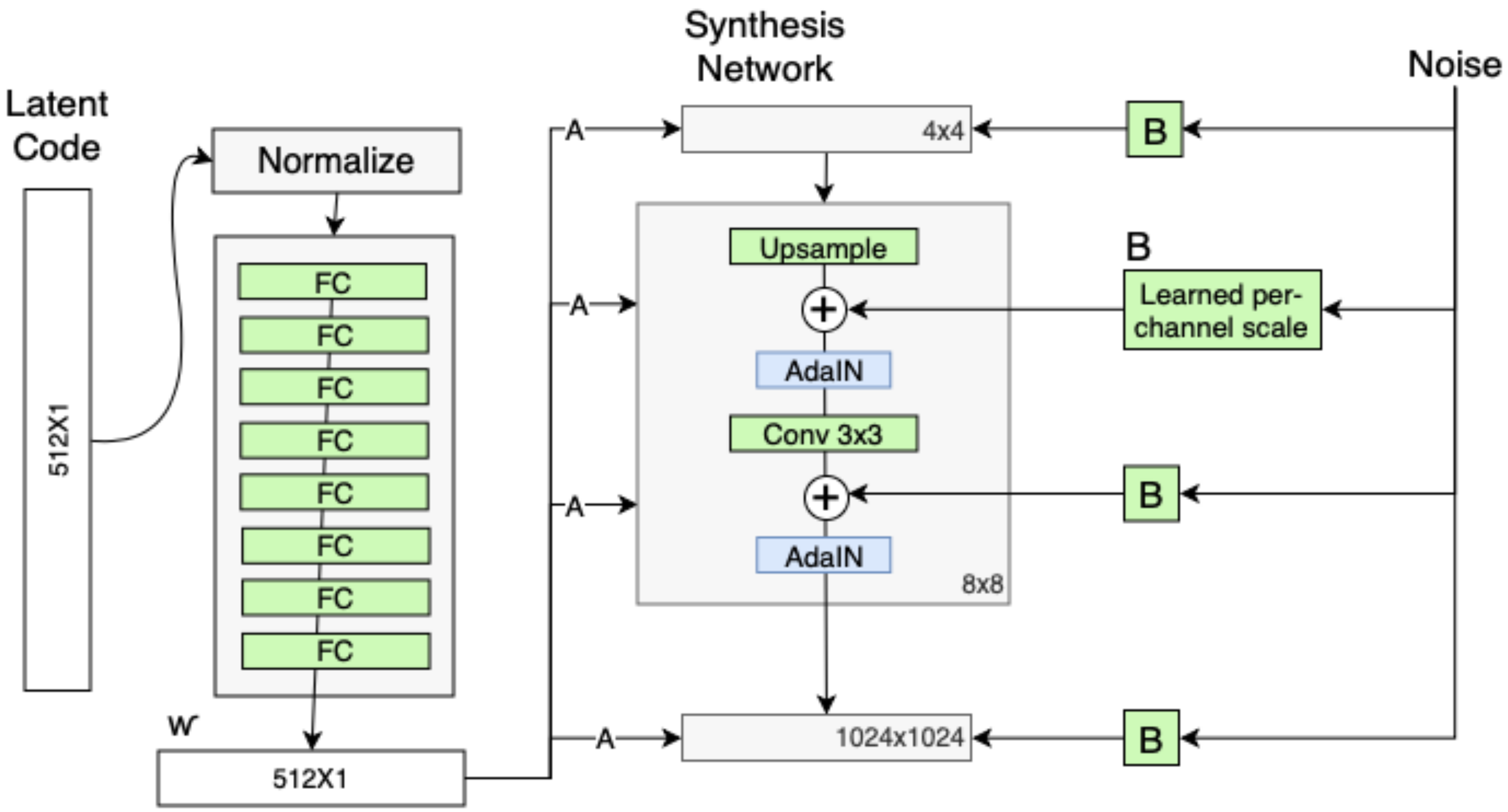
StyleGAN. Adaptive Instance Normalization

$$\mathbf{AdaIN}(x, y) = \sigma(y)\left(\frac{x - \mu(x)}{\sigma(x)}\right) + \mu(y)$$

StyleGAN. Adaptive Instance Normalization



StyleGAN. Stochastic Variation



StyleGAN. Mixing Regularization

StyleGAN. Mixing Regularization

StyleGAN. Mixing Regularization

StyleGAN. Evaluation Metrics

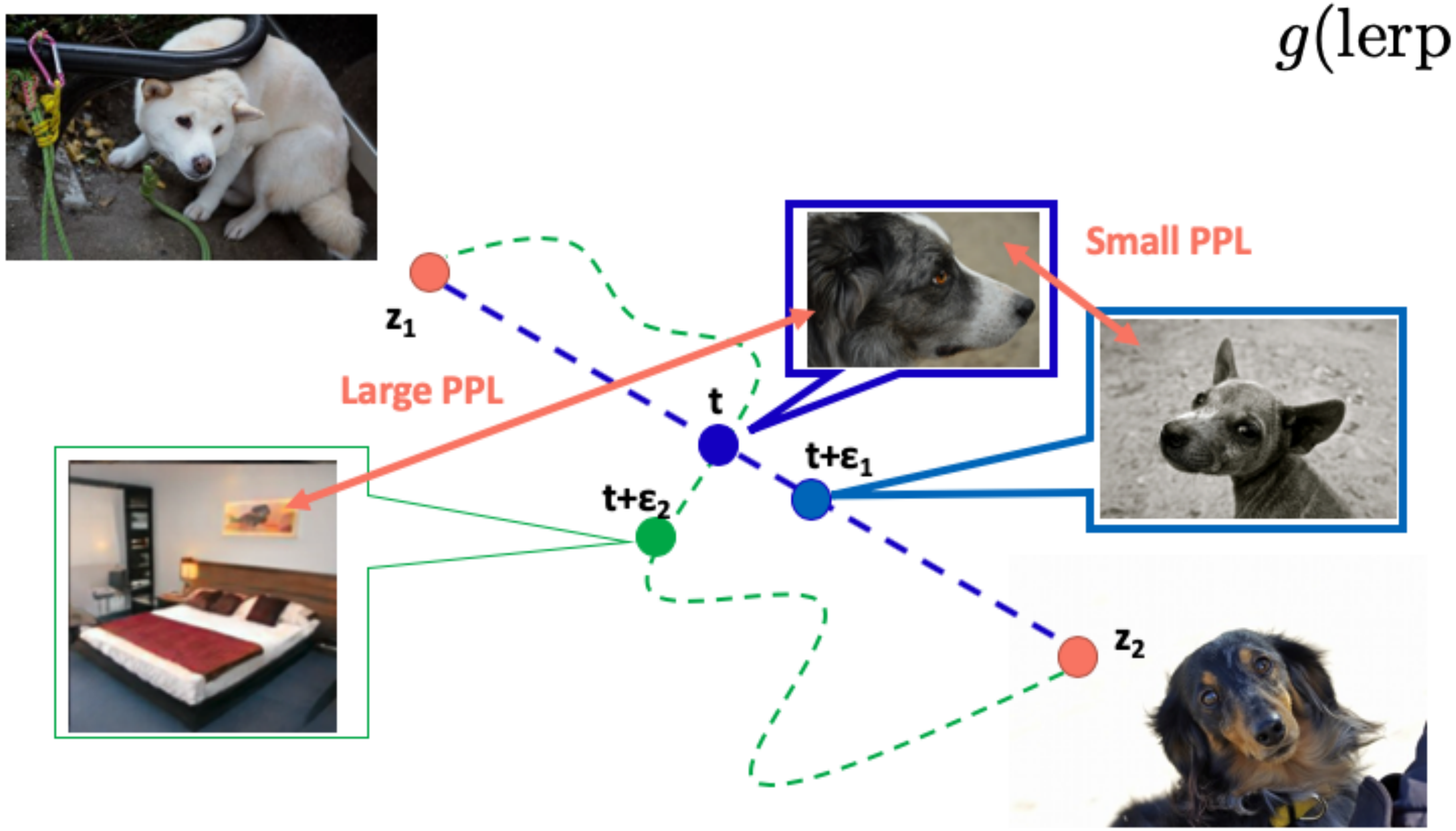
The **<u>Fréchet Inception Distance</u>** is a measure of similarity between curves that depends on the location and ordering of the points along the curves (a.k.a Wasserstein-2 distance between two multi-variate Gaussians)

$$\|\boldsymbol{m} - \boldsymbol{m}_w\|_2^2 + \mathrm{Tr}\left(\boldsymbol{C} + \boldsymbol{C}_w - 2(\boldsymbol{C}\boldsymbol{C}_w)^{1/2}\right)$$

StyleGAN. Evaluation Metrics

Perceptual Path Lengths

$$l_{\mathcal{W}} = \mathbb{E}\Big[\frac{1}{\epsilon^2} d\big(g(\mathrm{lerp}(f(\mathbf{z}_1), f(\mathbf{z}_2); t)), g(\mathrm{lerp}(f(\mathbf{z}_1), f(\mathbf{z}_2); t+\epsilon))\big)\Big],$$

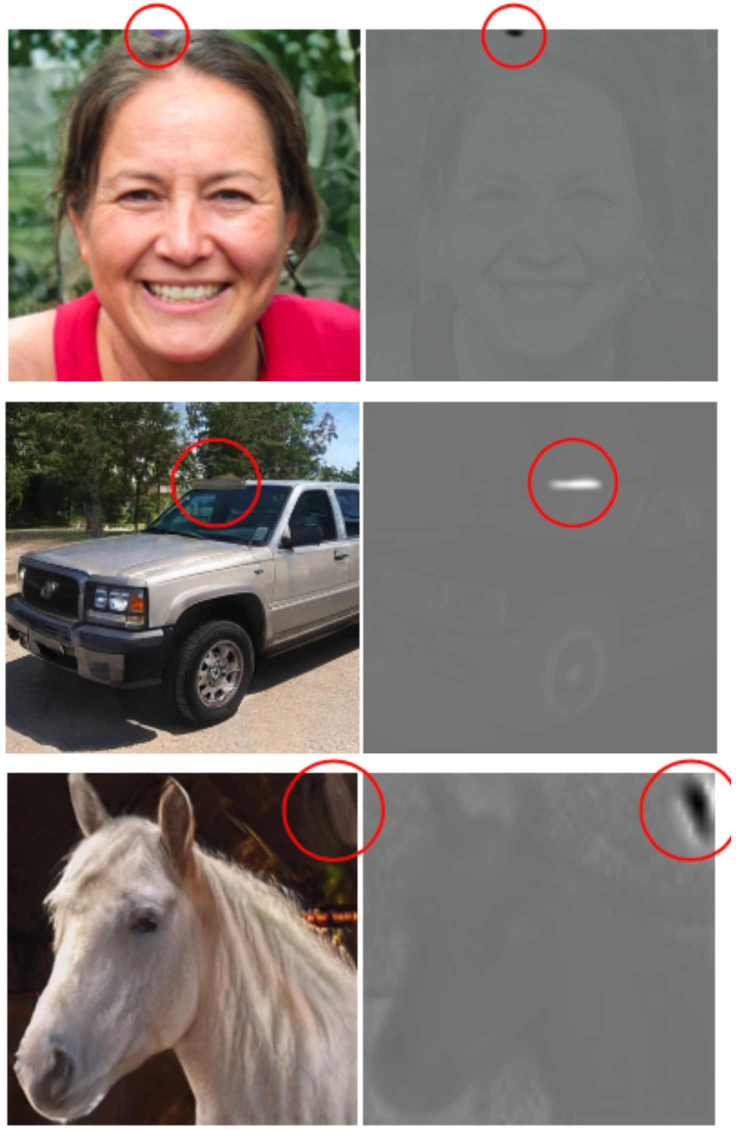


StyleGAN. Results

	Method	CelebA-HQ	FFHQ
A	Baseline Progressive GAN [30]	7.79	8.04
B	+ Tuning (incl. bilinear up/down)	6.11	5.25
C	+ Add mapping and styles	5.34	4.85
D	+ Remove traditional input	5.07	4.88
E	+ Add noise inputs	**5.06**	4.42
F	+ Mixing regularization	5.17	**4.40**

Table 1. Fréchet inception distance (FID) for various generator designs (lower is better). In this paper we calculate the FIDs using 50,000 images drawn randomly from the training set, and report the lowest distance encountered over the course of training.

StyleGAN. Problems



StyleGAN. Summary

- Architecture: ProGAN, AdaIN, Noise
- Regularization: Mixing Reg
- Metrics: FID, PPL
- Results: Nice quality, but it has some problems

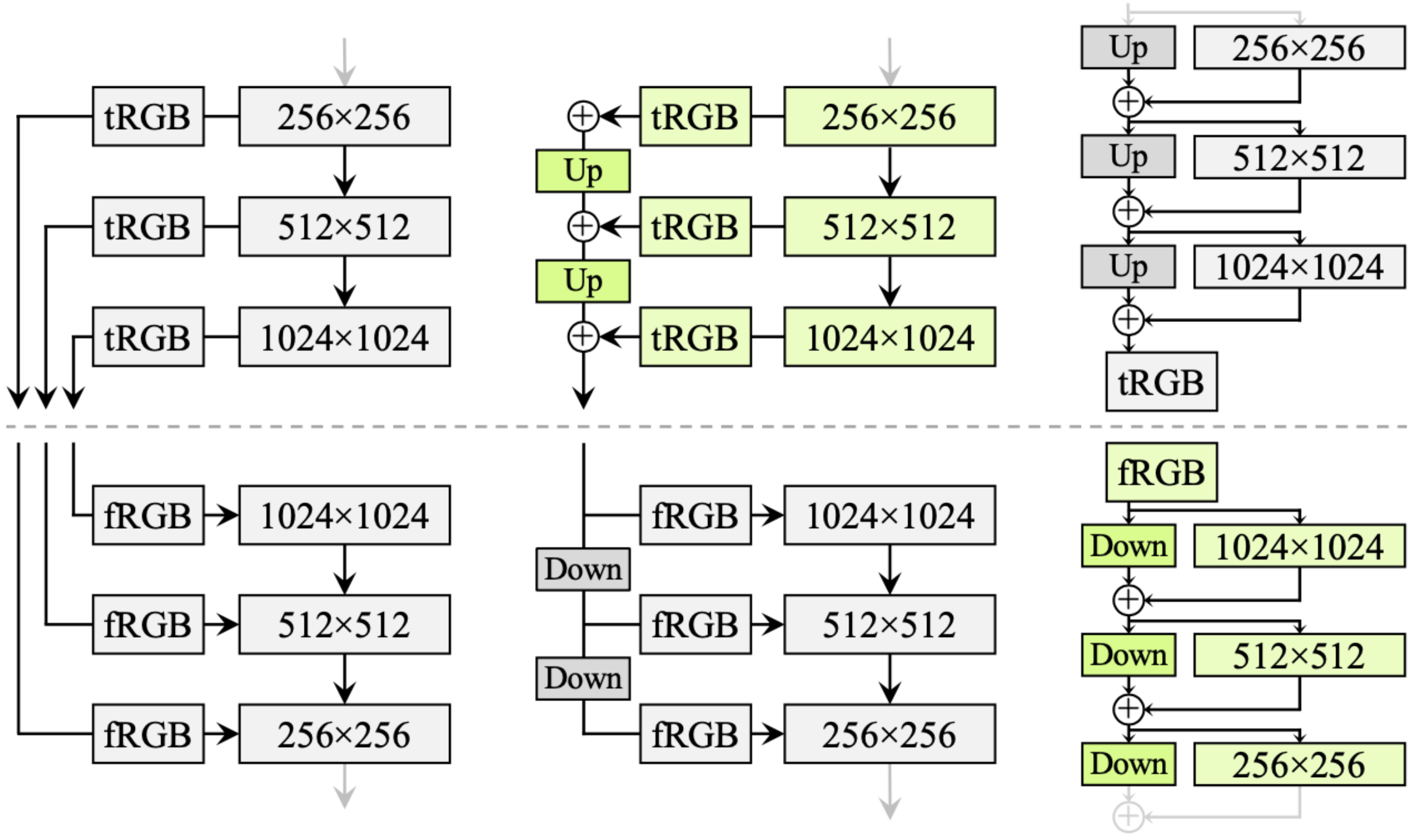
StyleGAN2. Trailer

- Architecture: ~~ProGAN~~, ~~AdaIN~~, Noise; **MSG-GAN, Custom Norm**
- Regularization: Mixing Reg; **Path Length Reg, Lazy Reg**
- Metrics: FID, **PPL**
- Results: Nice quality, ~~but it has some problems~~

StyleGAN2

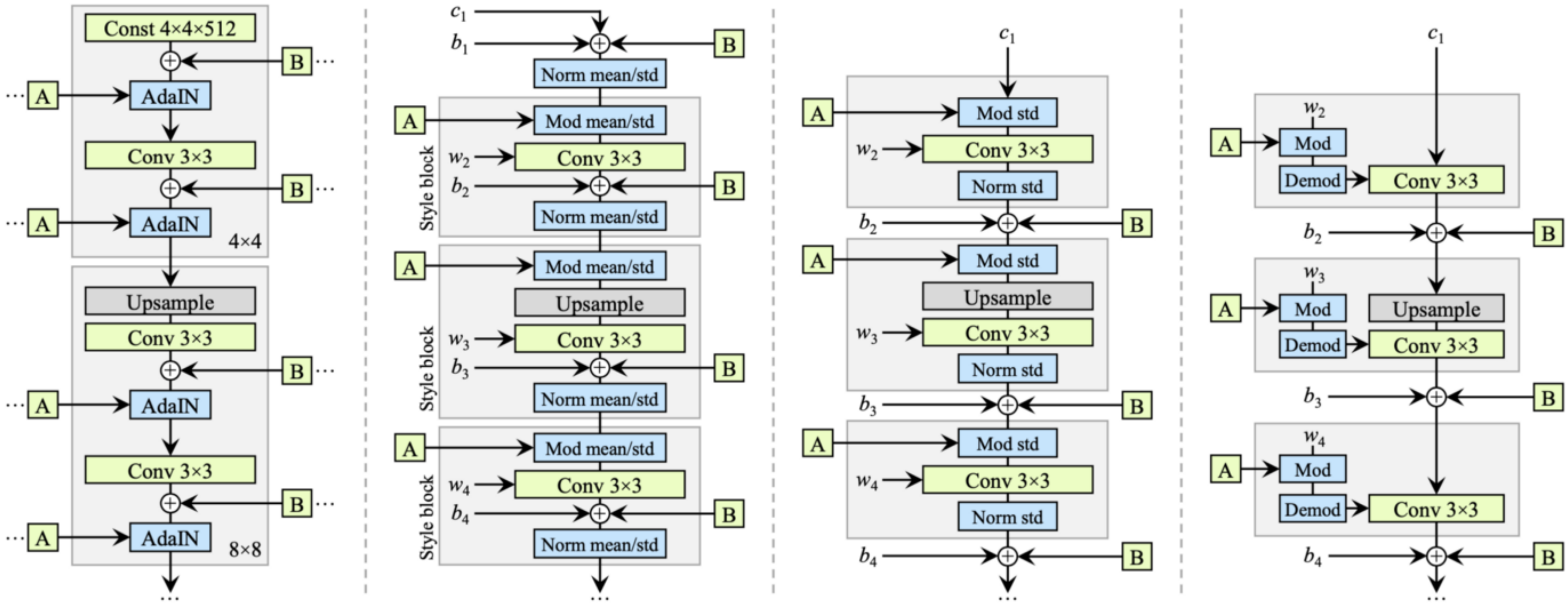
- Architecture
- Regularization
- Metrics
- Results

StyleGAN2. Architecture



(a) MSG-GAN (b) Input/output skips (c) Residual nets

StyleGAN2. Custom Normalization



(a) StyleGAN (b) StyleGAN (detailed) (c) Revised architecture (d) Weight demodulation

StyleGAN2. Path Length Regularization

(a) Low PPL scores

(b) High PPL scores

StyleGAN2. Path Length Regularization

$$\mathbb{E}_{\mathbf{w},\mathbf{y}\sim\mathcal{N}(0,\mathbf{I})}\left(\left\|\mathbf{J}_{\mathbf{w}}^{T}\mathbf{y}\right\|_2 - a\right)^2,$$

$$\mathbf{J}_{\mathbf{w}} = \partial g(\mathbf{w})/\partial\mathbf{w}$$

Now we can invert the synthesis network!

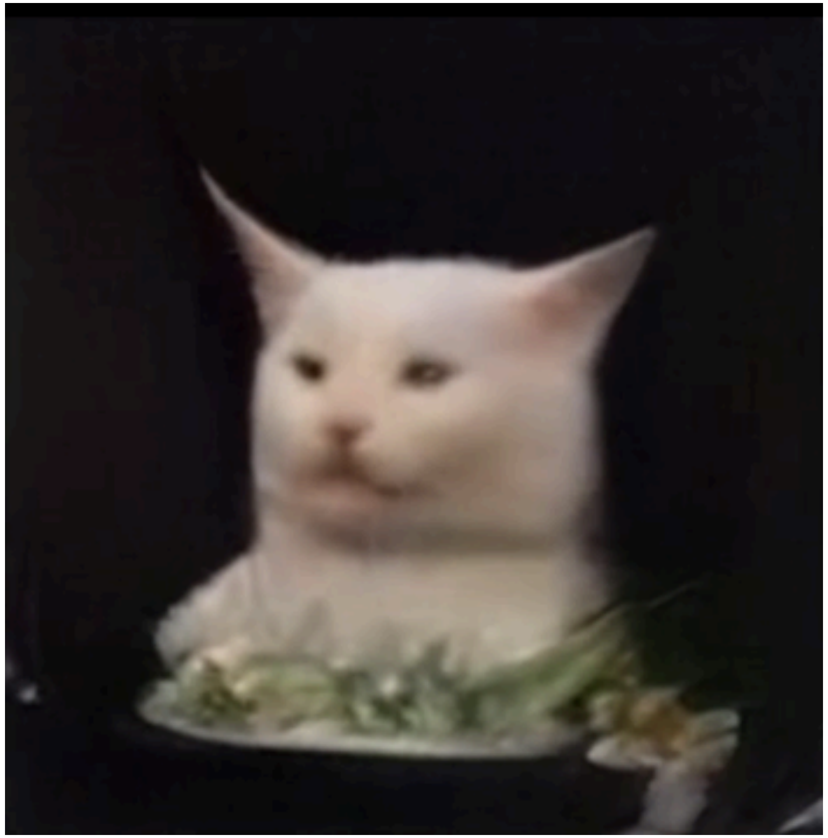
StyleGAN2. Path Length Regularization

StyleGAN — generated images

StyleGAN2. Path Length Regularization

StyleGAN2 — real images

StyleGAN2. Path Length Regularization



StyleGAN2. Lazy Regularization

We enable the regularization every k^{th} batch

k = 16 for Discriminator

k = 8 for Generator

StyleGAN2. Results

	Configuration	FFHQ, 1024×1024 FID	Path length	Precision	Recall	LSUN Car, 512×384 FID	Path length	Precision	Recall
A	Baseline StyleGAN [24]	4.40	195.9	**0.721**	0.399	3.27	1484.5	**0.701**	0.435
B	+ Weight demodulation	4.39	173.8	0.702	0.425	3.04	862.4	0.685	0.488
C	+ Lazy regularization	4.38	167.2	0.719	0.427	2.83	981.6	0.688	0.493
D	+ Path length regularization	4.34	139.2	0.715	0.418	3.43	651.2	0.697	0.452
E	+ No growing, new G & D arch.	3.31	**116.7**	0.705	0.449	3.19	471.2	0.690	0.454
F	+ Large networks	**2.84**	129.4	0.689	**0.492**	**2.32**	**415.5**	0.678	**0.514**

Table 1. Main results. For each training run, we selected the training snapshot with the lowest FID. We computed each metric 10 times with different random seeds and report their average. The "path length" column corresponds to the PPL metric, computed based on path endpoints in $\mathcal{W}$ [24]. For LSUN datasets, we report path lengths without the center crop that was originally proposed for FFHQ. The FFHQ dataset contains 70k images, and we showed the discriminator 25M images during training. For LSUN CAR the corresponding numbers were 893k and 57M.

StyleGAN2. Results

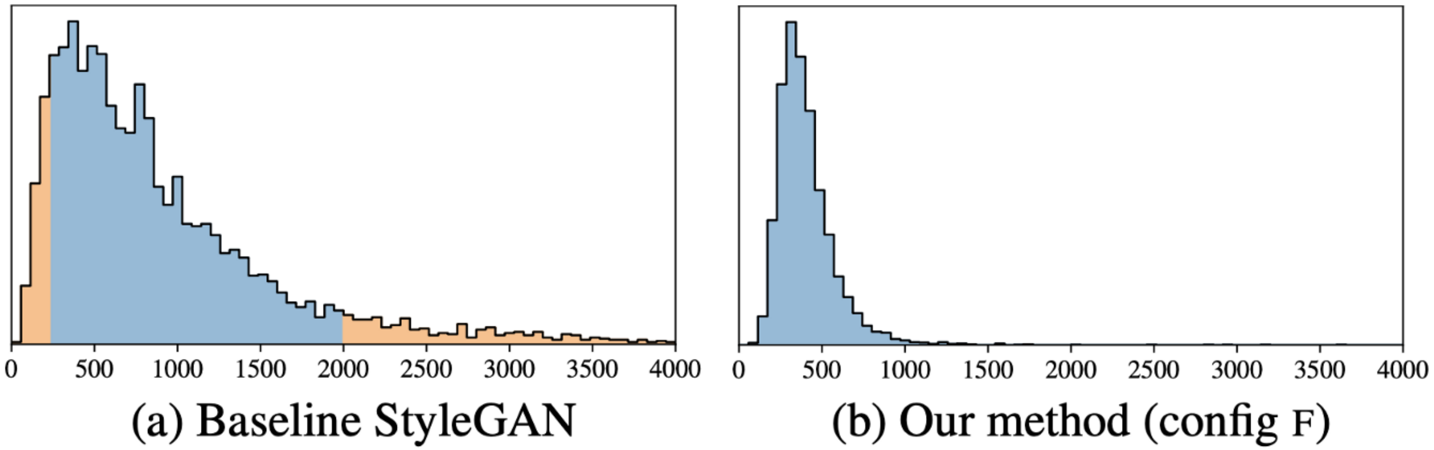


Figure 5. (a) Distribution of PPL scores of individual images generated using a baseline StyleGAN (config A in Table 1, FID = 8.53, PPL = 924). The percentile ranges corresponding to Figure 4 are highlighted in orange. (b) Our method (config F) improves the PPL distribution considerably (showing a snapshot with the same FID = 8.53, PPL = 387).

True Summary

- Architecture: ProGAN, AdaIN, Noise; **MSG-GAN, Custom Norm**
- Regularization: Mixing Reg; **Path Length Reg, Lazy Reg**
- Metrics: FID, **PPL**
- Results: Nice quality, but it has some problems

Bibliography

ProGAN (2 - 6). https://arxiv.org/abs/1710.10196

StyleGAN (7 - 21). https://arxiv.org/abs/1812.04948

StyleGAN2 (22 - 32). https://arxiv.org/abs/1912.04958