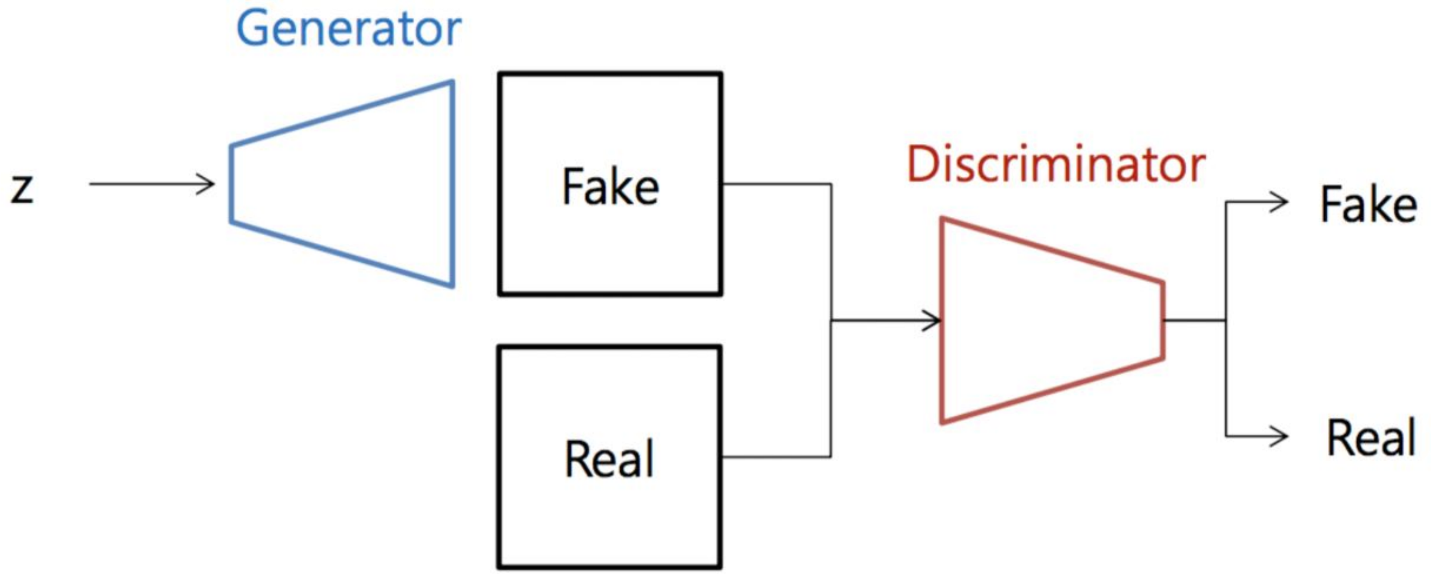
StyleGANs

Bonich Dmitrij

Presentation plan

1. Classic GAN reminder
2. Progressive GAN
3. StyleGAN-1
4. StyleGAN-2

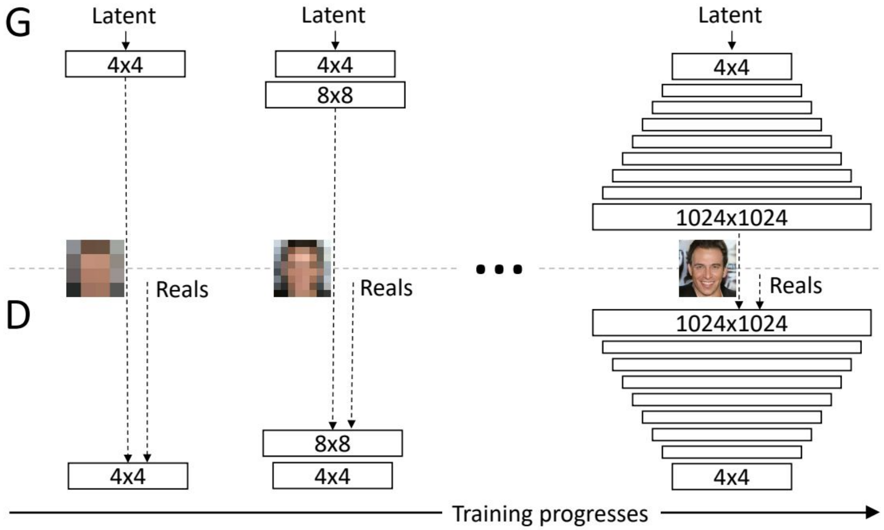
Classic GAN



$$\min_G \max_D V(D, G) = \mathbb{E}_{\boldsymbol{x} \sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z} \sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))]$$

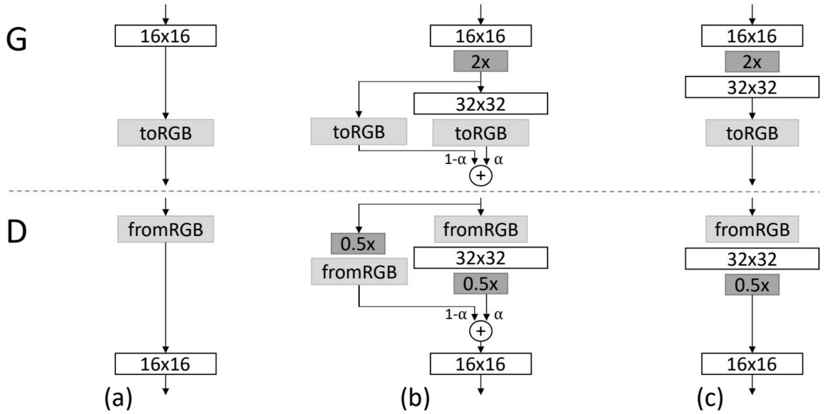
Progressive GAN

- Adds higher resolution layers to generator and discriminator progressively
- Uses Wasserstein loss for training
- Inserts minibatch statistics into features to discourage mode collapse
- Other tricks: equalized learning rate, normalization layers, etc

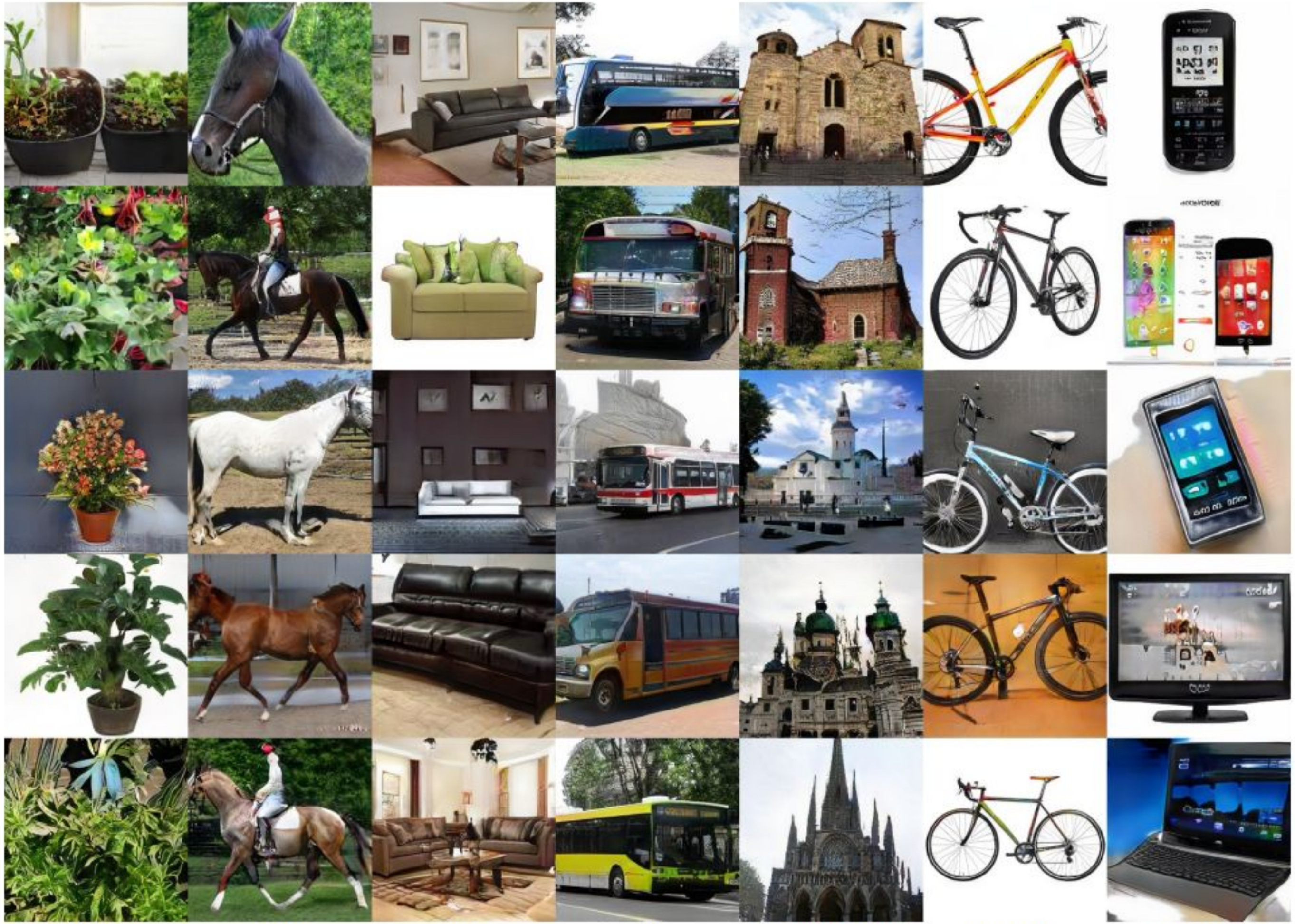


Resolution transition

1. Add new resolution block
2. Add skip connection from old resolution block
3. Linearly increase new resolution weight **α** from **0** to **1**
4. When **α=1** remove skip connection



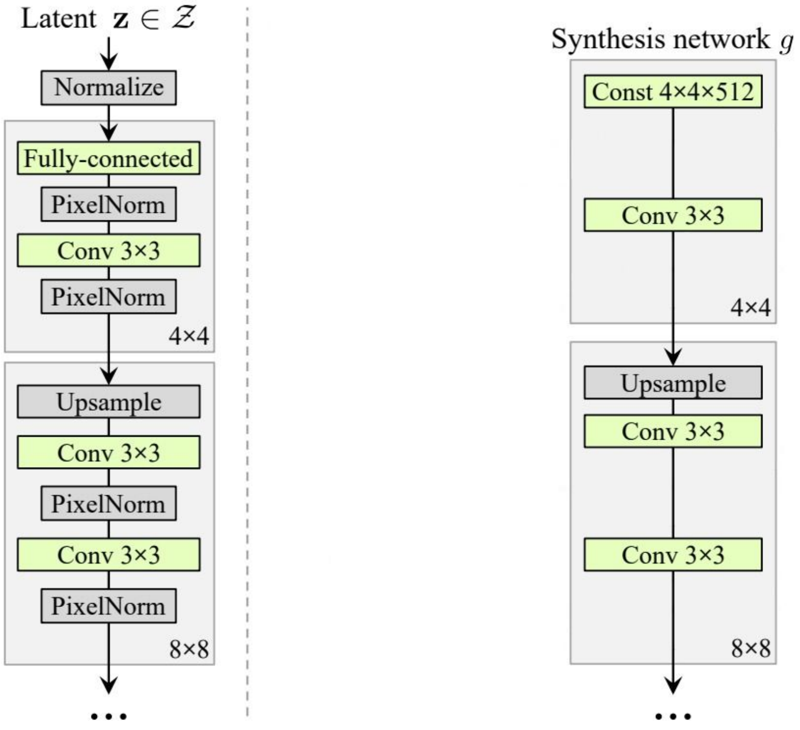
Generation examples



POTTEDPLANT HORSE SOFA BUS CHURCHOUTDOOR BICYCLE TVMONITOR

StyleGAN-1 (well, not quite)

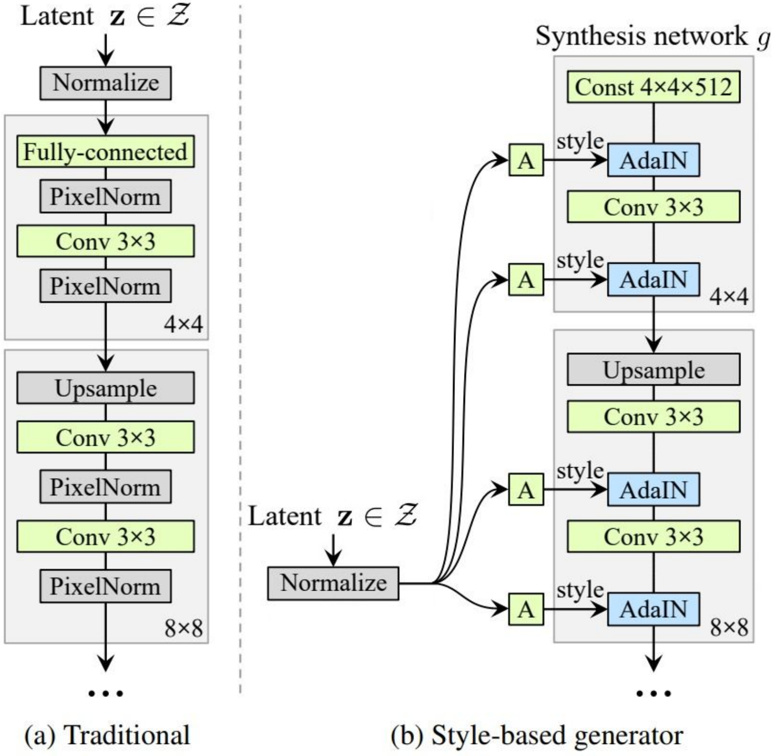
- Remove latent input
- Generator network takes const learnable tensor as input



(a) Traditional (b) Style-based generator

Adding styles

- Now latent vector **z** impacts generator via **AdaIN** layer
- Block “**A**” denotes linear layer. It transforms **z** into into **2** vectors: $\mathbf{y_s}$, $\mathbf{y_b}$, which we call styles

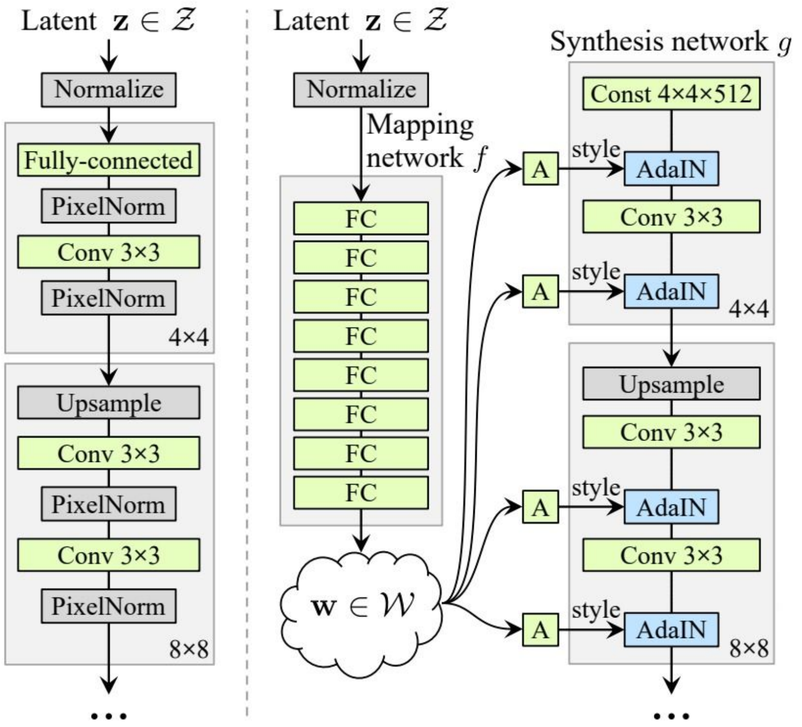


AdaIN

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

StyleGAN-1 (almost there)

- Map **z** to **w** using 8-layer MLP
- Space of **w** should be more disentangled than space of **z**

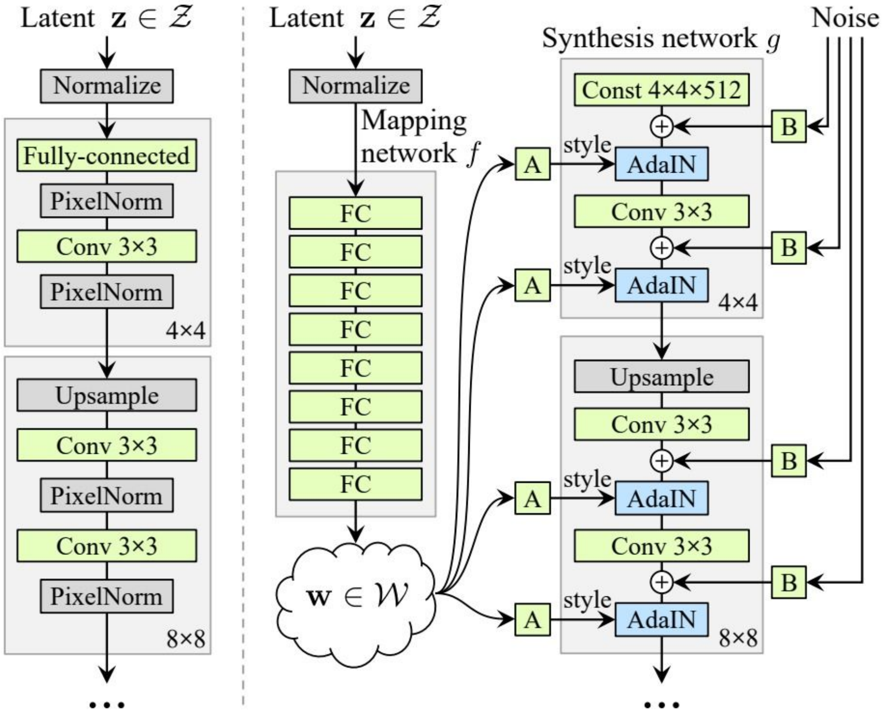


(a) Traditional

(b) Style-based generator

StyleGAN-1 (finally)

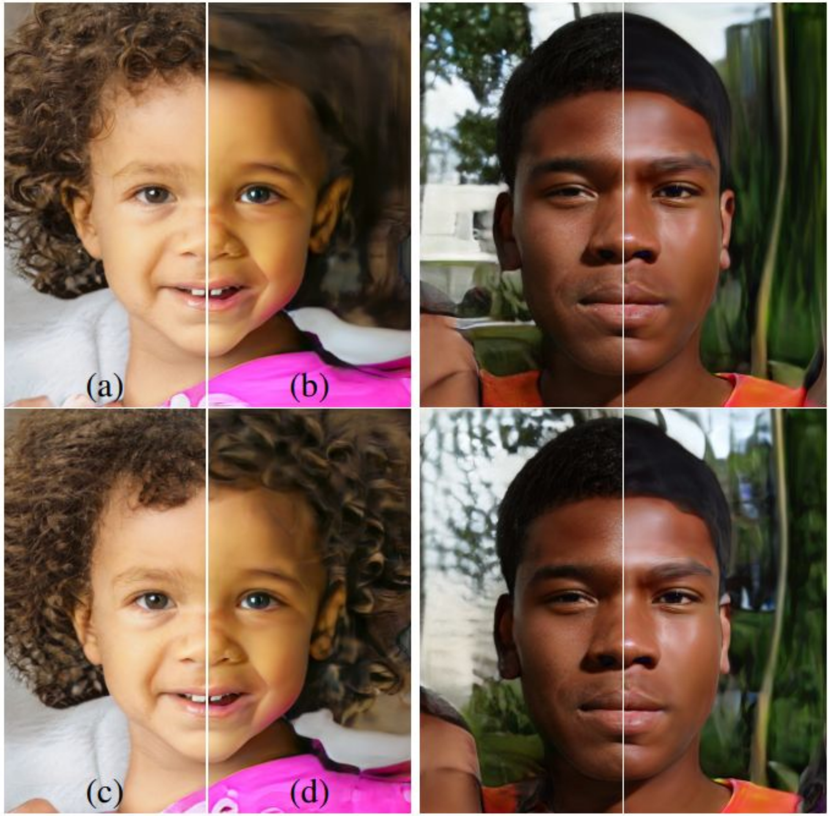
- Different noise $\mathcal{N}(\mathbf{0}, \mathtt{I})$ is passed into each “**B**” block
- “**B**” block implements per-channel scaling by learnable coefficients



(a) Traditional
(b) Style-based generator

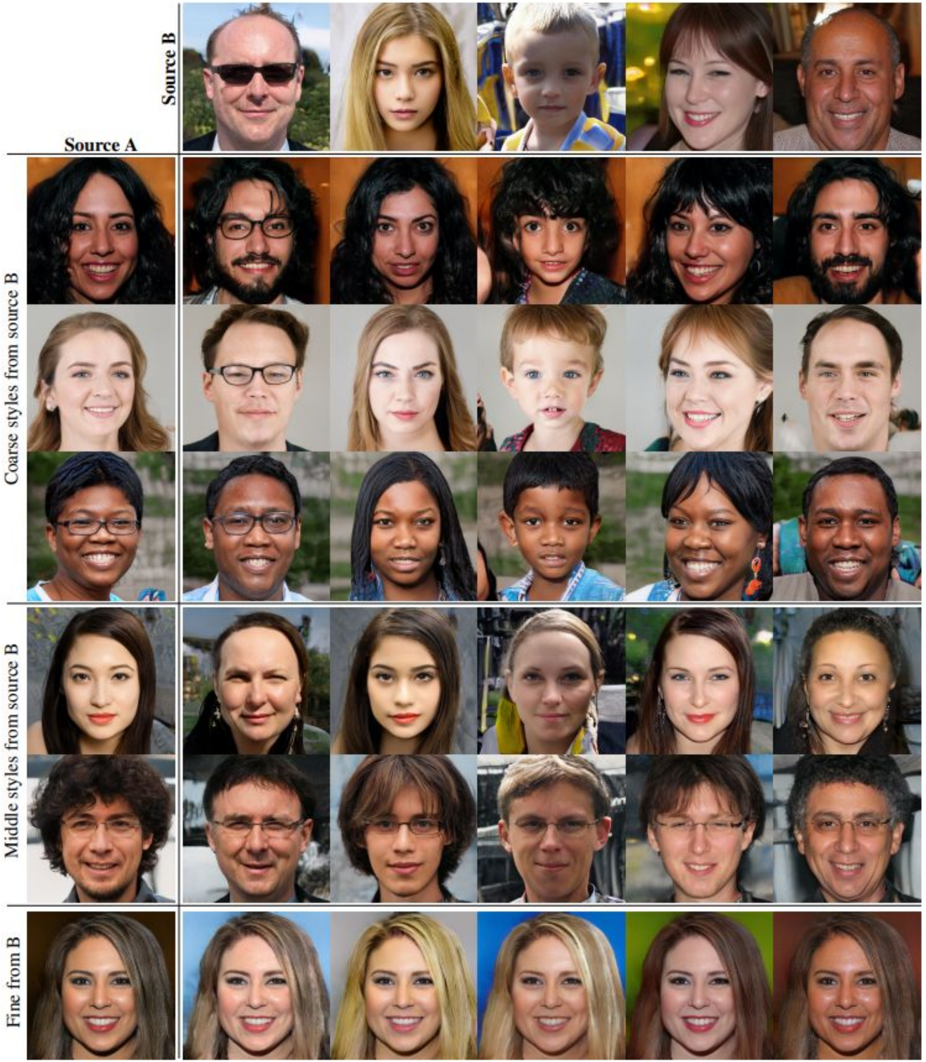
Noise matters

a) Noise is applied to all layers
b) No noise
c) Noise in fine layers only
d) Noise in coarse layers only



Style mixing

1. Generate 2 latent vectors $\mathbf{z_1}$, $\mathbf{z_2}$
2. Map them: $\mathbf{w_1 = f(z_1)}$, $\mathbf{w_2 = f(z_2)}$
3. Pass $\mathbf{w_1}$ to first **k** "**A**" blocks and $\mathbf{w_2}$ to the rest to generate new image


Source B
Source A
Coarse styles from source B
Middle styles from source B
Fine from B

Ablations compared using FID

	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**

Perceptual Path Length (PPL)

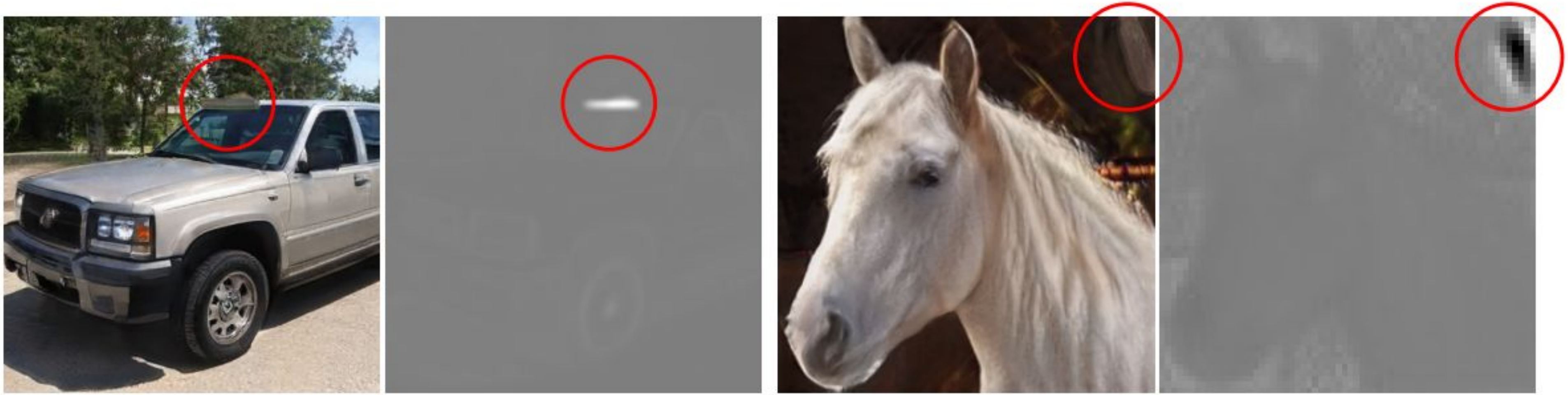
- A measure of disentanglement

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

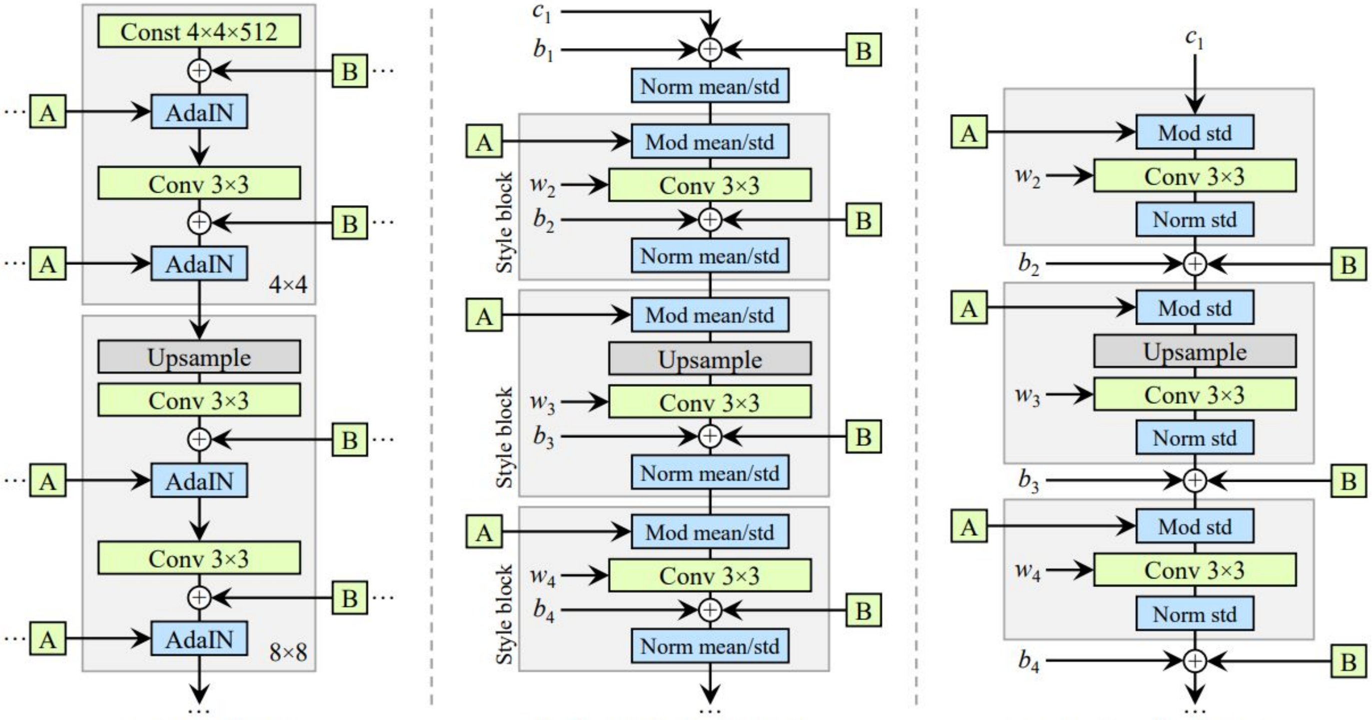
Mapping network impact

	Method		FID	Path length full	Path length end	Separability
B	Traditional 0	$\mathcal{Z}$	5.25	412.0	415.3	10.78
	Traditional 8	$\mathcal{Z}$	4.87	896.2	902.0	170.29
	Traditional 8	$\mathcal{W}$	4.87	324.5	212.2	6.52
	Style-based 0	$\mathcal{Z}$	5.06	283.5	285.5	9.88
	Style-based 1	$\mathcal{W}$	4.60	219.9	209.4	6.81
	Style-based 2	$\mathcal{W}$	4.43	**217.8**	199.9	6.25
F	Style-based 8	$\mathcal{W}$	**4.40**	234.0	**195.9**	**3.79**

StyleGAN-1: Blob-like artifacts



StyleGAN-2: Revising architecture (INFRA)

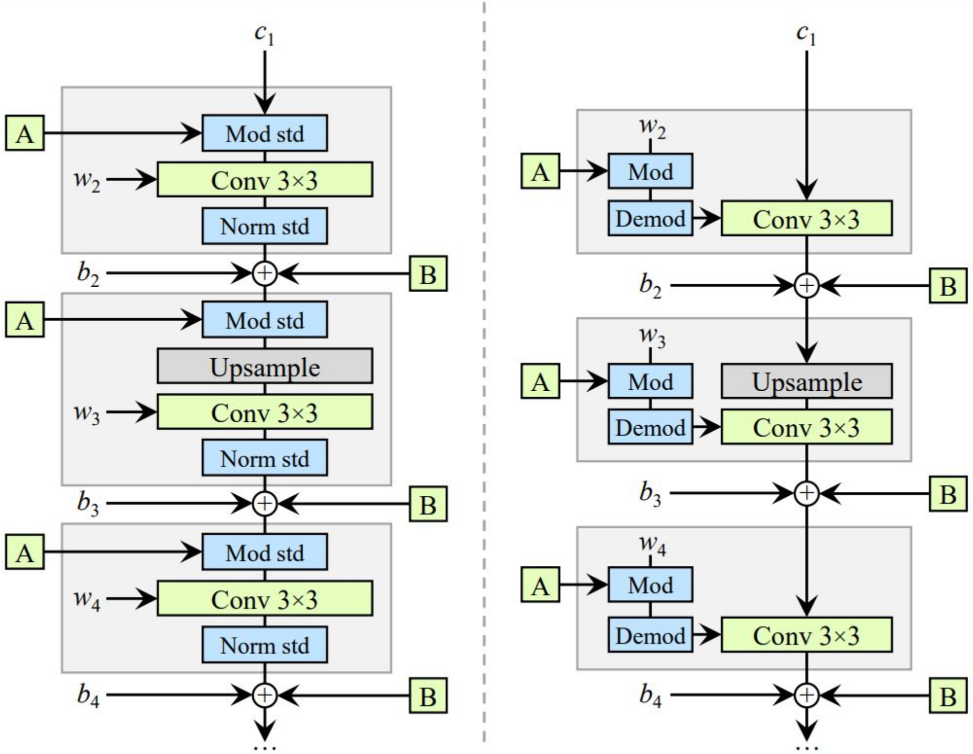


(a) StyleGAN (b) StyleGAN (detailed) (c) Revised architecture

Weight demodulation

$$w'_{ijk} = s_i \cdot w_{ijk}$$

$$w''_{ijk} = w'_{ijk} \Big/ \sqrt{\sum_{i,k} {w'_{ijk}}^2 + \epsilon}$$



(c) Revised architecture

(d) Weight demodulation

PPL as image quality metric

(a) Low PPL scores

(b) High PPL scores

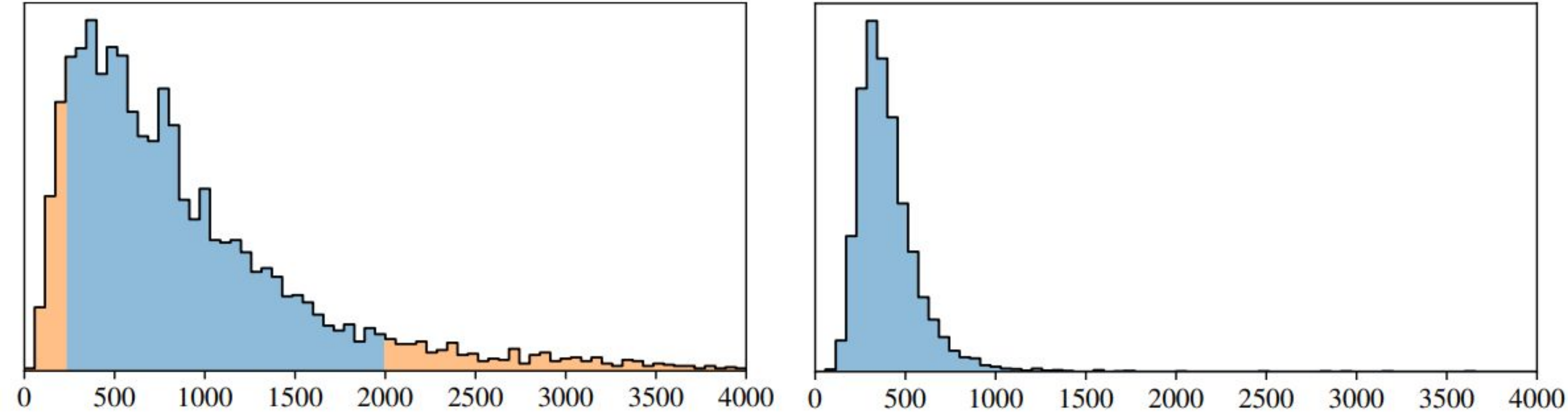
PPL Regularizer

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

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

- Optimal $\mathbf{J}_{\mathbf{w}}$ is orthogonal up to a global scale
- **a** is an exponential moving average of $\|\mathbf{J}_{\mathbf{w}}^{T}\mathbf{y}\|_2$

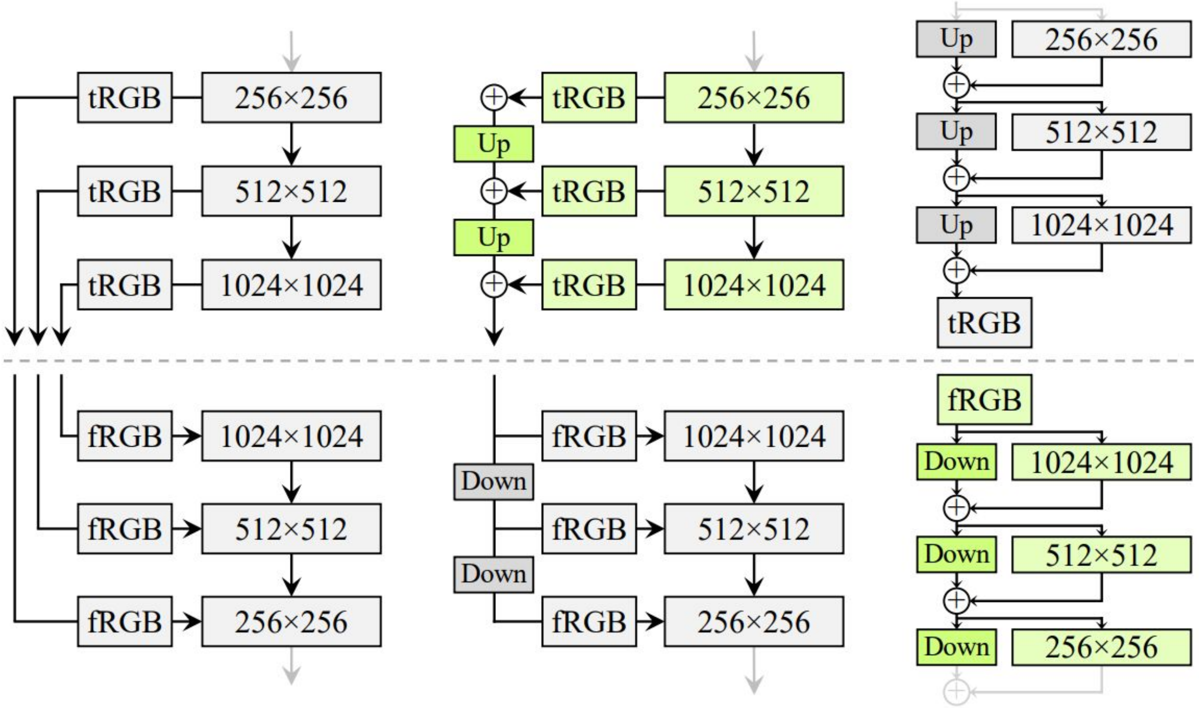
PPL distributions without/with PPL Regularizer



(a) StyleGAN (config A) (b) StyleGAN2 (config F)

Yet another StyleGAN-1 artifact

StyleGAN-2: Removing progressive growing



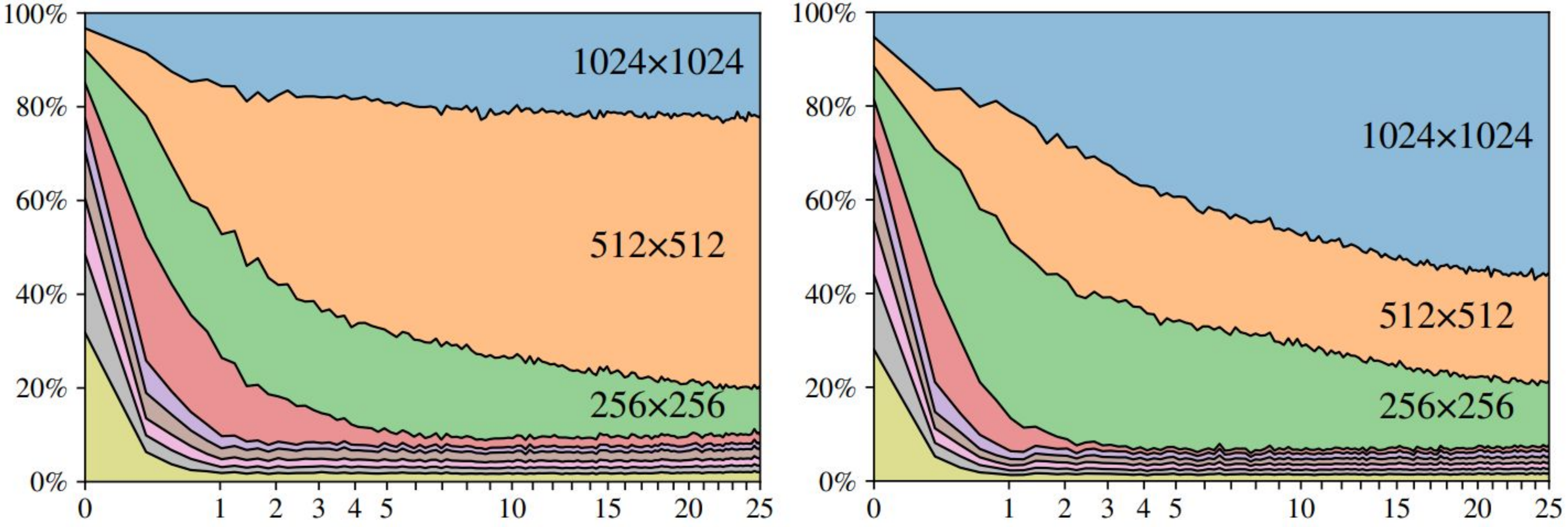
(a) MSG-GAN (b) Input/output skips (c) Residual nets

Experiments with alternative architectures

FFHQ	D original		D input skips		D residual	
	FID	PPL	FID	PPL	FID	PPL
G original	4.32	265	4.18	235	3.58	269
G output skips	4.33	169	3.77	127	**3.31**	**125**
G residual	4.35	203	3.96	229	3.79	243

LSUN Car	D original		D input skips		D residual	
	FID	PPL	FID	PPL	FID	PPL
G original	3.75	905	3.23	758	3.25	802
G output skips	3.77	544	3.86	**316**	3.19	471
G residual	3.93	981	3.40	667	**2.66**	645

Contribution of resolution and capacity problem



(a) StyleGAN-sized (config E)

(b) Large networks (config F)

StyleGAN-2: Final results

Configuration	FFHQ, 1024×1024				LSUN Car, 512×384			
	FID ↓	Path length ↓	Precision ↑	Recall ↑	FID ↓	Path length ↓	Precision ↑	Recall ↑
A Baseline StyleGAN [24]	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: ~~Futexes~~ GANs are tricky

- Progressive GAN introduced progressive growing, which enabled generation of high resolution images
- StyleGAN-1 made it possible to control image synthesis, namely combining styles via image mixing
- StyleGAN-2 removed some of the StyleGAN-1 image artifacts and modernized the architectures of the generator and discriminator