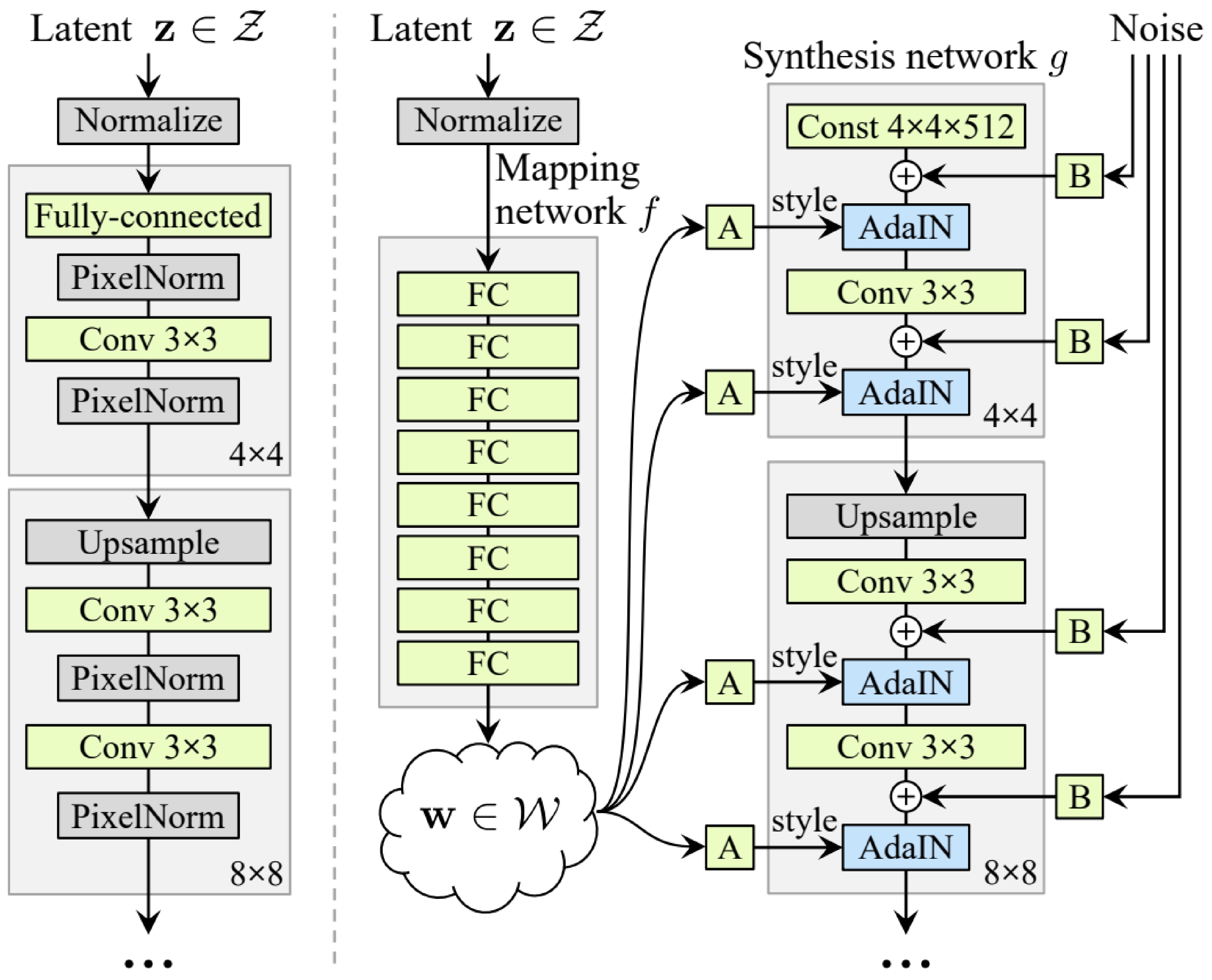
StyleGAN2

ANALYZING AND IMPROVING THE IMAGE QUALITY OF STYLEGAN

Мария Поклонская

АРХИТЕКТУРА **STYLEGAN**



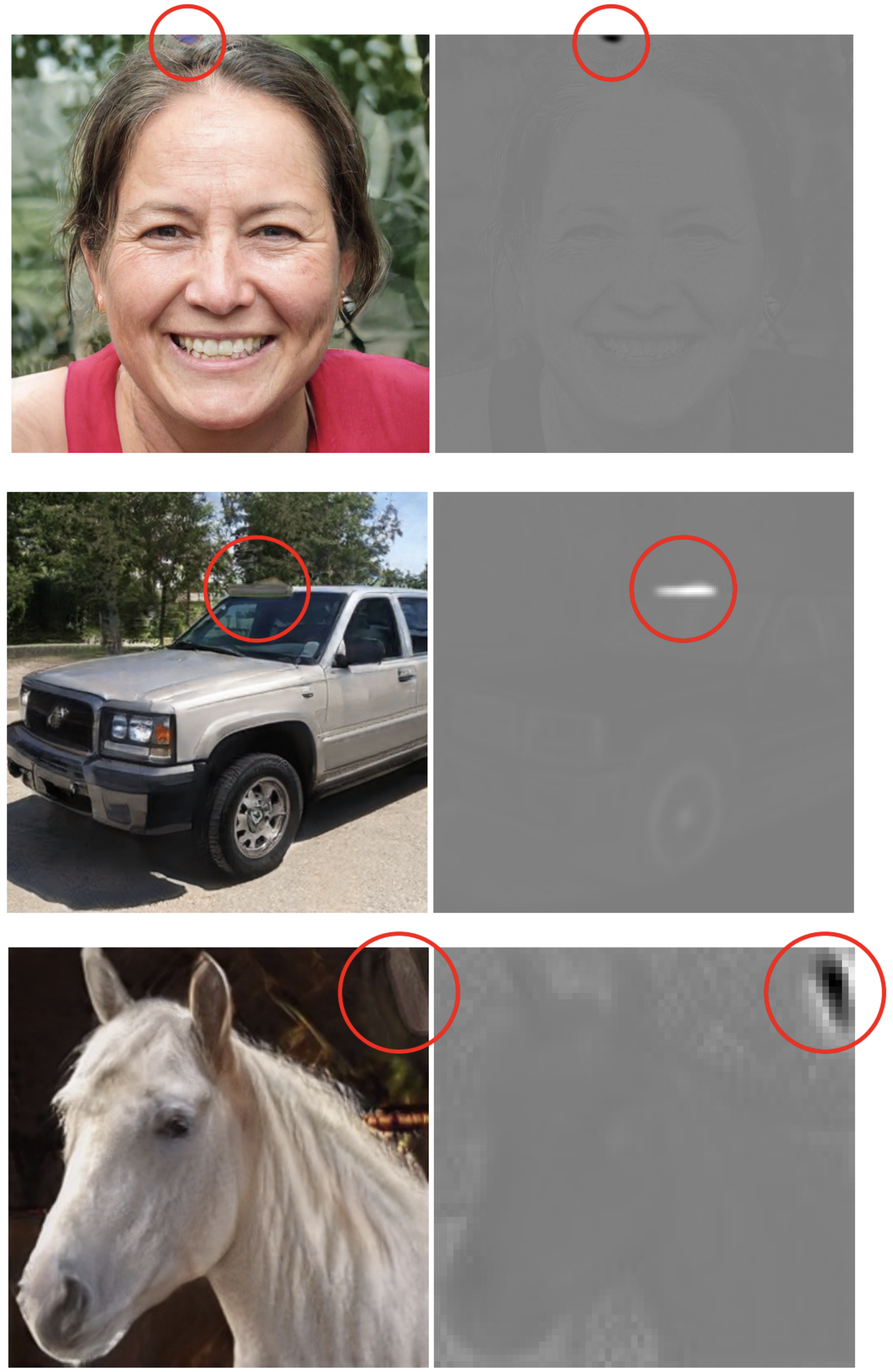
(a) Traditional (b) Style-based generator

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.

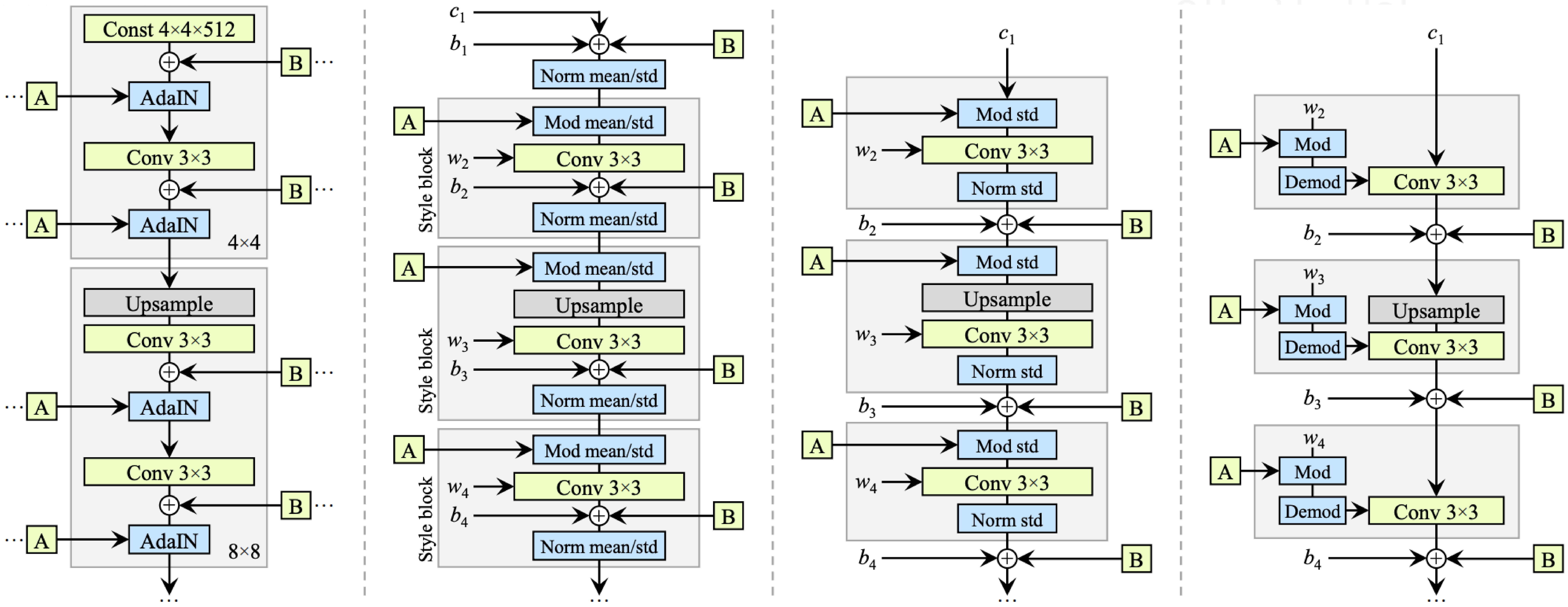
МОТИВАЦИЯ **STYLEGAN 2**

PHASE ARTIFACTS



DROPLET ARTIFACTS

Уходим от AdaIN



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

ДЕМОДУЛЯЦИЯ ВЕСОВ

БЫЛО

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

СТАЛО

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

дисперсия на выходе из сверточного слоя: $\sigma_j = \sqrt{\sum_{i,k} {w'_{ijk}}^2}$

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

w_{ijk} – веса сверточной сети

s_i – масштаб, соответствующий iому input feature map

j – индекс output feature map

k – индекс spatial footprint of the convolution

Исчезли 'DROPLET' ARTIFACTS

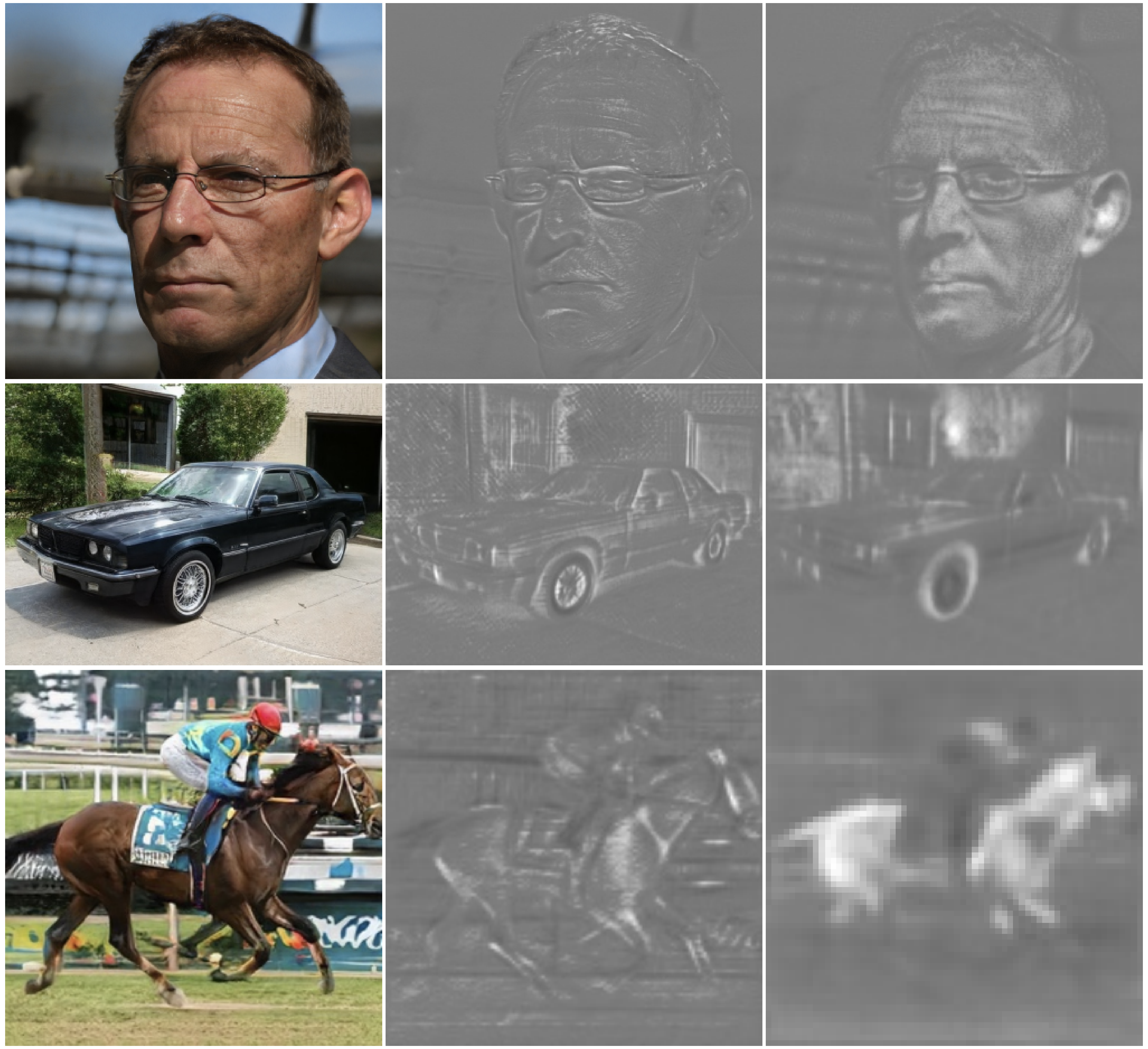


Figure 3. Replacing normalization with demodulation removes the characteristic artifacts from images and activations.

МЕТРИКИ

БЫЛО

FRECHET INCEPTION DISTANCE (FID)

Model 1: FID = 3.27, P = 0.70, R = 0.44, PPL = 1485

Model 2: FID = 3.27, P = 0.67, R = 0.48, PPL = 437

МЕТРИКИ

СТАЛО

FRECHET INCEPTION DISTANCE (FID) + **PERCEPTUAL PATH LENGTH (PPL)**

(a) Low PPL scores

(b) High PPL scores

PATH LENGTH REGULARIZATION

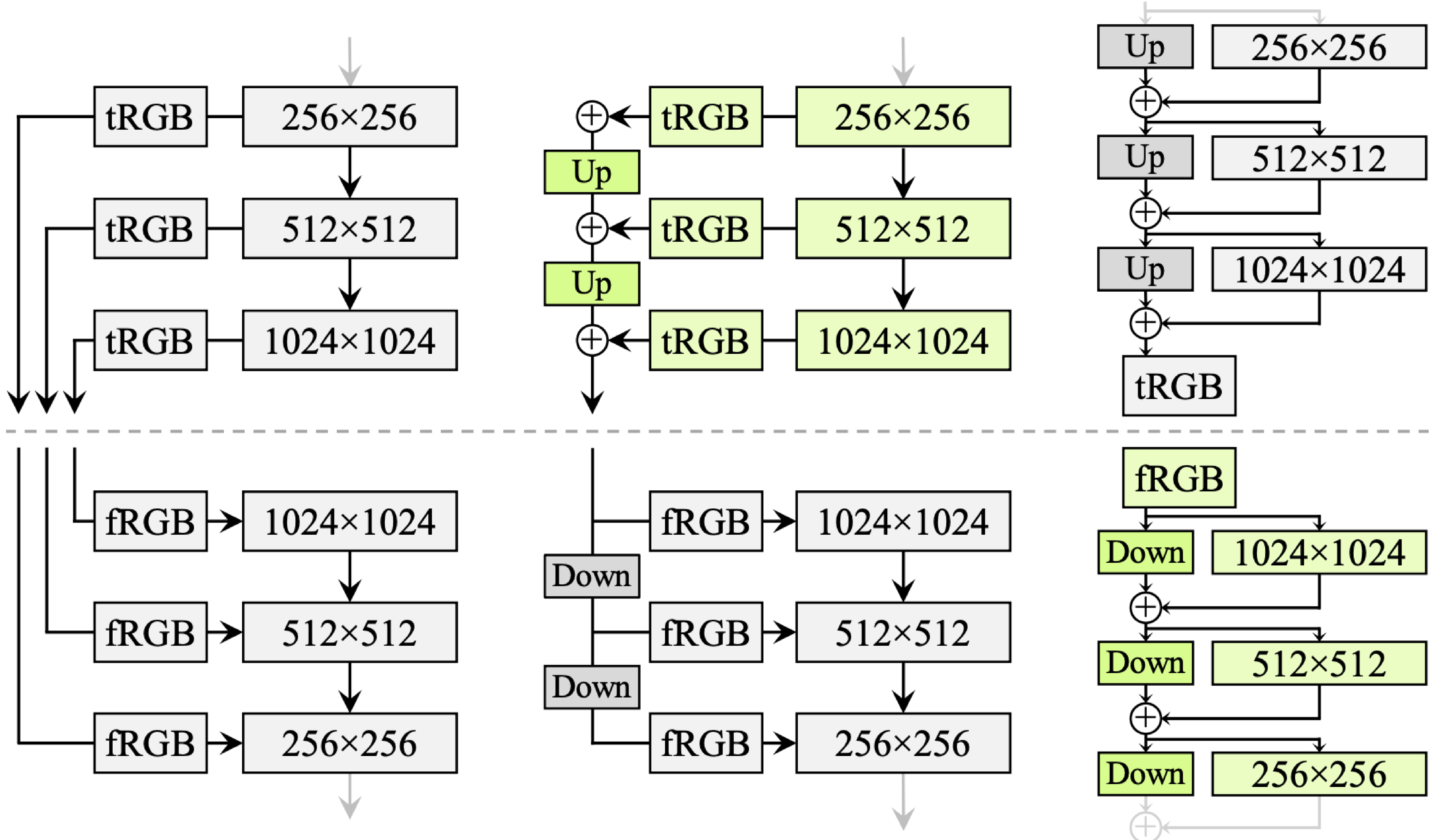
$$\mathbb{E}_{\mathbf{w},\mathbf{y}\sim\mathcal{N}(0,\mathbf{I})}\left(\left\|\mathbf{J}_{\mathbf{w}}^{T}\mathbf{y}\right\|_{2}-a\right)^{2}$$

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

+ LAZY REULARIZATION

Делаем **раз в 16 шагов** для уменьшения вычислительной стоимости и объема использкемой памяти

УХОДИМ ОТ PROGRESSIVE GROWING



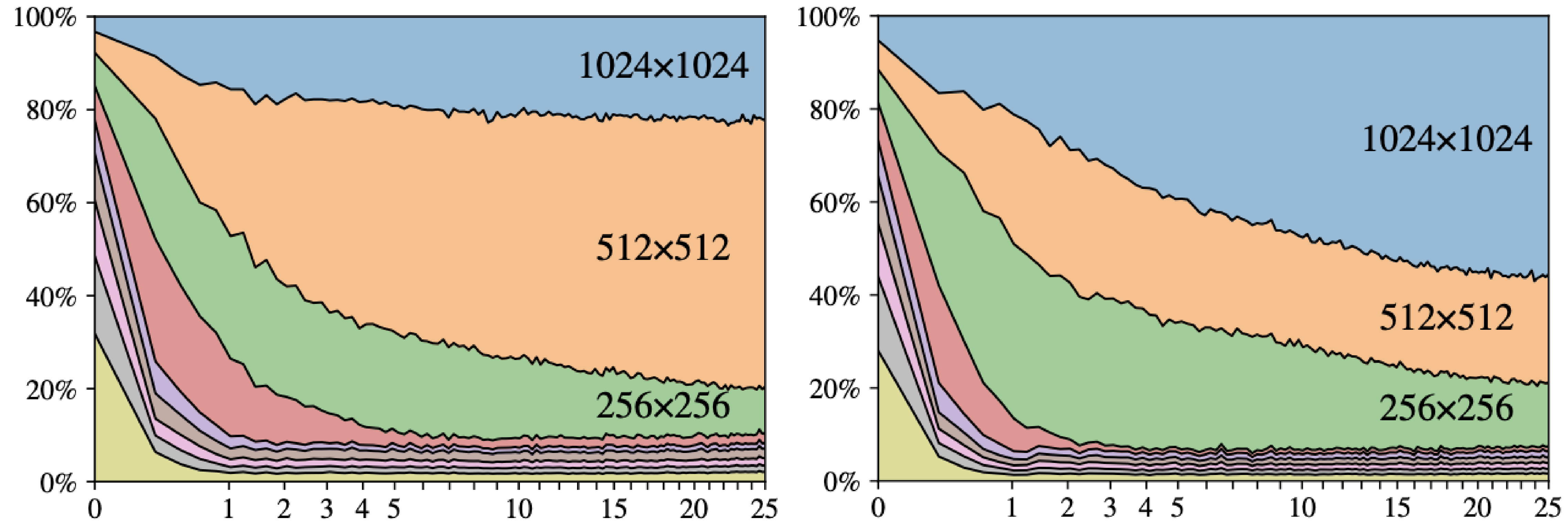
(a) MSG-GAN (b) Input/output skips (c) Residual nets

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

Table 2. Comparison of generator and discriminator architectures without progressive growing. The combination of generator with output skips and residual discriminator corresponds to configuration E in the main result table.

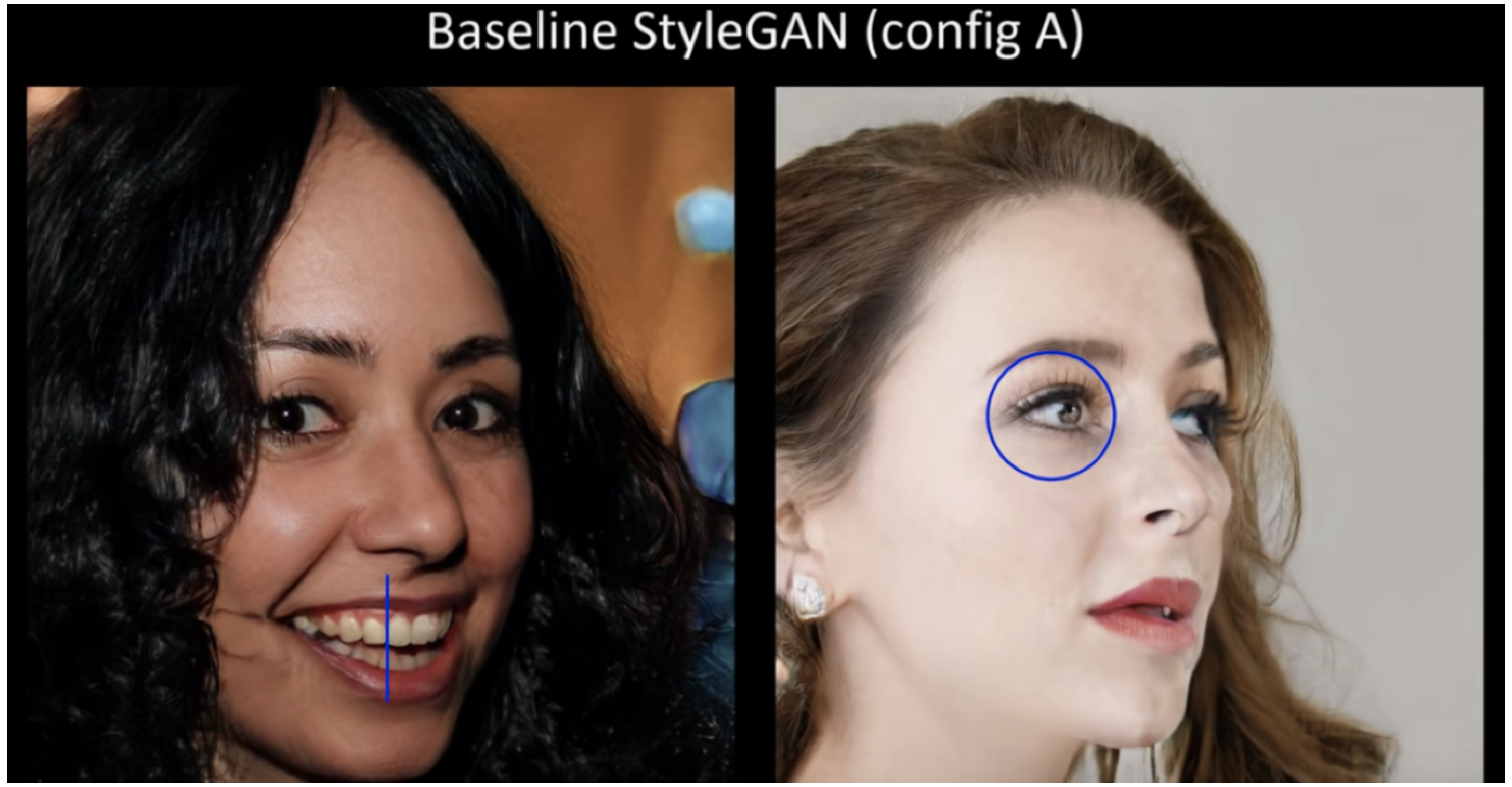
УХОДИМ ОТ **PROGRESSIVE GROWING**

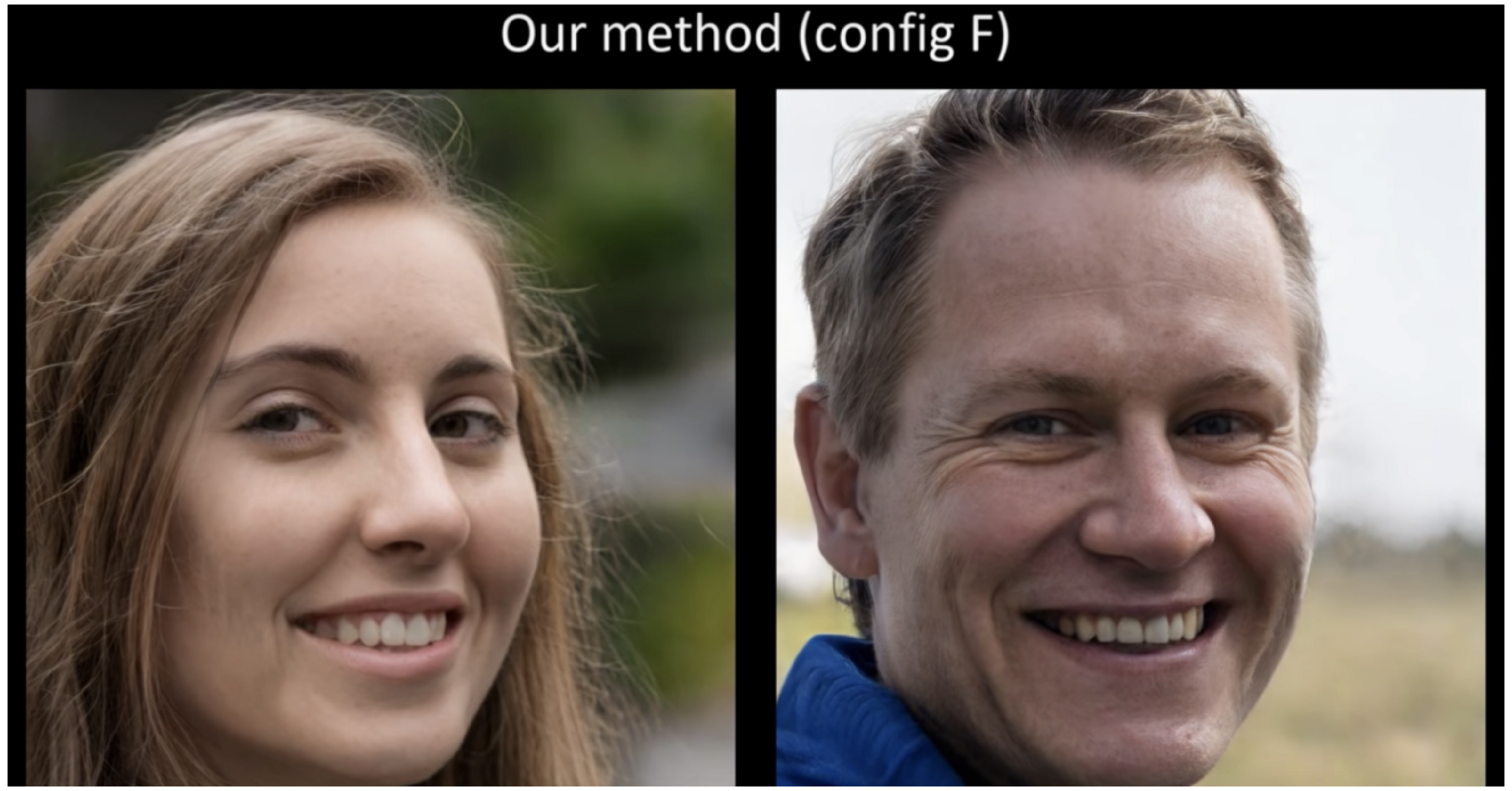


(a) StyleGAN-sized (config E) (b) Large networks (config F)

Figure 8. Contribution of each resolution to the output of the generator as a function of training time. The vertical axis shows a breakdown of the relative standard deviations of different resolutions, and the horizontal axis corresponds to training progress, measured in millions of training images shown to the discriminator. We can see that in the beginning the network focuses on low-resolution images and progressively shifts its focus on larger resolutions as training progresses. In (a) the generator basically outputs a 512^2 image with some minor sharpening for 1024^2, while in (b) the larger network focuses more on the high-resolution details.

Исчезли 'PHASE' ARTIFACTS





ПРОЕЦИРУЕМ ИЗОБРАЖЕНИЯ В ЛАТЕНТНОЕ ПРОСТРАНСТВО

StyleGAN — generated images StyleGAN2 — generated images StyleGAN2 — real images

Figure 9. Example images and their projected and re-synthesized counterparts. For each configuration, top row shows the target images and bottom row shows the synthesis of the corresponding projected latent vector and noise inputs. With the baseline StyleGAN, projection often finds a reasonably close match for generated images, but especially the backgrounds differ from the originals. The images generated using StyleGAN2 can be projected almost perfectly back into generator inputs, while projected real images (from the training set) show clear differences to the originals, as expected. All tests were done using the same projection method and hyperparameters.

ИТОГИ

Four hand-picked examples illustrating the image quality and diversity achievable using StylegGAN2 (config F).

	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	–	–	–	–