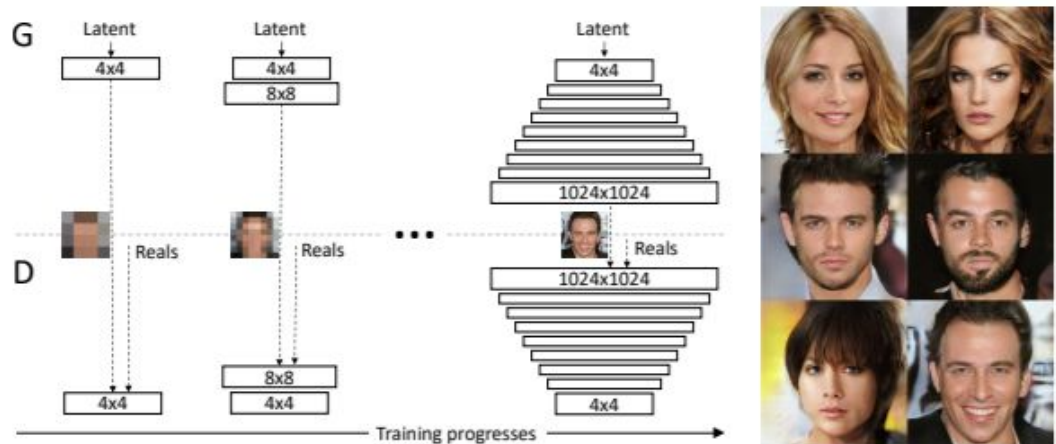
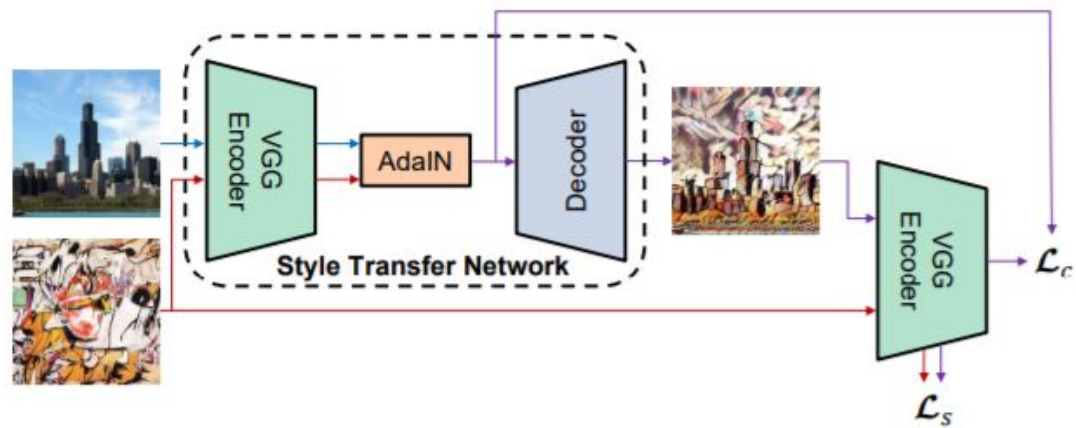


A Style-Based Generator Architecture for Generative Adversarial Networks (StyleGAN)

Щербинин Артём, 181

Основные архитектуры для переноски стиля



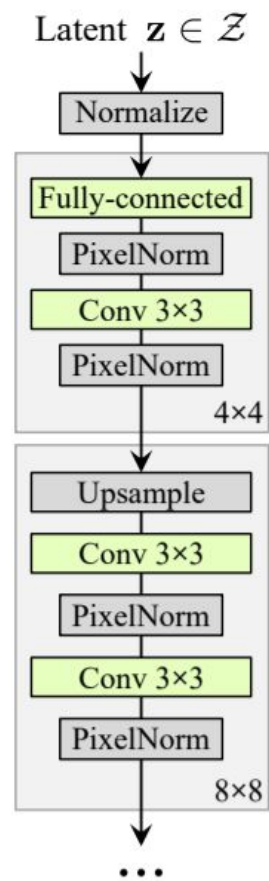
Набор данных FFHQ

- 1) 70000 картинок размера 1024x1024
- 2) Больше вариативности различных факторов



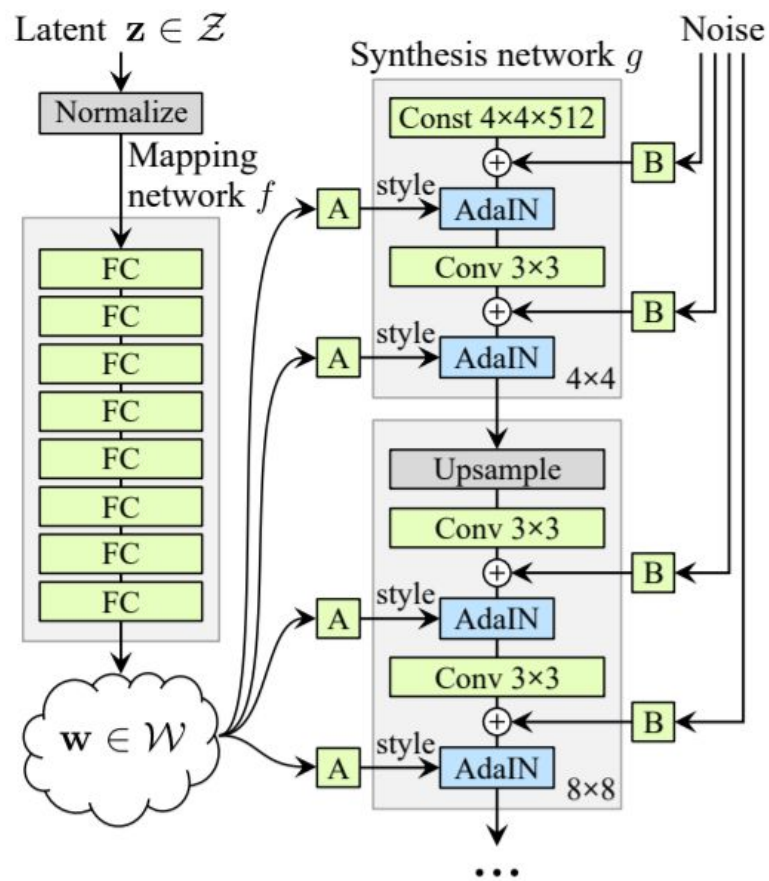
Основные изменения:

- 1) Новая нормализация - adaptive instance normalization (AdaIN)
- 2) Разделение на 2 подсети - Mapping network (f) и Synthesis network (g)
- 3) Добавление шума



(a) Traditional

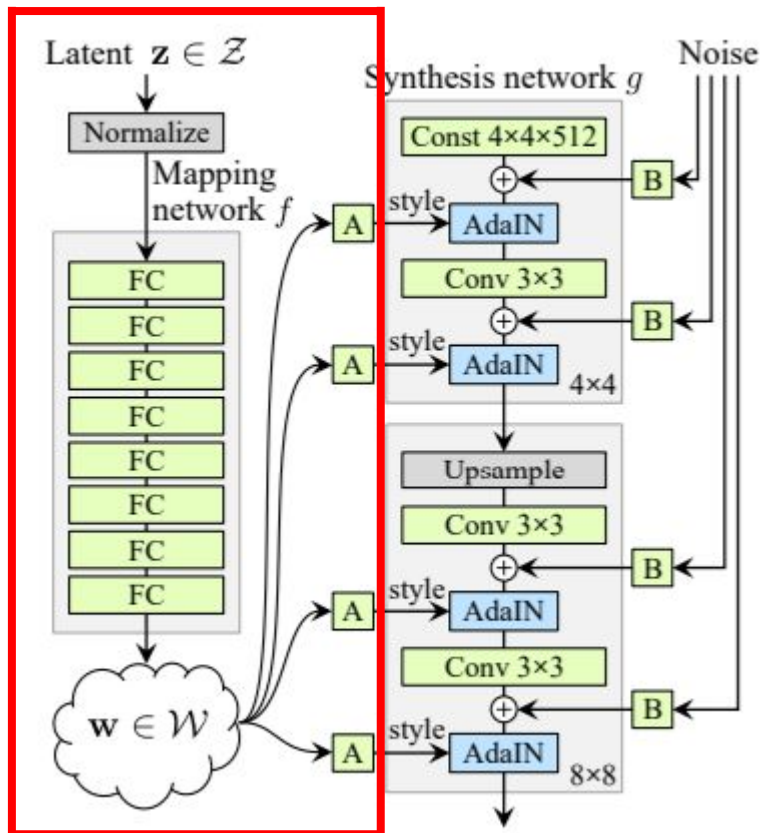
23.1M параметров



(b) Style-based generator

26.2M параметров

Преобразование входного слоя

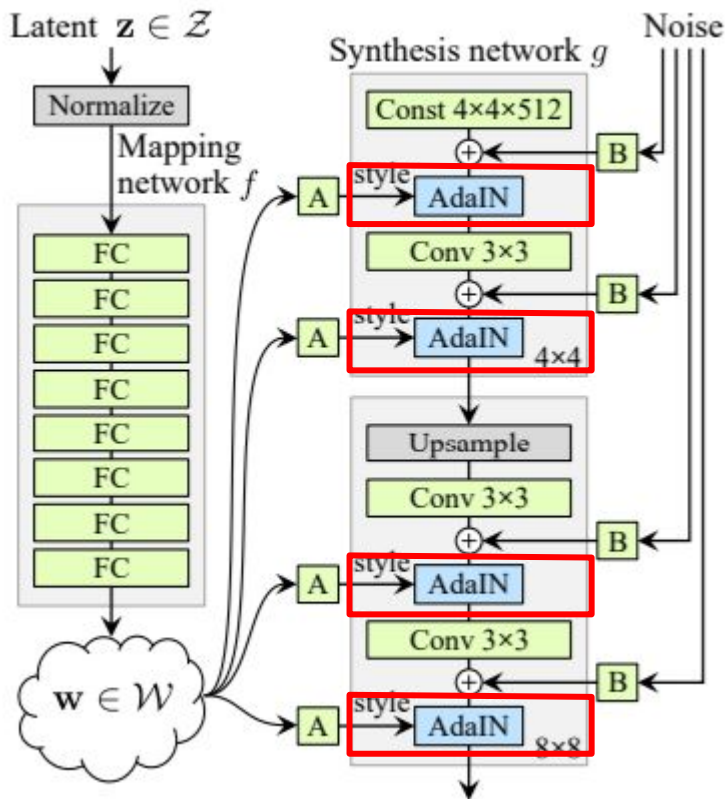


$$w \sim y = (y_s, y_b)$$

Особенности:

- 1) Вся сеть, а также A - тренируемые объекты
- 2) y размера $2 \times K(512)$, где K - количество каналов
- 3) Отвечает за "стиль" нашего изображения
- 4) Можно разделить преобразование на 2 части - W служит для распутывания скрытой информации, а A для сужения.

AdaIN



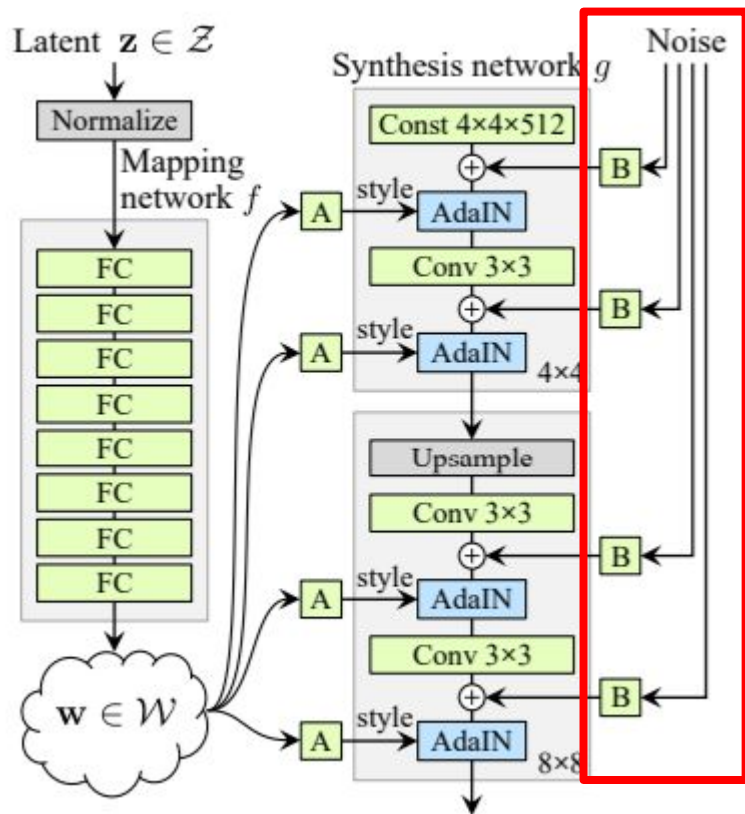
$$\text{AdaIN}(\mathbf{x}_i, \mathbf{y}) = \mathbf{y}_{s,i} \frac{\mathbf{x}_i - \mu(\mathbf{x}_i)}{\sigma(\mathbf{x}_i)} + \mathbf{y}_{b,i}$$

$\mu(x_i)$ — математическое ожидание i -ого канала

$\sigma(x_i)$ — стандартное отклонение i -ого канала

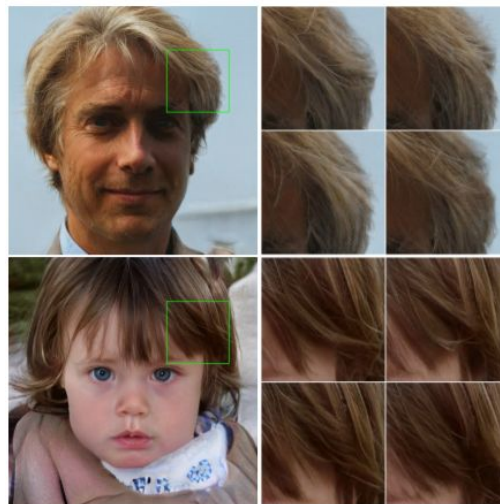
$y_{s,i}/y_{b,i}$ — соответствующие компоненты вектора стиля \mathbf{y}

Шум



Особенности:

- 1) $N(18)$ - одноканальных картинок, состоящие из некоррелированного нормального шума
- 2) B - обучаемое скалирование для каждого канала нашего изображения
- 3) Отвечает за случайность мелких деталей изображений



(a) Generated image (b) Stochastic variation

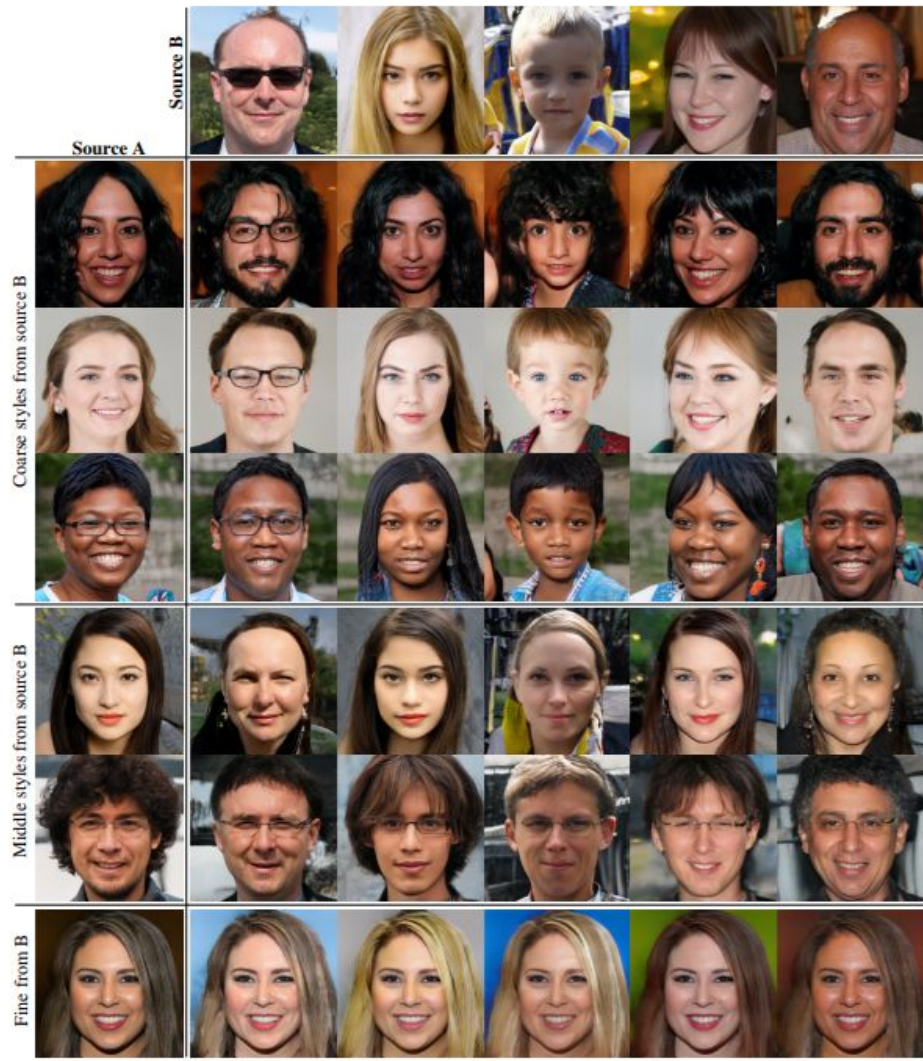


(c) (d)

Style mixing

Особенности:

- 1) в зависимости от количества раз и последовательности применения разных стилей меняется кардинально результат



Метрика Perceptual path length

$$l_W = \mathbb{E} \left[\frac{1}{\epsilon^2} d(g(\text{lerp}(f(\mathbf{z}_1), f(\mathbf{z}_2); t)), g(\text{lerp}(f(\mathbf{z}_1), f(\mathbf{z}_2); t + \epsilon))) \right]$$

$f()$ — функция преобразования

t, ϵ — коэффициенты

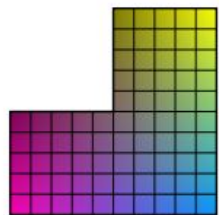
$\text{lerp}()$ — линейная интерполяция

$g()$ — функция синтеза

$d(., .)$ — Евклидово расстояние между пикселями

Качества и интересные моменты

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



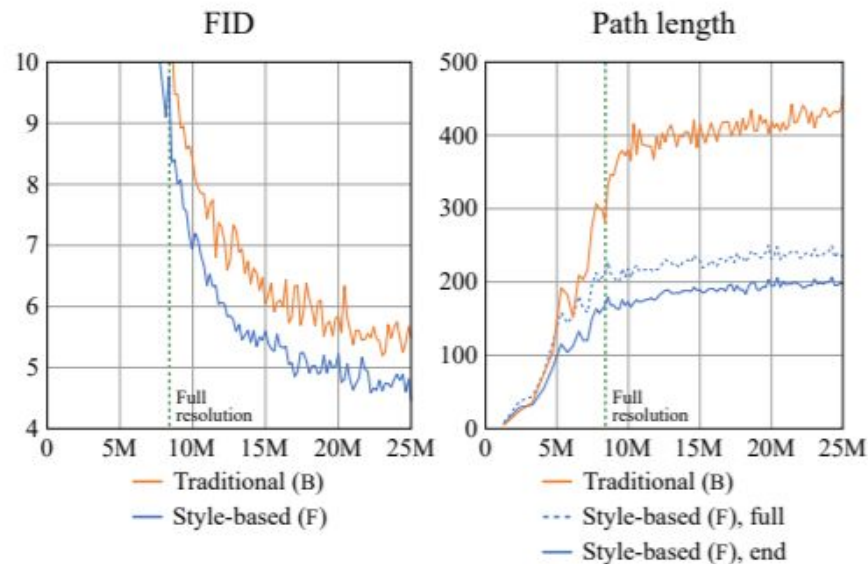
(a) Distribution of features in training set



(b) Mapping from \mathcal{Z} to features

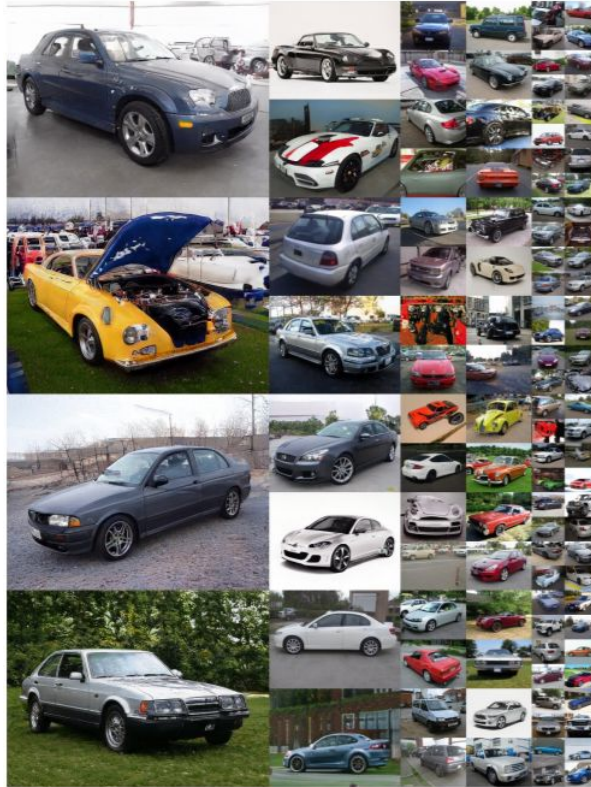


(c) Mapping from \mathcal{W} to features



- 1) Обучалось неделю на NVIDIA DGX-1 с 8 Tesla V100 GPUs.
- 2) Оптимизатор Adam
- 3) Функция потерь non-saturating loss с R1 регуляризаций
- 4) Данные CelebA-HQ и их данные FFHQ

Результаты на других данных



Проблемы StyleGAN



- 1) Артефакты / водяные знаки
- 2) Качество изображений
- 3) Резкость мягкость изображений

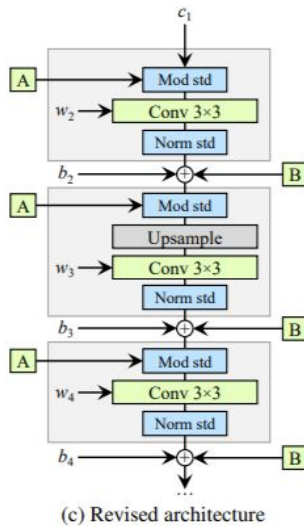
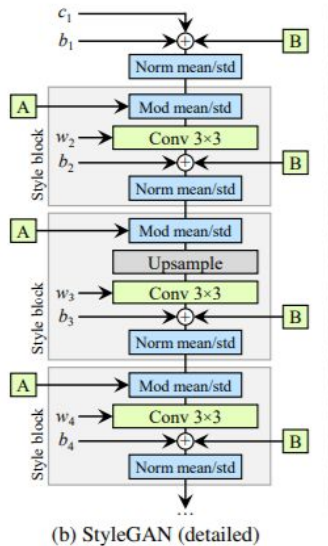
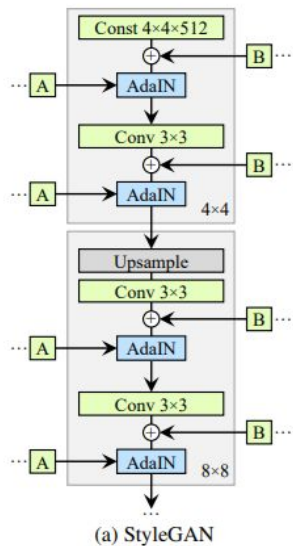
В чём заключались проблемы:

- 1) Артефакты появлялись из-за нормализации
- 2) Также появлялись артефакты из-за увеличения размера модели
- 3) Не было регуляризации для лучшего качества картинки

Способы решения проблем:

- 1) Изменение функции нормализации
- 2) Изменение основного блока архитектуры
- 3) Добавить регуляризации

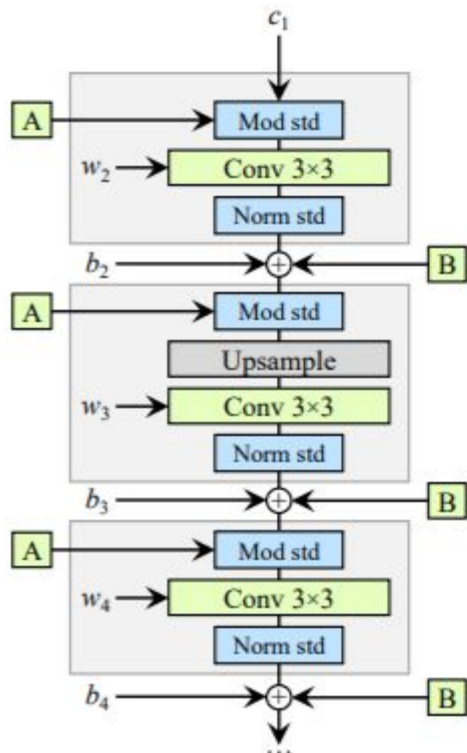
Изменение функции нормализации и блока



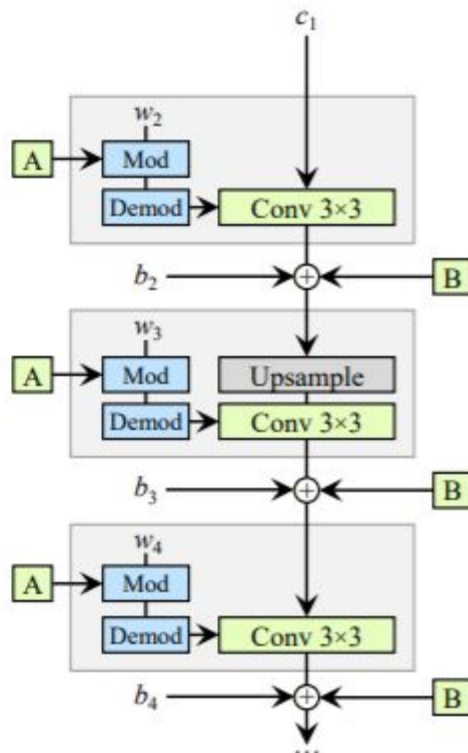
Особенности:

- 1) Пересобрали блок стиля
- 2) Убрали мат. ожидания (так как они не влияли на результат)

Изменение функции нормализации и блока



(c) Revised architecture



(d) Weight demodulation

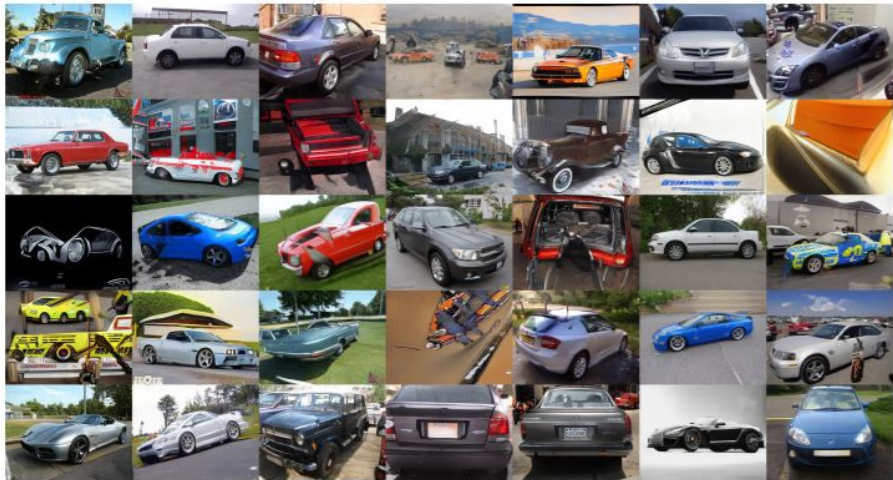
$$w'_{ijk} = s_i \cdot w_{ijk}$$

$$\sigma_j = \sqrt{\sum_{i,k} w'_{ijk}{}^2}$$

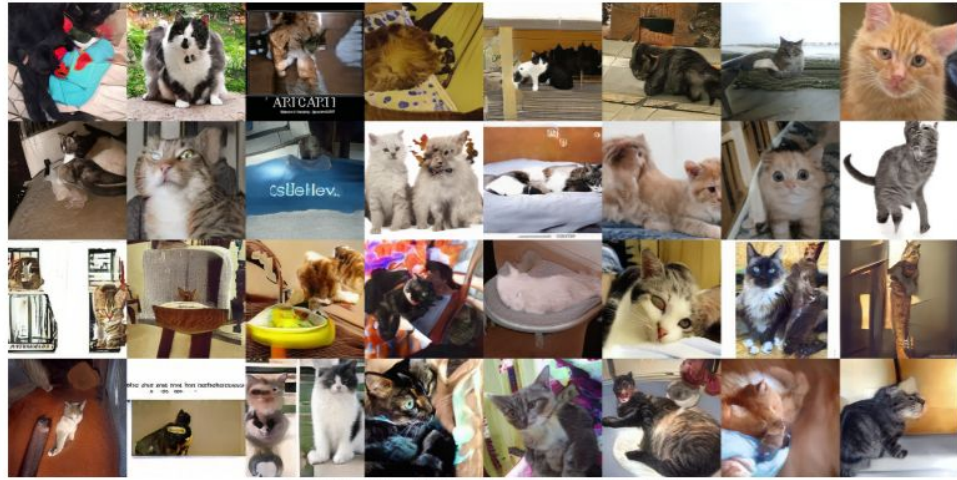
$$w''_{ijk} = w'_{ijk} / \sqrt{\sum_{i,k} w'_{ijk}{}^2 + \epsilon}$$

Особенности:

- 1) Меньший разброс от значений в векторе стиля



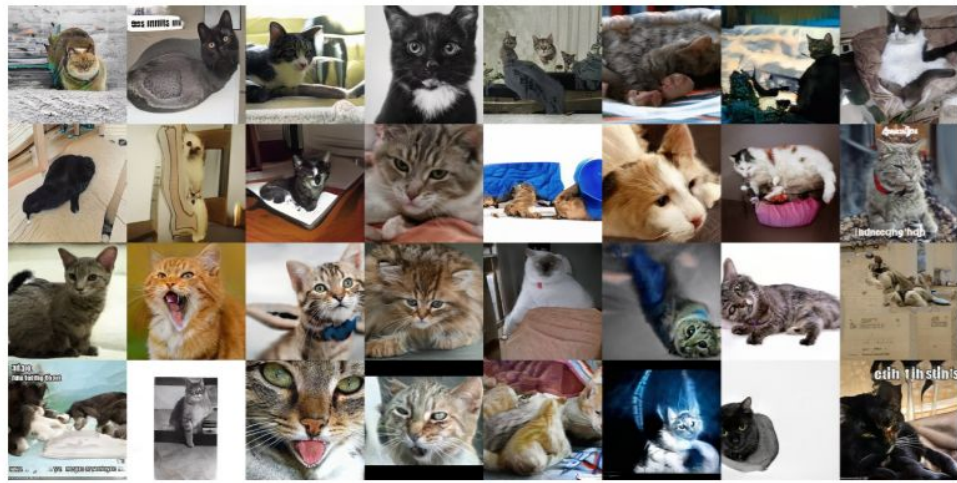
Model 1: FID = 3.27, P = 0.70, R = 0.44, PPL = 1485



Model 1: FID = 8.53, P = 0.64, R = 0.28, PPL = 924



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



Model 2: FID = 8.53, P = 0.62, R = 0.29, PPL = 387

Lazy regularization и Path length regularization

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

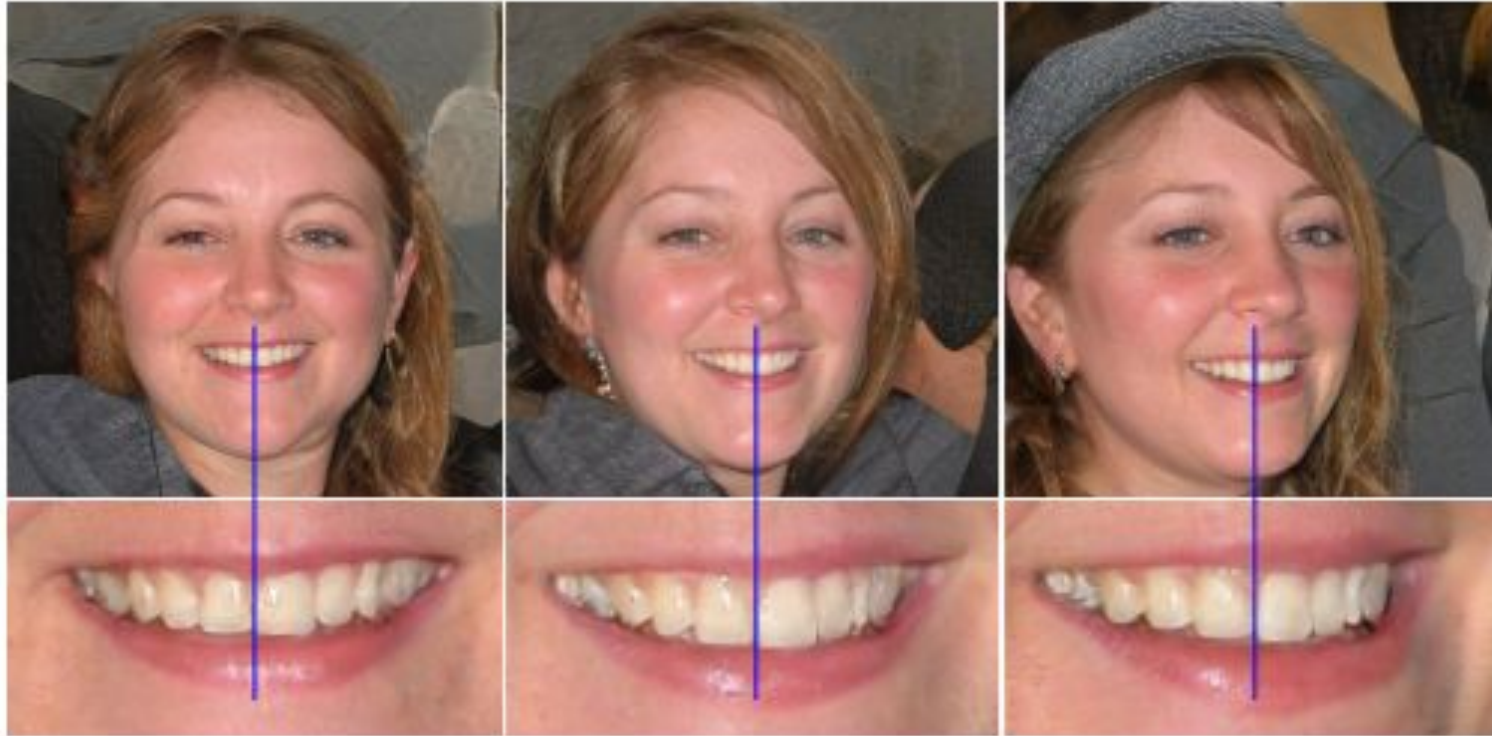
J_w — Якобиан: $\partial g(w) / \partial w$

y — случайная картинка из нормального распределения

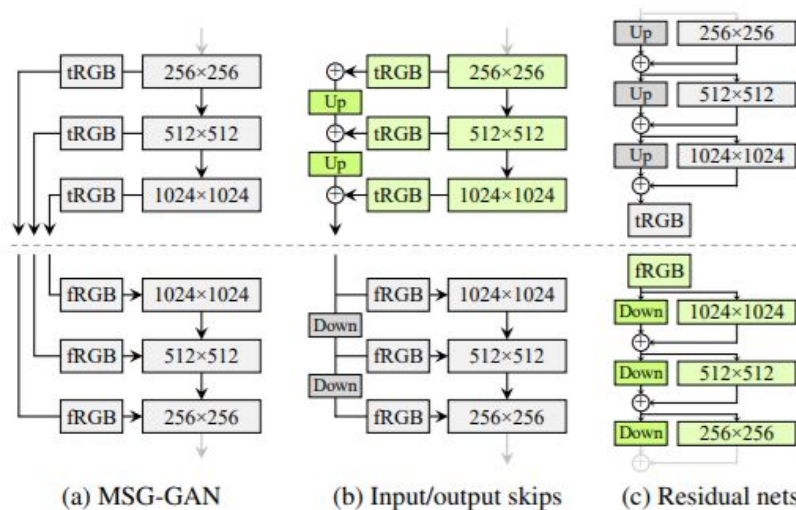
$w \sim f(z)$, где z из нормального распределения

a — скользящее среднее от $\|J_w^T\|_2$

Нерешаемая проблема StyleGAN



Другие модели и модификации



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

Метрики качества

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	–	–	–	–

Dataset	Resolution	StyleGAN (A)		StyleGAN2 (F)	
		FID	PPL	FID	PPL
LSUN CAR	512×384	3.27	1485	2.32	416
LSUN CAT	256×256	8.53	924	6.93	439
LSUN CHURCH	256×256	4.21	742	3.86	342
LSUN HORSE	256×256	3.83	1405	3.43	338



FFHQ



LSUN CAR



LSUN CAT



LSUN CHURCH



LSUN HORSE



СПИСОК ИСТОЧНИКОВ

1. [A Style-Based Generator Architecture for Generative Adversarial Networks \(StyleGAN\)](#)
2. [Analyzing and Improving the Image Quality of StyleGAN \(StyleGAN2\)](#)
3. [Training Generative Adversarial Networks with Limited Data \(StyleGAN-ADA\)](#)