

StyleGAN

Tim Senior

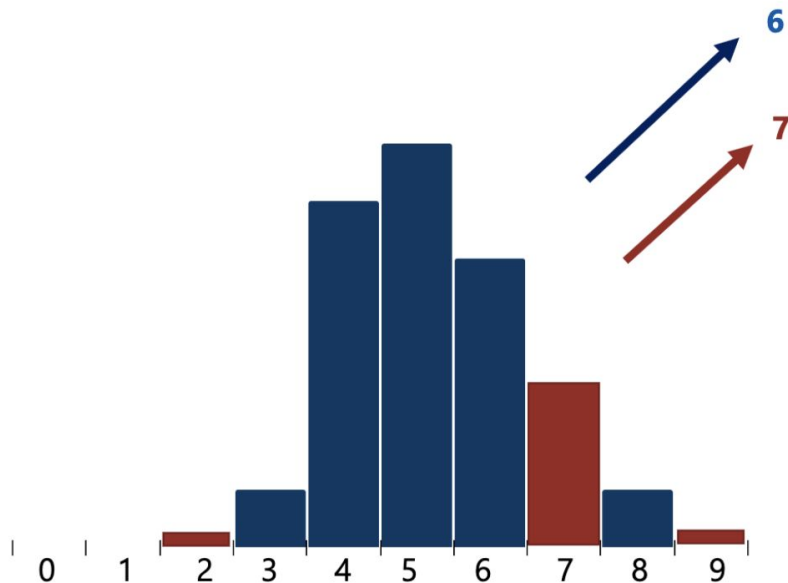
План доклада

1. Задача генеративного моделирования
2. Дивергенции
3. Идея работы GAN
4. Проблемы GAN
5. StyleGAN
6. StyleGAN2

Генеративное моделирование

Random Number Generation

- ▶ Assume there is a probability density $p_{true}(x)$.
- ▶ **Choose interpolation model.**
- ▶ Try to estimate $p_{true}(x)$ using data and obtain $p_{data}(x)$.
- ▶ Sample from $p_{data}(x)$.



KL дивергенция

$$KL(p||q_{\theta}) = \int p(x) \log \left(\frac{p(x)}{q_{\theta}(x)} \right) dx$$

- не симметрична
- не ограничена сверху

$$\theta^* = \underset{\theta}{\operatorname{argmin}} KL(p(x)||q_{\theta}(x))$$

Дивергенция Йенсона-Шеннона

$$JS(p, q) = \frac{1}{2} \left(KL(p(x) || \frac{p(x) + q_{\theta}(x)}{2}) + KL(q_{\theta}(x) || \frac{p(x) + q_{\theta}(x)}{2}) \right)$$

- симметрична
- ограничена $0 \leq JS(P, Q) \leq \ln(2)$
- можно превратить в расстояние $\sqrt{JS(p, q)}$

Generative Adversarial Networks

- Назовем модель G **генератором** и обучим ее выдавать объекты из распределения

$$z_j \sim \mathcal{N}(0, I)$$

$$x_j = G_\theta(z_j)$$

- Цель модели:

$$\min_G \left\{ \mathbb{E}_{z \sim \mathcal{N}(0, I)} [\log(1 - D(G(z)))] \right\}$$

Generative Adversarial Networks

- Введем модель **дискриминатора**, которая будет различать настоящее изображение от сгенерированного
- Цель модели:

$$\max_D \left\{ \mathbb{E}_{x \sim p_{data}} [\log D(x)] + \mathbb{E}_{z \sim \mathcal{N}(0, I)} [\log(1 - D(G(z)))] \right\}$$

Generative Adversarial Networks

- Цель дискриминатора:

$$\max_D \left\{ \mathbb{E}_{x \sim p_{data}} [\log D(x)] + \mathbb{E}_{z \sim \mathcal{N}(0, I)} [\log(1 - D(G(z)))] \right\}$$

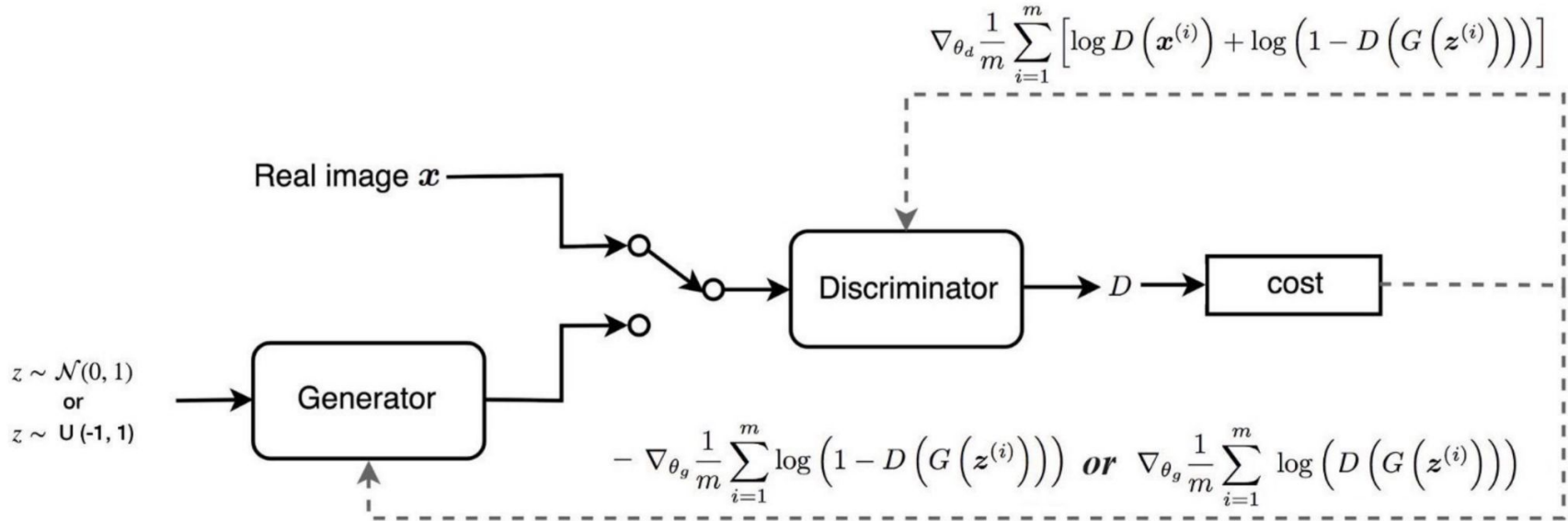
- Цель генератора:

$$\min_G \left\{ \mathbb{E}_{z \sim \mathcal{N}(0, I)} [\log(1 - D(G(z)))] \right\}$$

- Итоговая задача оптимизации:

$$\min_G \max_D \left\{ \mathbb{E}_{x \sim p_{data}} [\log D(x)] + \mathbb{E}_{z \sim \mathcal{N}(0, I)} [\log(1 - D(G(z)))] \right\} = L_{GAN} \rightarrow \min$$

Принцип работы



For D and G defined as neural networks, we can use backpropagation.

Алгоритм работы

for number of training iterations **do**

for k steps **do**

- Sample minibatch of m noise samples $\{\mathbf{z}^{(1)}, \dots, \mathbf{z}^{(m)}\}$ from noise prior $p_g(\mathbf{z})$.
- Sample minibatch of m examples $\{\mathbf{x}^{(1)}, \dots, \mathbf{x}^{(m)}\}$ from data generating distribution $p_{\text{data}}(\mathbf{x})$.
- Update the discriminator by ascending its stochastic gradient:

$$\nabla_{\theta_d} \frac{1}{m} \sum_{i=1}^m \left[\log D(\mathbf{x}^{(i)}) + \log \left(1 - D(G(\mathbf{z}^{(i)})) \right) \right].$$

end for

- Sample minibatch of m noise samples $\{\mathbf{z}^{(1)}, \dots, \mathbf{z}^{(m)}\}$ from noise prior $p_g(\mathbf{z})$.
- Update the generator by descending its stochastic gradient:

$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^m \log \left(1 - D(G(\mathbf{z}^{(i)})) \right).$$

end for

Математическое обоснование

- Для фиксированного генератора дискриминатор имеет вид

$$D_G^*(\mathbf{x}) = \frac{p_{data}(\mathbf{x})}{p_{data}(\mathbf{x}) + p_g(\mathbf{x})}$$

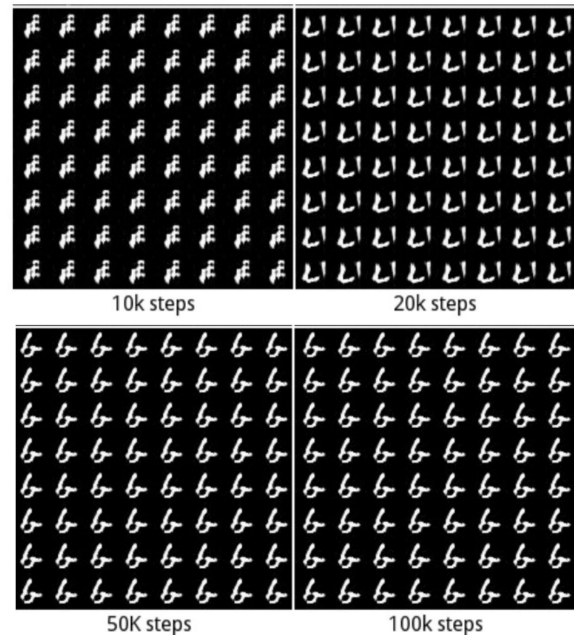
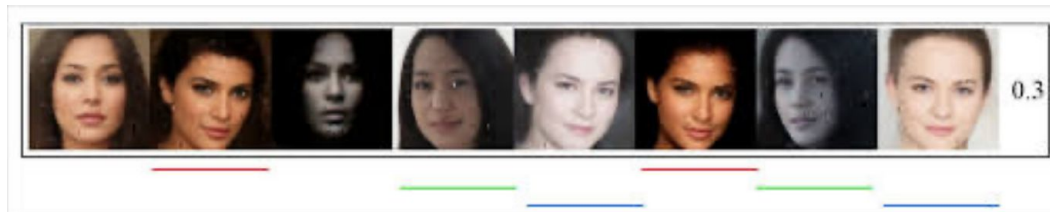
- Для оптимального дискриминатора loss принимает вид

$$L_{GAN} = 2JS(p_{data}, p_{gen}) - 2 \log 2$$

Проблемы GAN

- Mode collapse

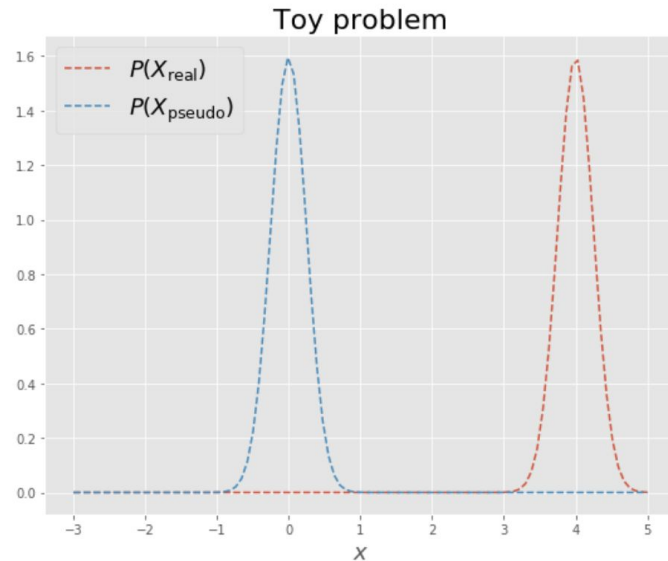
Генератор может долго генерировать объекты, пока дискриминатор не запомнит этот паттерн



Проблемы GAN

- Затухающие градиенты

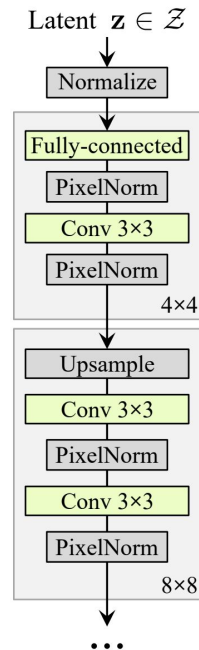
Генератор может генерировать объекты, которые дискриминатору легко отличить от настоящих, вследствие чего генератор не будет учиться



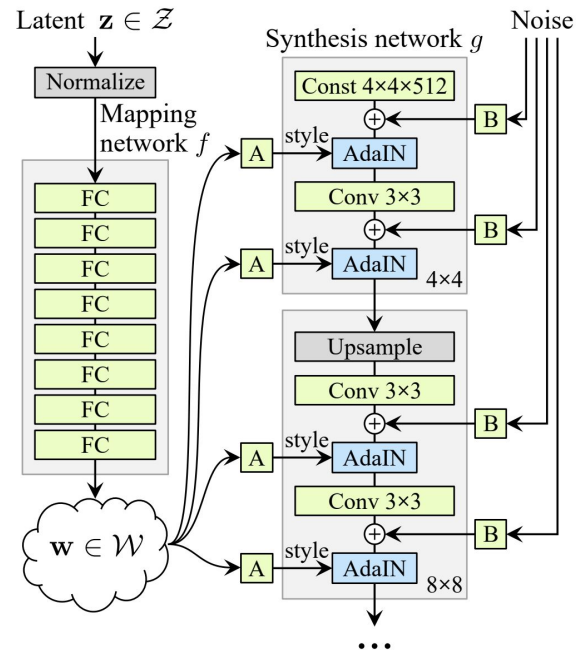
StyleGAN

- Основная идея - вектор w отвечает за стиль, передаваемый в изображения
- Шум, пропускаемый через B отвечает за вариативность получаемых объектов

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



(a) Traditional



(b) Style-based generator

FID (Fréchet inception distance)

Идея: сравнить распределения сгенерированных изображений и реальных

Алгоритм вычисления:

- Получить эмбединги изображений из внутренних слоев другой предобученной нейросети
- Предположить, что эмбединги распределены нормально
- Посчитать $FID = ||\mu_1 - \mu_2||^2 + Tr(\Sigma_1 + \Sigma_2 - 2(\Sigma_1 \Sigma_2)^{1/2})$

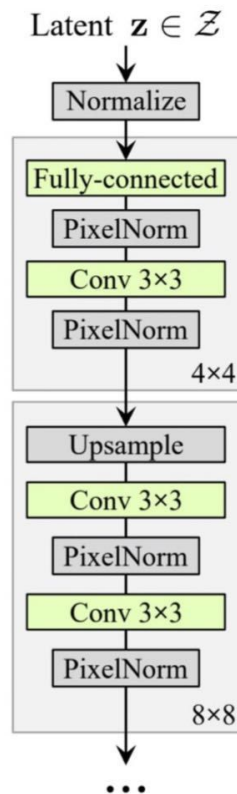
Численные результаты для StyleGAN

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

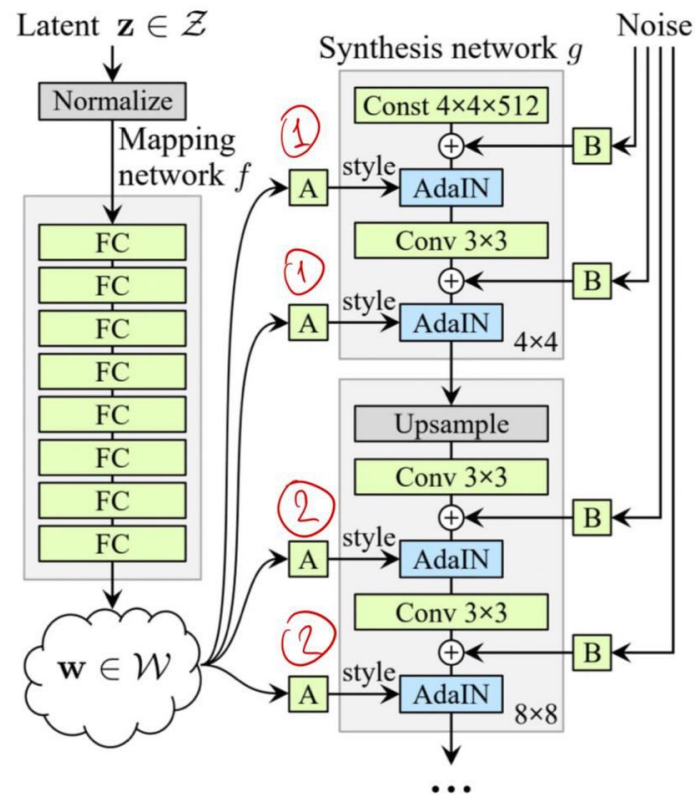
Table 1. Fréchet inception distance (FID) for various generator designs (lower is better). In this paper we calculate the FIDs using 50,000 images drawn randomly from the training set, and report the lowest distance encountered over the course of training.

Style mixing

Будем на вход синтезирующей сети с какого-то момента подавать другой латентный вектор, тогда стили с изображений будут комбинироваться



(a) Traditional



(b) Style-based generator

Style mixing

Будем на вход
синтезирующей сети с
какого-то момента
подавать другой
латентный вектор, тогда
стили с изображений будут
комбинироваться

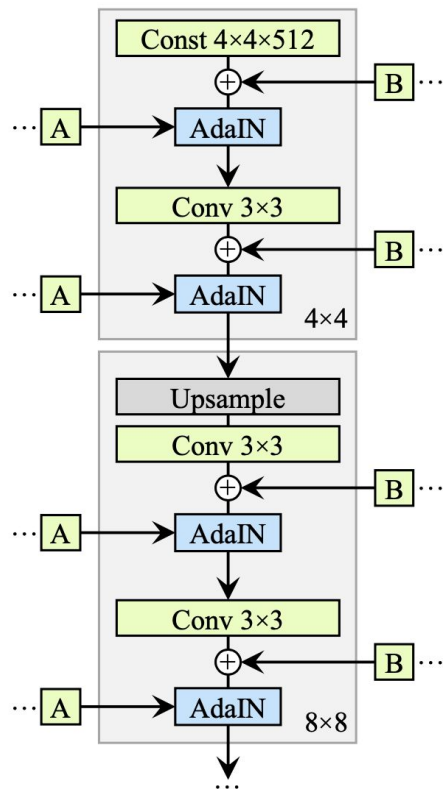


StyleGAN2

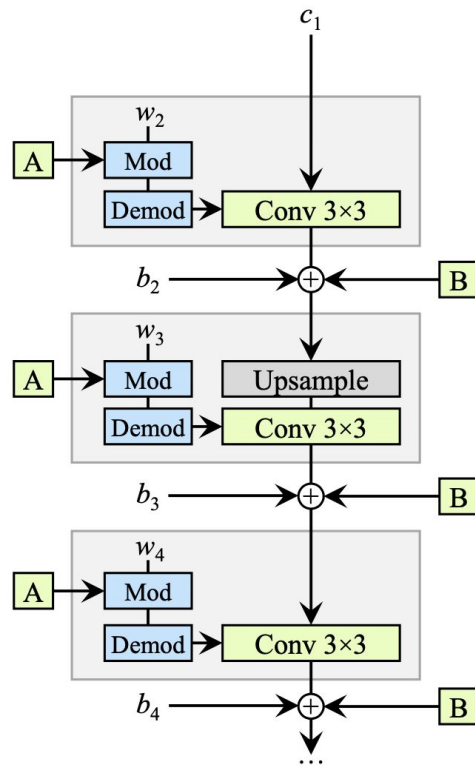
- В новом варианте нормализации встраиваем информацию о стиле в веса модели

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

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



(a) StyleGAN



(d) Weight demodulation

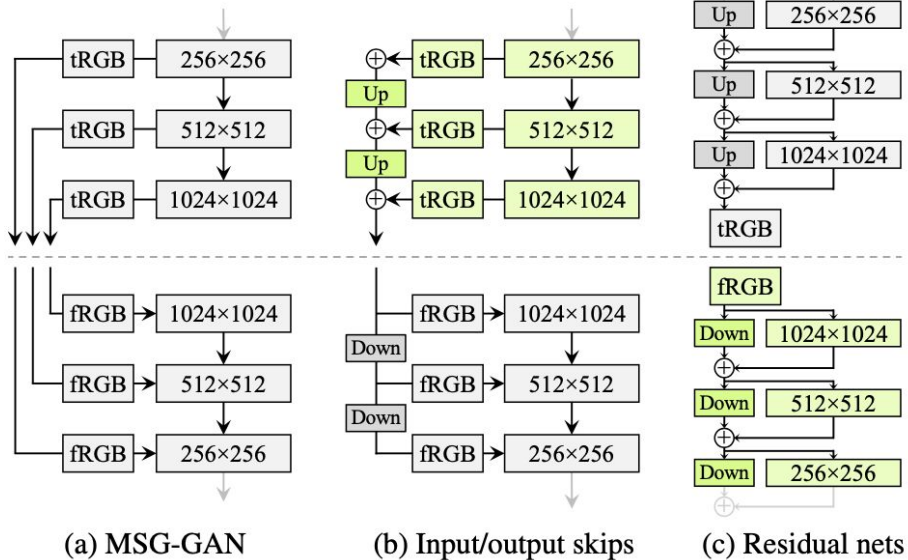
StyleGAN2



Артефакты в StyleGANv1

- Новая нормализация работает сразу с весами модели, что избавляет изображение от артефактов

StyleGAN2



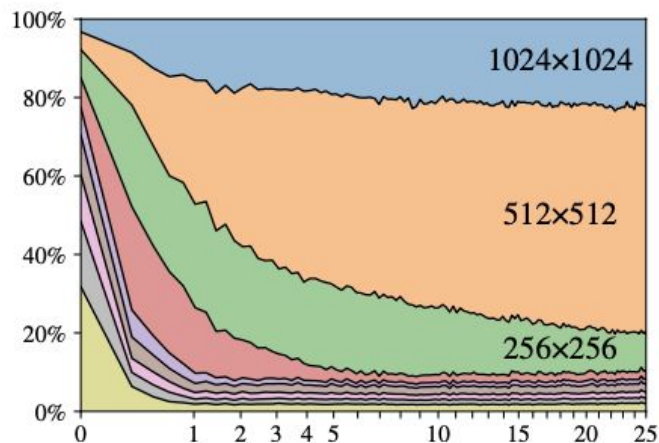
FFHQ	D original		D input skips		D residual	
	FID	PPL	FID	PPL	FID	PPL
G original	4.32	265	4.18	235	3.58	269
G output skips	4.33	169	3.77	127	3.31	125
G residual	4.35	203	3.96	229	3.79	243

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

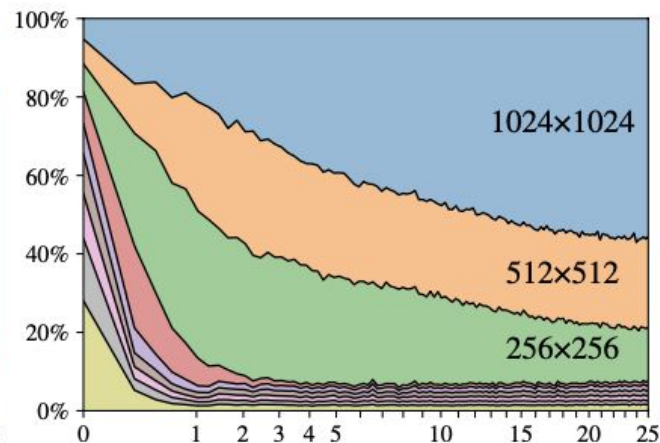
Численные
результаты

Архитектуры генератора (сверху)
и дискриминатора (снизу)

StyleGAN2



(a) StyleGAN-sized (config E)



(b) Large networks (config F)

Вклад разных в итоговое
изображение

StyleGAN2

- В новой архитектуре используется дополнительная регуляризация
- Идея в том, что изменения в латентном пространстве не должны быть слишком резкими, т.е. небольшое изменение входного шума должно чуть-чуть изменить изображение

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

\mathbf{J} - матрица якоби, a - скользящее эксп. среднее

Численные результаты для StyleGAN2

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

Table 3. Improvement in LSUN datasets measured using FID and PPL. We trained CAR for 57M images, CAT for 88M, CHURCH for 48M, and HORSE for 100M images.

ИСТОЧНИКИ

- Karras, Tero, et al. "A Style-Based Generator Architecture for Generative Adversarial Networks." arXiv:1812.04948, 29 Mar. 2019, arxiv.org/abs/1812.04948.
- Karras, Tero, et al. "Analyzing and Improving the Image Quality of StyleGAN." arXiv:1912.04958v2, 23 Mar. 2020, arxiv.org/abs/1912.04958v2.
- Лекции курса по генеративным моделям ШАД