Генеративно-состязательные нейронные сети и их друзья

Перминов Андрей

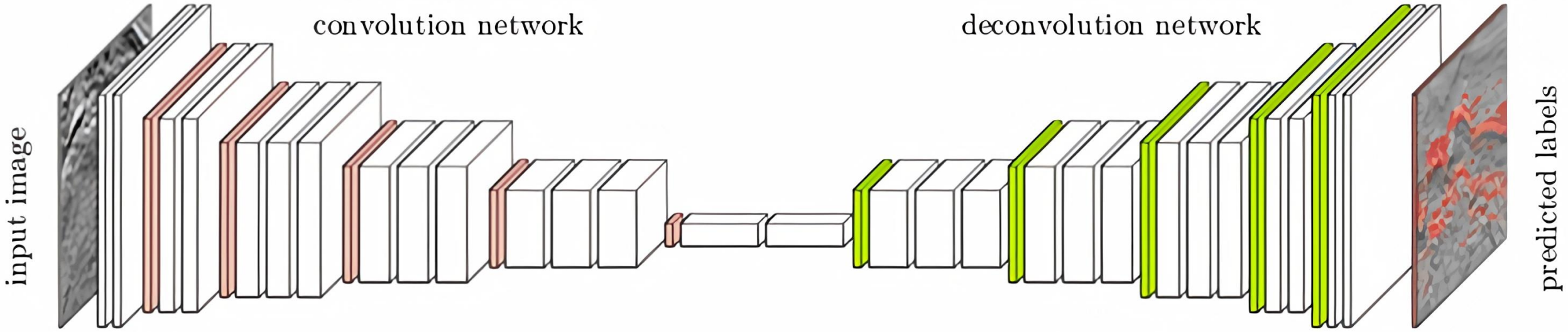
13 октября 2021

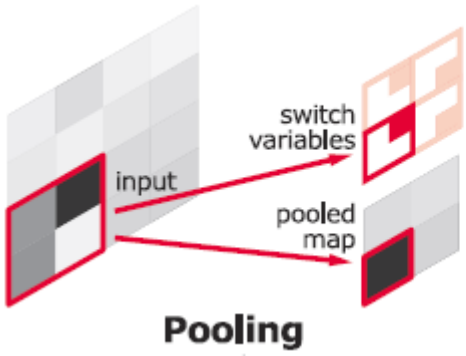
ИСП РАН

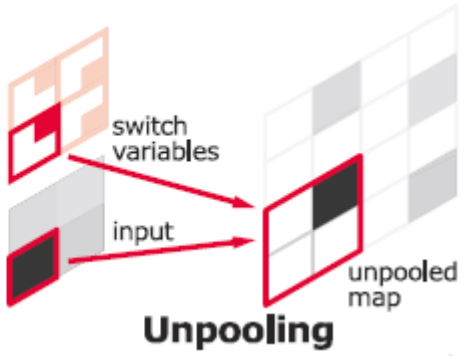
They don't appear to want to compete. They just want to dance.

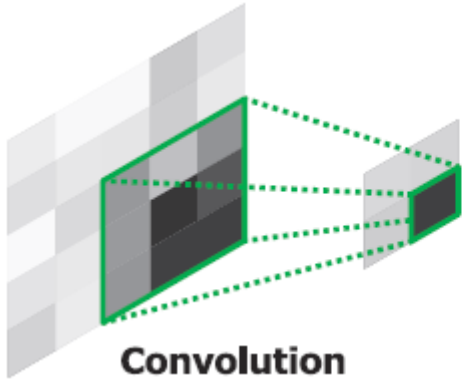
Деконволюционная нейронная сеть

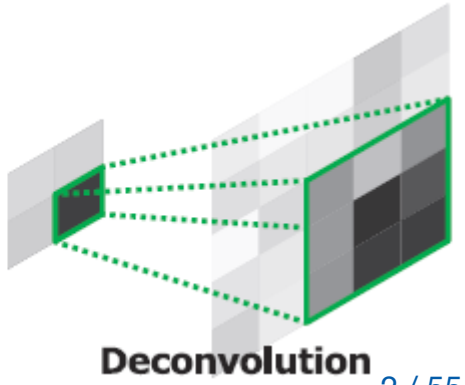
ИСП РАН







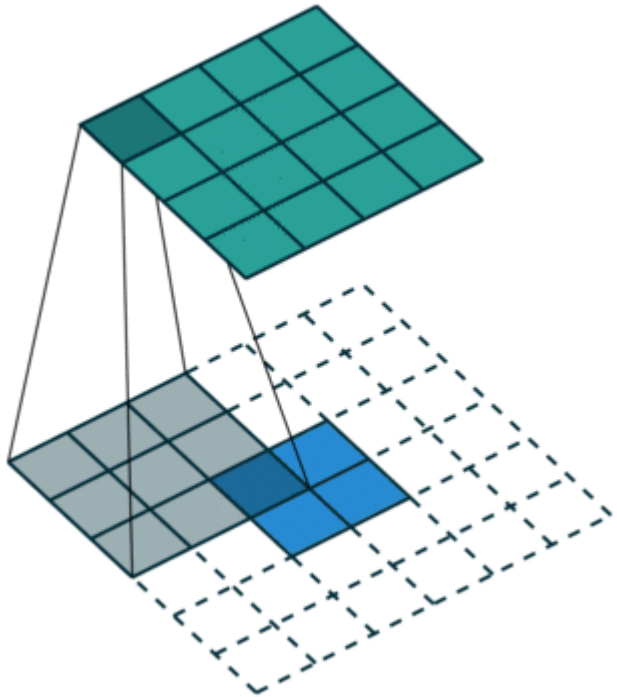




Обратная (транспонированная) свёртка

ИСП РАН

- позволяет увеличивать размер входного тензора;
- работает аналогично шагу обратного распространения ошибки в свёрточном слое;
- в отличие от обычных интерполяционных слоёв обучается и показывает более высокие результаты;
- при вычислении фильтры поворачиваются на 180°



Размерность выхода слоя обратной свёртки

ИСП РАН

Если к изображению размера $W_{in} \times H_{in}$ применить обратную свертку размера $F_w \times F_h$ с дополнением нулями P и шагом S, то размер выходного изображения будет равен $W_{out} \times H_{out}$, где:

$$W_{out} = (W_{in} - 1) \cdot S + F_w - 2P$$
$$H_{out} = (H_{in} - 1) \cdot S + F_h - 2P$$

Примеры:

$F = 3 \times 3, P = 1, S = 1$:

$W_{out} = (W_{in} - 1) \cdot 1 + 3 - 2 = W_{in}$

$H_{out} = (H_{in} - 1) \cdot 1 + 3 - 2 = H_{in}$

$F = 5 \times 3, P = 2, S = 2$:

$W_{out} = (W_{in} - 1) \cdot 2 + 5 - 2 \cdot 2 = 2 \cdot W_{in} - 1$

$H_{out} = (H_{in} - 1) \cdot 2 + 3 - 2 \cdot 2 = 2 \cdot H_{in} - 3$

Обратная свёртка (пример)

2	-4	-3
1	3	-2

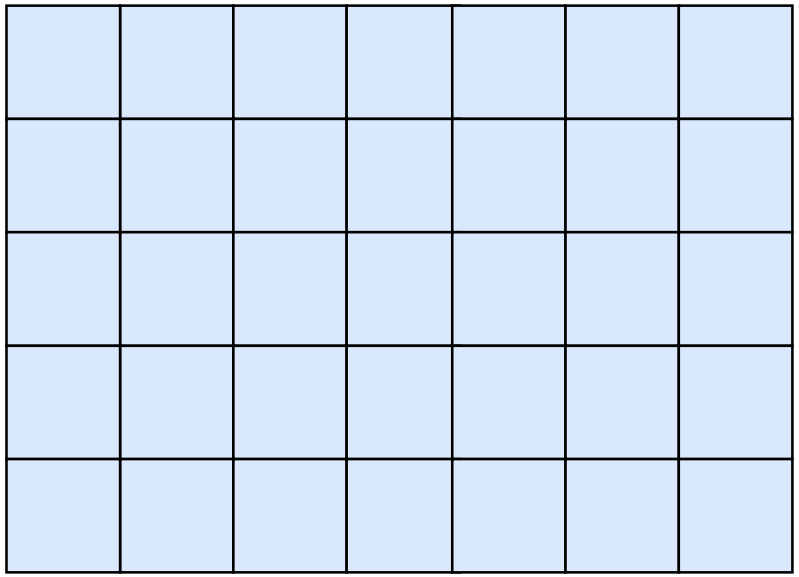
Изображение (2x3)

×

3	2	-1
-1	1	4
1	3	2

Фильтр

=



Результат свёртки: 5x7

Обратная свёртка (пример)

ИСП РАН

2	0	-4	0	-3
0	0	0	0	0
1	0	3	0	-2

Изображение (3x5)

×

2	3	1
4	1	-1
-1	2	3

Фильтр (повёрнутый)

=

Результат свёртки: 5x7

Обратная свёртка (пример)

ИСП РАН

0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	2	0	-4	0	-3	0	0
0	0	0	0	0	0	0	0	0
0	0	1	0	3	0	-2	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

Изображение + full padding (7x9)

×

2	3	1
4	1	-1
-1	2	3

Фильтр
(повёрнутый)

=

Результат свёртки: 5x7

Обратная свёртка (пример)

ИСП РАН

0 (2)	0 (3)	0 (1)	0	0	0	0	0	0
0 (4)	0 (1)	0 (-1)	0	0	0	0	0	0
0 (-1)	0 (2)	2 (3)	0	-4	0	-3	0	0
0	0	0	0	0	0	0	0	0
0	0	1	0	3	0	-2	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

Изображение + full padding (7x9)

×

2	3	1
4	1	-1
-1	2	3

Фильтр (повёрнутый)

=

6						

Результат свёртки: 5x7

Обратная свёртка (пример)

ИСП РАН

0	0_{2}	0_{3}	0_{1}	0	0	0	0	0
0	0_{4}	0_{1}	0_{-1}	0	0	0	0	0
0	0_{-1}	2_{2}	0_{3}	-4	0	-3	0	0
0	0	0	0	0	0	0	0	0
0	0	1	0	3	0	-2	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

Изображение + full padding (7x9)

×

2	3	1
4	1	-1
-1	2	3

Фильтр (повёрнутый)

=

6	4					

Результат свёртки: 5x7

Обратная свёртка (пример)

ИСП РАН

0	0	0 (2)	0 (3)	0 (1)	0	0	0	0
0	0	0 (4)	0 (1)	0 (-1)	0	0	0	0
0	0	2 (-1)	0 (2)	-4 (3)	0	-3	0	0
0	0	0	0	0	0	0	0	0
0	0	1	0	3	0	-2	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

Изображение + full padding (7x9)

×

2	3	1
4	1	-1
-1	2	3

Фильтр (повёрнутый)

=

6	4	-14				

Результат свёртки: 5x7

Обратная свёртка (пример)

ИСП РАН

0	0	0	0 (2)	0 (3)	0 (1)	0	0	0
0	0	0	0 (4)	0 (1)	0 (-1)	0	0	0
0	0	2	0 (-1)	-4 (2)	0 (3)	-3	0	0
0	0	0	0	0	0	0	0	0
0	0	1	0	3	0	-2	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

Изображение + full padding (7x9)

×

2	3	1
4	1	-1
-1	2	3

Фильтр
(повёрнутый)

=

6	4	-14	-8			

Результат свёртки: 5x7

Обратная свёртка (пример)

ИСП РАН

0	0	0	0	0 (2)	0 (3)	0 (1)	0	0
0	0	0	0	0 (4)	0 (1)	0 (-1)	0	0
0	0	2	0	-4 (-1)	0 (2)	-3 (3)	0	0
0	0	0	0	0	0	0	0	0
0	0	1	0	3	0	-2	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

Изображение + full padding (7x9)

×

2	3	1
4	1	-1
-1	2	3

Фильтр (повёрнутый)

=

6	4	-14	-8	-5		

Результат свёртки: 5x7

Обратная свёртка (пример)

ИСП РАН

0	0	0	0	0	0_{2}	0_{3}	0_{1}	0
0	0	0	0	0	0_{4}	0_{1}	0_{-1}	0
0	0	2	0	-4	0_{-1}	-3_{2}	0_{3}	0
0	0	0	0	0	0	0	0	0
0	0	1	0	3	0	-2	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

Изображение + full padding (7x9)

×

2	3	1
4	1	-1
-1	2	3

Фильтр
(повёрнутый)

=

6	4	-14	-8	-5	-6	

Результат свёртки: 5x7

Обратная свёртка (пример)

ИСП РАН

0	0	0	0	0	0	0_{2}	0_{3}	0_{1}
0	0	0	0	0	0	0_{4}	0_{1}	0_{-1}
0	0	2	0	-4	0	-3_{-1}	0_{2}	0_{3}
0	0	0	0	0	0	0	0	0
0	0	1	0	3	0	-2	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

Изображение + full padding (7x9)

×

2	3	1
4	1	-1
-1	2	3

Фильтр (повёрнутый)

=

6	4	-14	-8	-5	-6	3

Результат свёртки: 5x7

Обратная свёртка (пример)

ИСП РАН

0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	2	0	-4	0	-3	0	0
0	0	0	0	0	0	0	0	0
0	0	1	0	3	0	-2_{2}	0_{3}	0_{1}
0	0	0	0	0	0	0_{4}	0_{1}	0_{-1}
0	0	0	0	0	0	0_{-1}	0_{2}	0_{3}

Изображение + full padding (7x9)

×

2	3	1
4	1	-1
-1	2	3

Фильтр (повёрнутый)

=

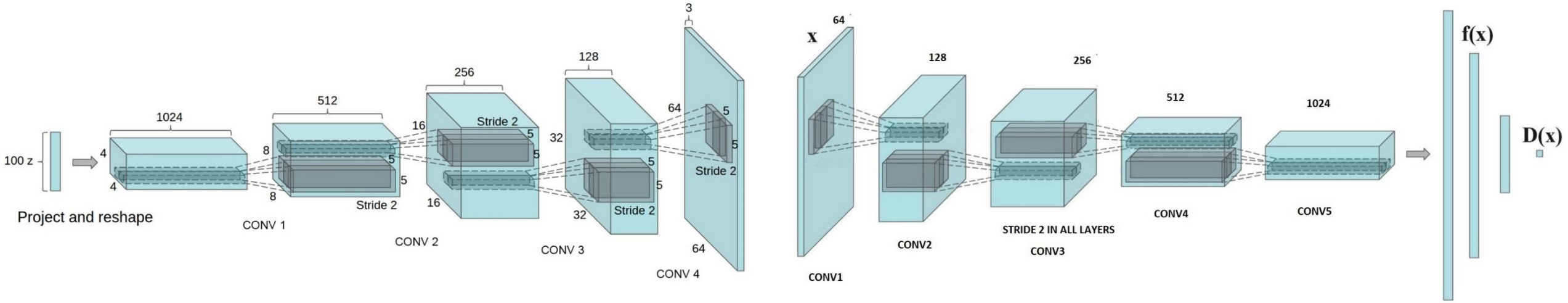
6	4	-14	-8	-5	-6	3
-2	2	12	-4	-13	-3	-12
5	8	8	-6	-20	-13	-4
-1	1	1	3	14	-2	-8
1	3	5	9	4	-6	-4

Результат свёртки: 5x7

Генеративно-состязательная сеть (GAN)

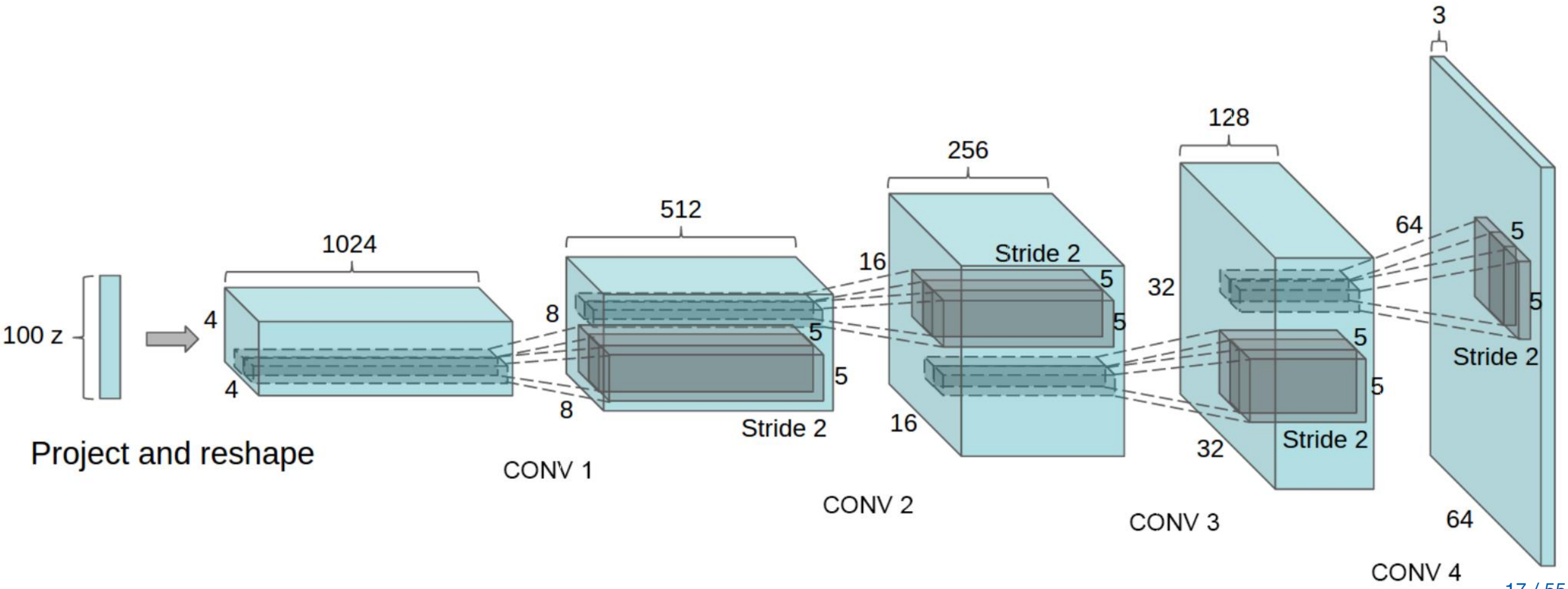
ИСП РАН

Генеративно-состязательная нейросеть (generative adversarial network, GAN) — архитектура, состоящая из двух независимых нейронных сетей, настроенных на работу друг против друга: генератора и дискриминатора.



ИСП РАН

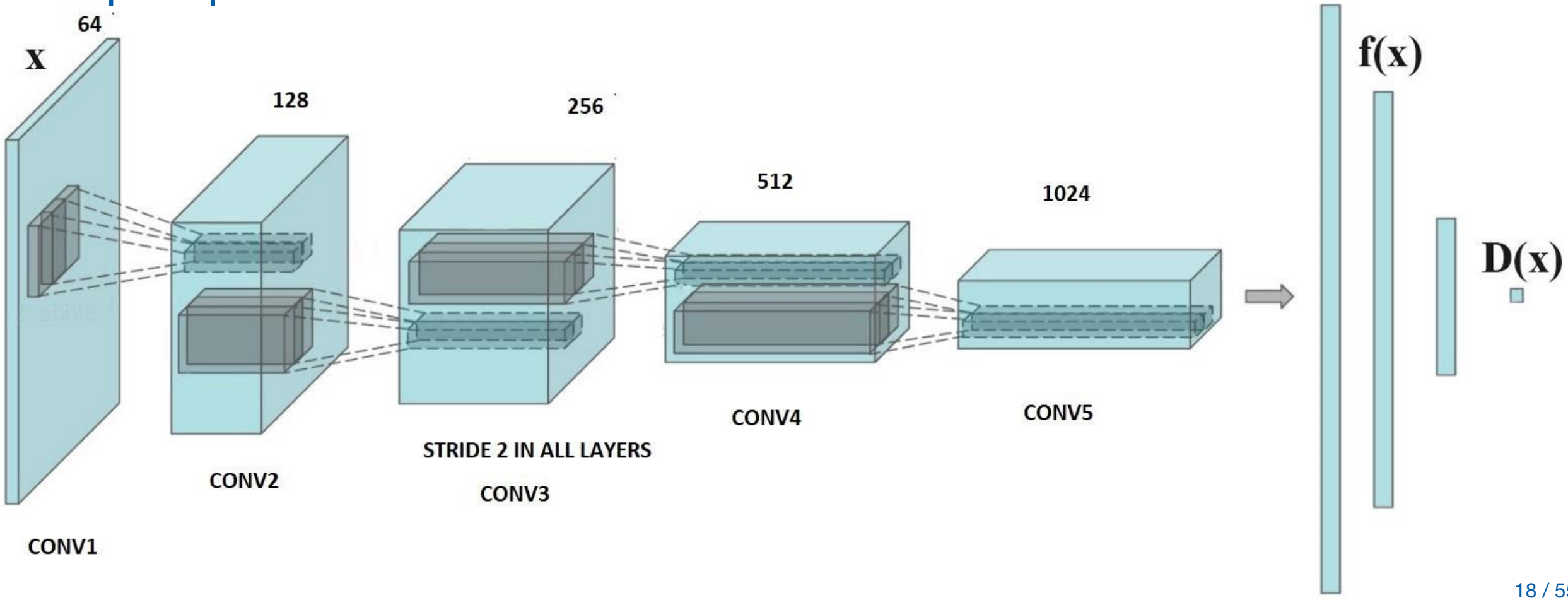
Генератор – сеть, создающая изображения, используя произвольный вектор небольшой размерности (обычно около 64-256).



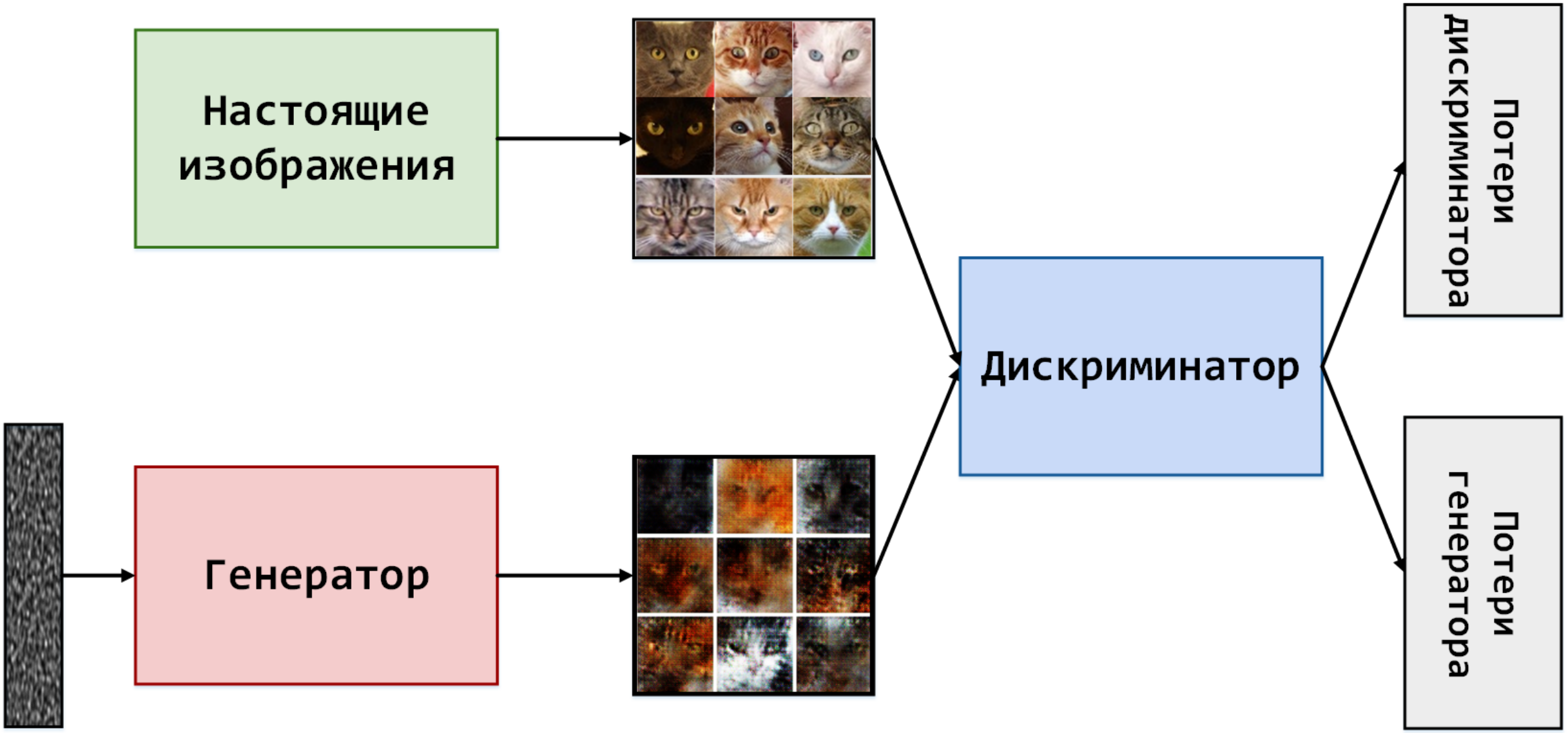
Дискриминатор

ИСП РАН

Дискриминатор – бинарный классификатор, который учится отличать настоящие изображения от изображений, созданных генератором.



Обучение генеративно-состязательной сети



Обучение генеративно-состязательной сети

ИСП РАН

Обучаем дискриминатор предсказывать метку «1» на батче из реальных изображений.

Обучение генеративно-состязательной сети

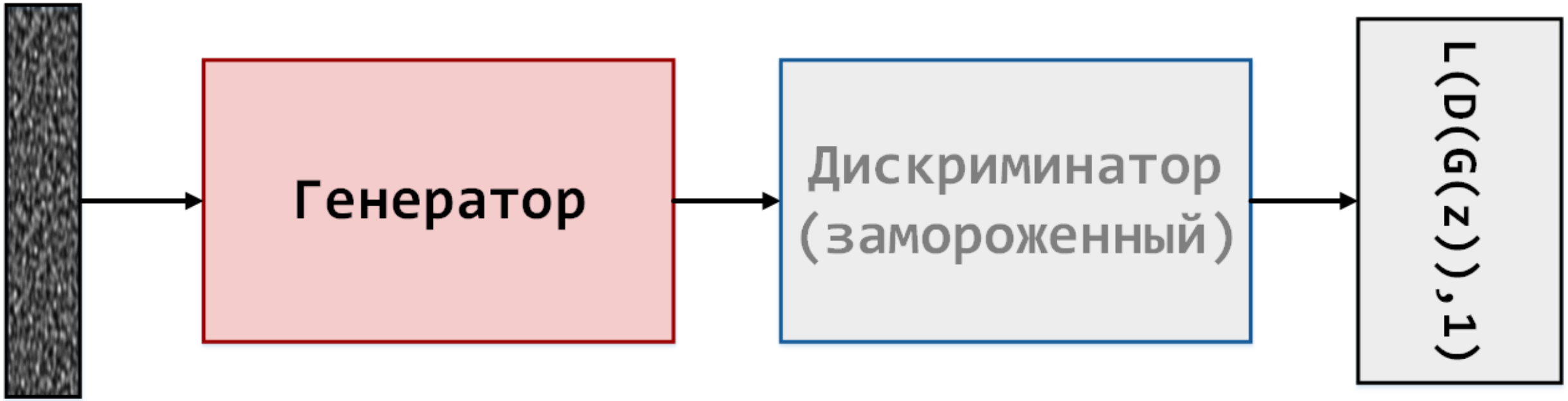
ИСП РАН

Обучаем дискриминатор предсказывать метку «0» на батче из изображений, созданных генератором.

Обучение генеративно-состязательной сети

ИСП РАН

Замораживаем весовые коэффициенты дискриминатора и обучаем полную сеть предсказывать метку «1» для произвольного входного вектора.



Вопрос: зачем замораживать веса дискриминатора?

Функция потерь дискриминатора

Функция потерь может быть получена из формулы бинарной перекрёстной энтропии:

$$L(\hat{y}, y) = y \cdot ln\, \hat{y} + (1 - y) \cdot \ln(1 - \hat{y})$$

При обучении дискриминатора на настоящих данных $y = 1, \hat{y} = D(x)$, следовательно:

$$L(\hat{y}, y) = L(D(x), 1) = \ln D(x)$$

При обучении дискриминатора на изображениях, созданных генератором $y = 0, \hat{y} = D\big(G(z)\big)$:

$$L(\hat{y}, y) = L(D(G(z)), 0) = \ln\Big(1 - D\big(G(z)\big)\Big)$$

Функция потерь дискриминатора

Цель дискриминатора – правильно классифицировать настоящий и поддельный набор данных, для этого значения функции должны быть максимизированы, а окончательная функция потерь будет выглядеть так:

$$L^{(D)} = \max\left[\ln D(x) + \ln\left(1 - D\left(G(z)\right)\right)\right]$$

Функция потерь генератора

Генератор постоянно борется с дискриминатором и пытается минимизировать уравнение:

$$L^{(G)} = \min\left[\ln D(x) + \ln\left(1 - D\left(G(z)\right)\right)\right]$$

Комбинированная функция потерь

ИСП РАН

Оба уравнения можно объединить в одно:

$$L = min_G\, max_D \left[\ln D(x) + \ln\left(1 - D\left(G(z)\right)\right)\right]$$

В реальности используют две разных функции:

$$\mathrm{L}_{discriminator} = \ln D(x) + \ln\left(1 - D\left(G(z)\right)\right)$$

$$L_{generator} = \ln\left(D\left(G(z)\right)\right)$$

Алгоритм обучения GAN

ИСП РАН

for количество_итераций **do**

сгенерировать набор из ***m*** случайных векторов $\{z_1, z_2, \dots, z_m\}$

выбрать ***m*** настоящих изображений $\{x_1, x_2, \dots, x_m\}$

обновить веса дискриминатора в сторону возрастания градиента:

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

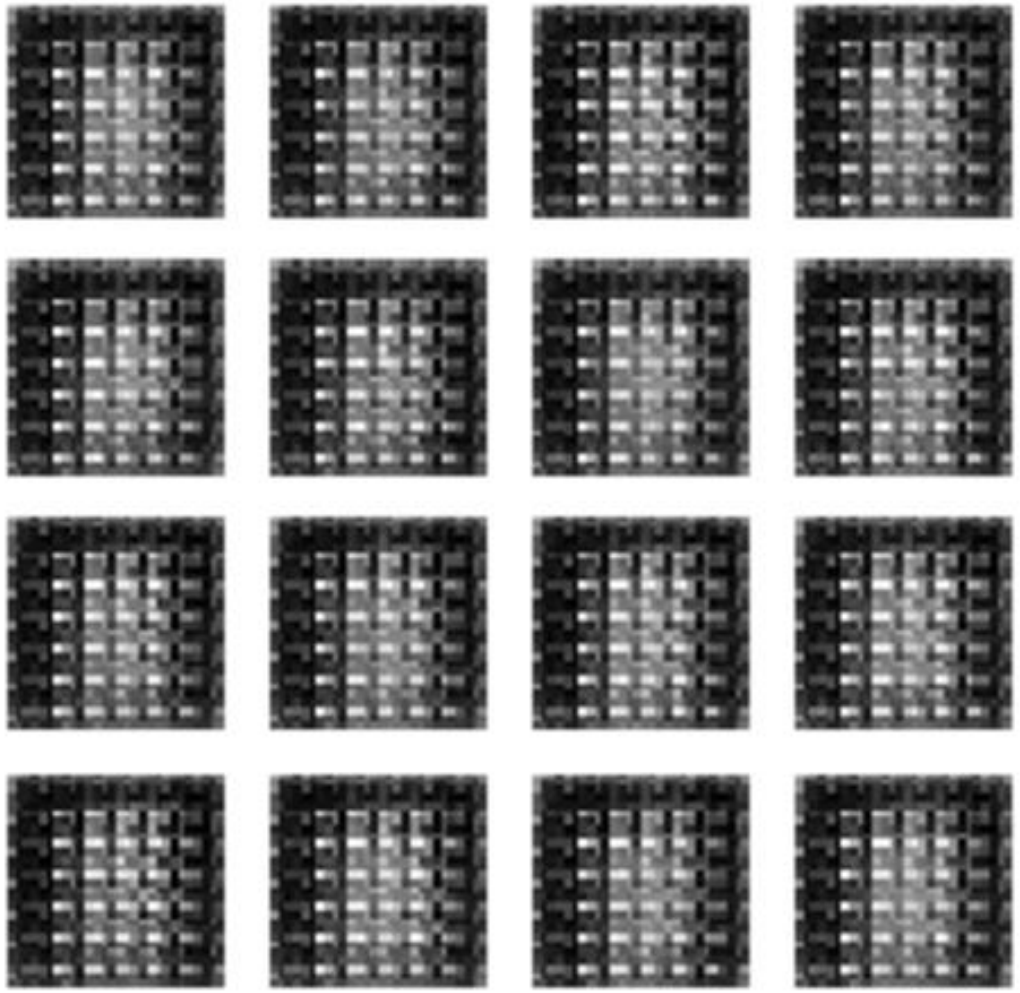
сгенерировать набор из ***m*** случайных векторов $\{z_1, z_2, \dots, z_m\}$

обновить веса дискриминатора в сторону уменьшения градиента:

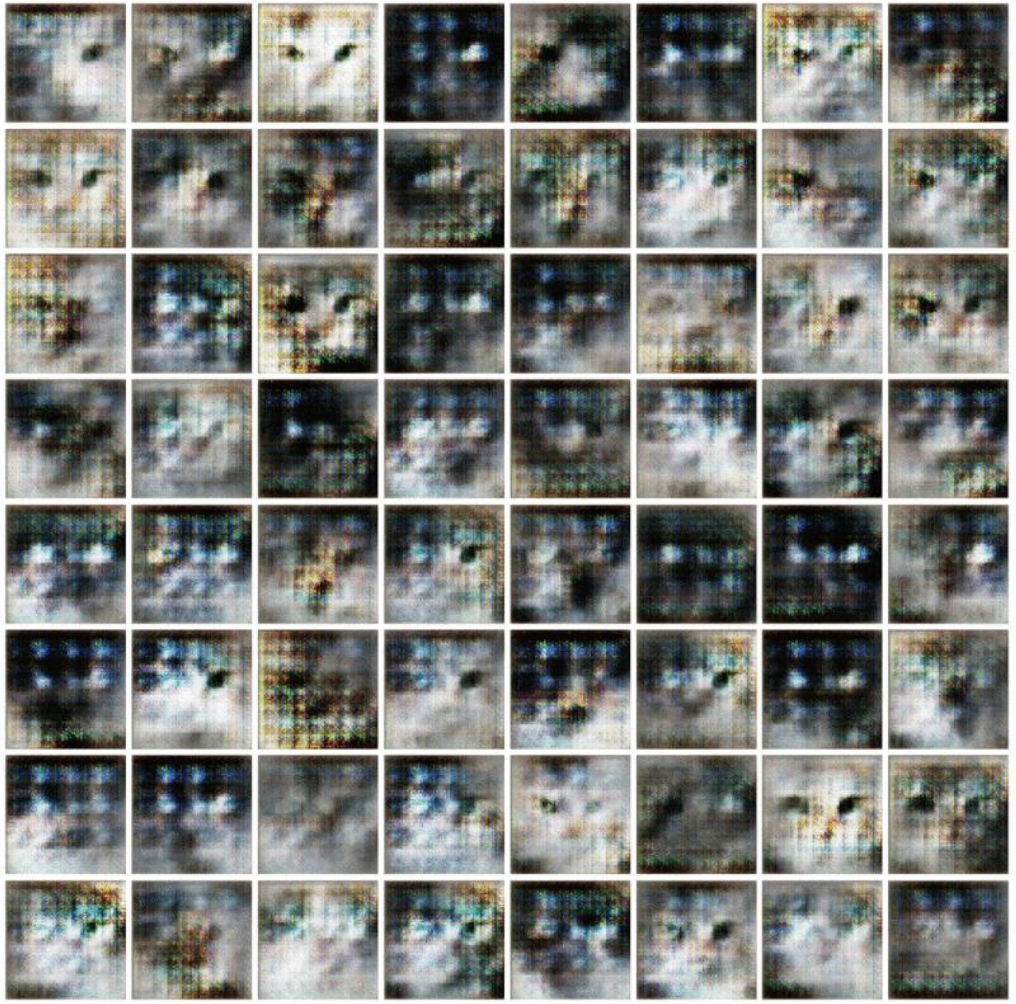
$$\nabla_{\theta_g} \frac{1}{m} \sum_{i=1}^{m} \left[\ln\left(1 - D\left(G\left(z^{(i)}\right)\right)\right) \right]$$

end for

Процесс обучения GAN



Обучение на наборе данных MNIST (.gif)



Обучение на наборе данных Cat faces (.gif)

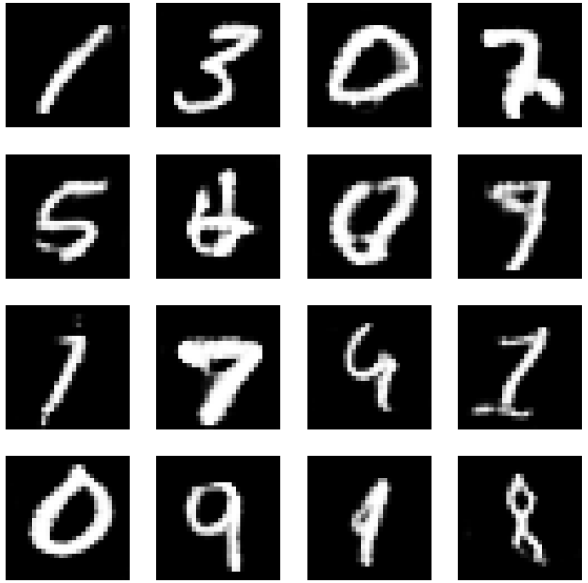
Качественная оценка модели

ИСП РАН

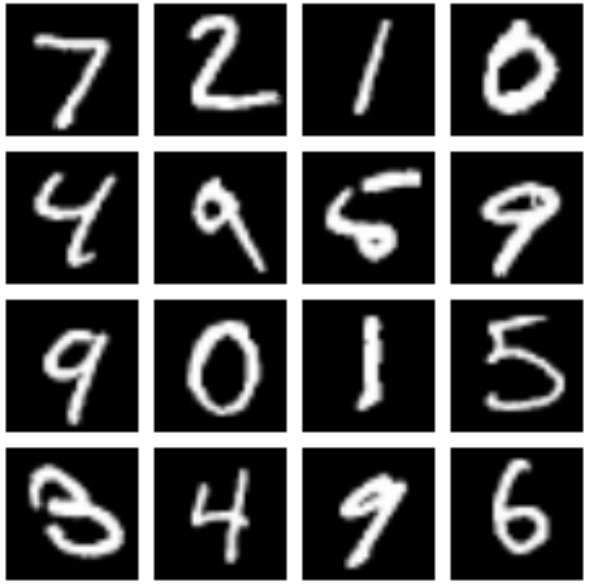
Для качественной оценки изображений, созданных генератором, можно использовать следующие характеристики:

- сходство с изображениями обучающей выборки;
- отсутствие дубликатов из обучающей выборки;
- разнообразие изображений;
- отсутствие артефактов;

Низкое качество



Высокое качество



Обучающие данные

Количественная оценка модели. FID score

ИСП РАН

Frechet Inception Distance score (FID) – метрика, позволяющая оценить, насколько похожи два набора изображений между собой.

$$FID = \|\mu_r - \mu_g\|^2 + Tr\left(\Sigma_r + \Sigma_g - 2\left(\Sigma_r\Sigma_g\right)^{\frac{1}{2}}\right)$$

В качестве значений r и g используется результат работы сети InceptionV3 без последнего (классифицирующего) слоя:

$$r = InceptionV_3(x_{real})$$
$$g = InceptionV_3(x_{gen})$$

Вопрос: при каких значения FID модель лучше?

Количественная оценка модели. FID score

ИСП РАН

Frechet Inception Distance score (FID) – метрика, позволяющая оценить, насколько похожи два набора изображений между собой.

$$FID = \|\mu_r - \mu_g\|^2 + Tr\left(\Sigma_r + \Sigma_g - 2\left(\Sigma_r\Sigma_g\right)^{\frac{1}{2}}\right)$$

В качестве значений r и g используется результат работы сети InceptionV3 без последнего (классифицирующего) слоя:

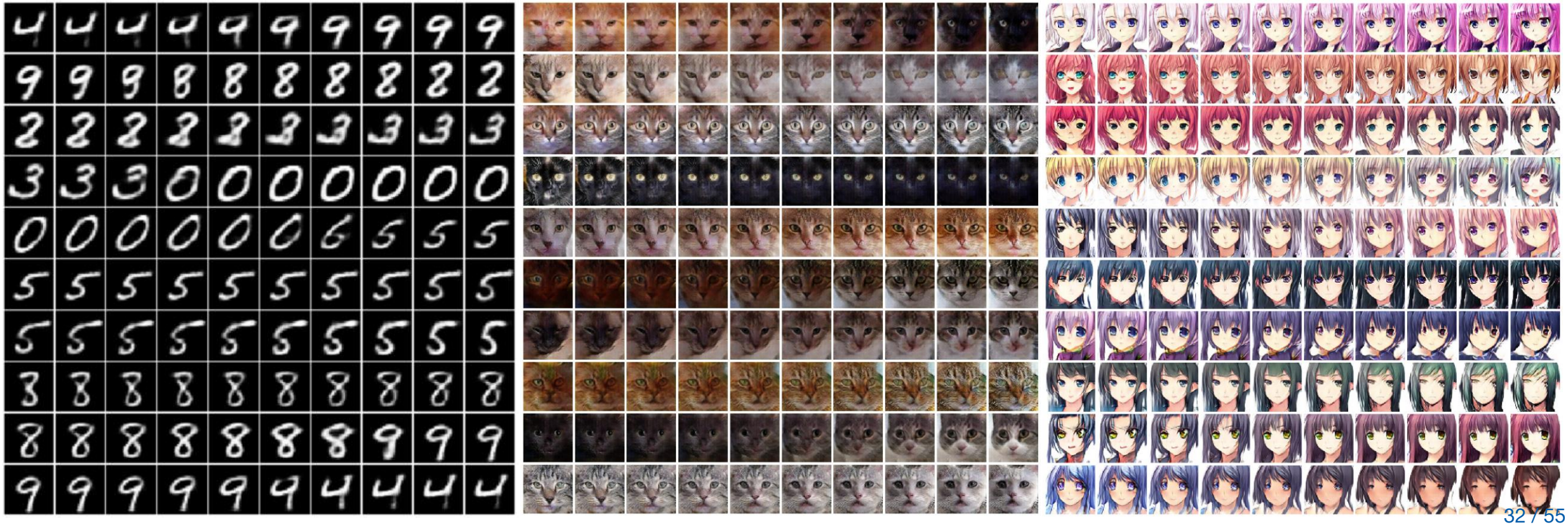
$$r = InceptionV_3(x_{real})$$
$$g = InceptionV_3(x_{gen})$$

FID – расстояние, поэтому модель тем лучше, чем ниже значение метрики.

Интерполяция в скрытом пространстве

ИСП РАН

Выбирая два вектора (z_1 и z_2) в скрытом пространстве и выполняя линейную интерполяцию между ними, можно получить интерполяцию в пространстве изображений: $z = t \cdot z_1 + (1 - t) \cdot z_2$, где $t \in [0, 1]$

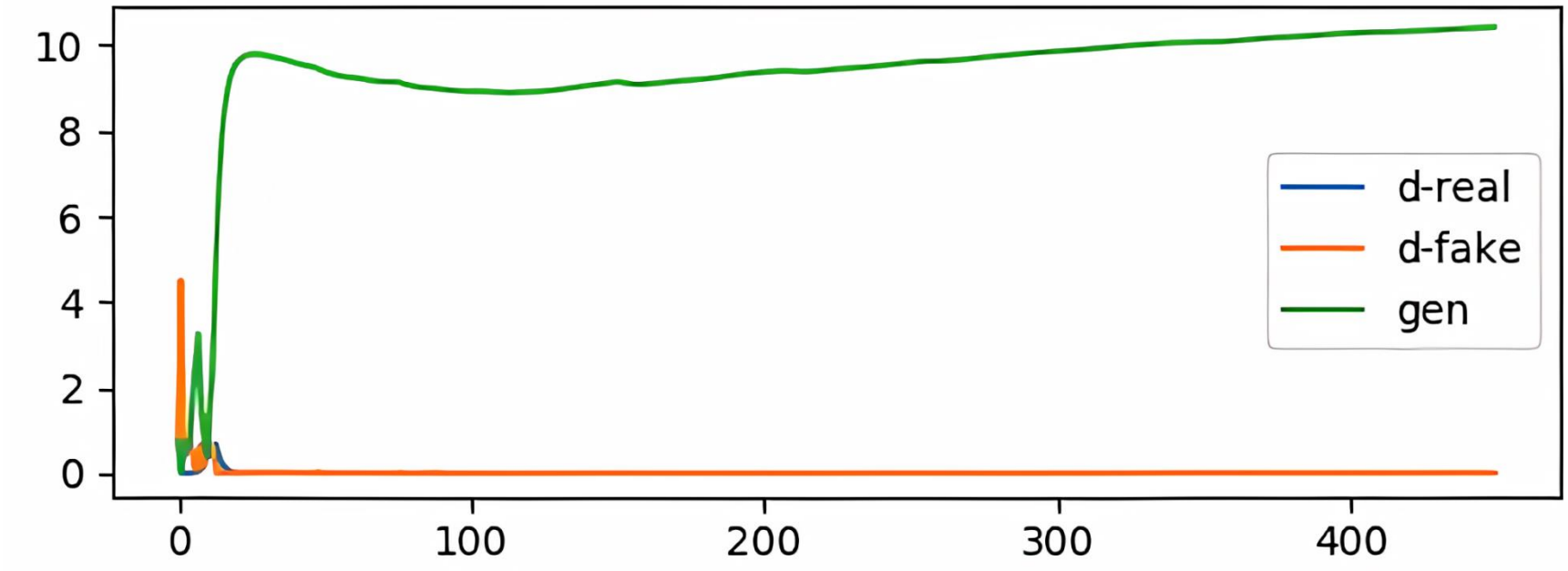


Ошибки при обучении. Затухающие градиенты

ИСП РАН

Если дискриминатор слишком хорош, то обучение генератора может не сработать из-за затухающих градиентов.

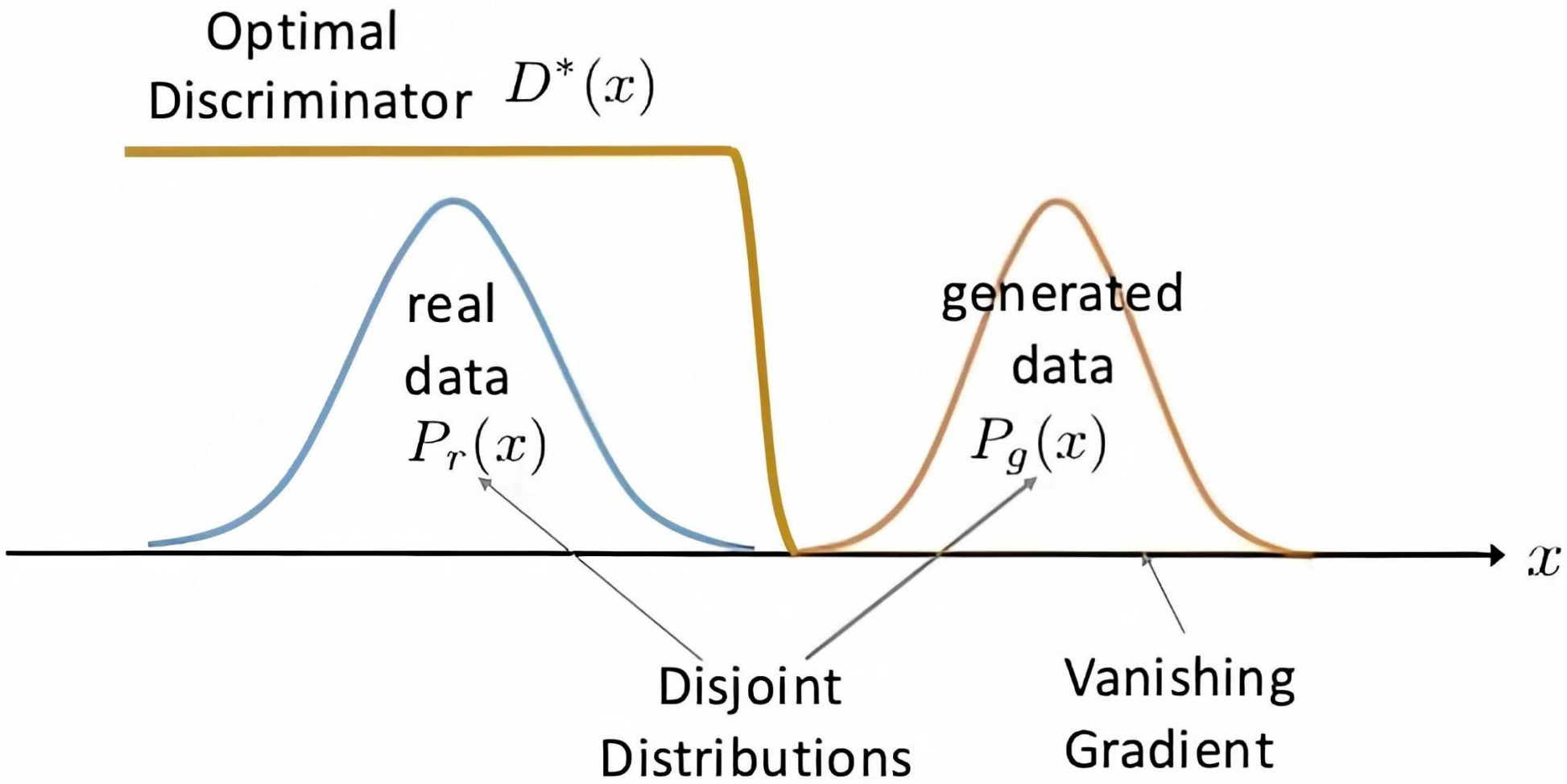
Оптимальный дискриминатор не предоставляет достаточно информации для работы генератора.



Ошибки при обучении. Неспособность сойтись

ИСП РАН

GAN’ы сложно обучать. Многие генеративно-состязательные модели вовсе не способны сойтись в приемлемую точку.



Ошибки при обучении. Коллапс моды

ИСП РАН

Режим работы GAN, при котором генератор выдаёт лишь сильно ограниченное количество изображений.

Приёмы для стабильного обучения

ИСП РАН

- нормализация входных данных (изображения в диапазоне [-1, 1])
- tanh в качестве функции активации генератора;
- нормальное распределение вместо равномерного для шума;
- раздельные батчи для настоящих и сгенерированных изображений;
- использование слоёв батч-нормализации;
- использование LeakyReLU вместо ReLU;
- добавление шума в метки ([0, 0.3] и [0.7, 1.2] вместо 0 и 1)
- использование Adam со значением момента 0.5
- использование свёрточных слоёв с шагом > 1 вместо макспулинга;

Друзья GAN’ов

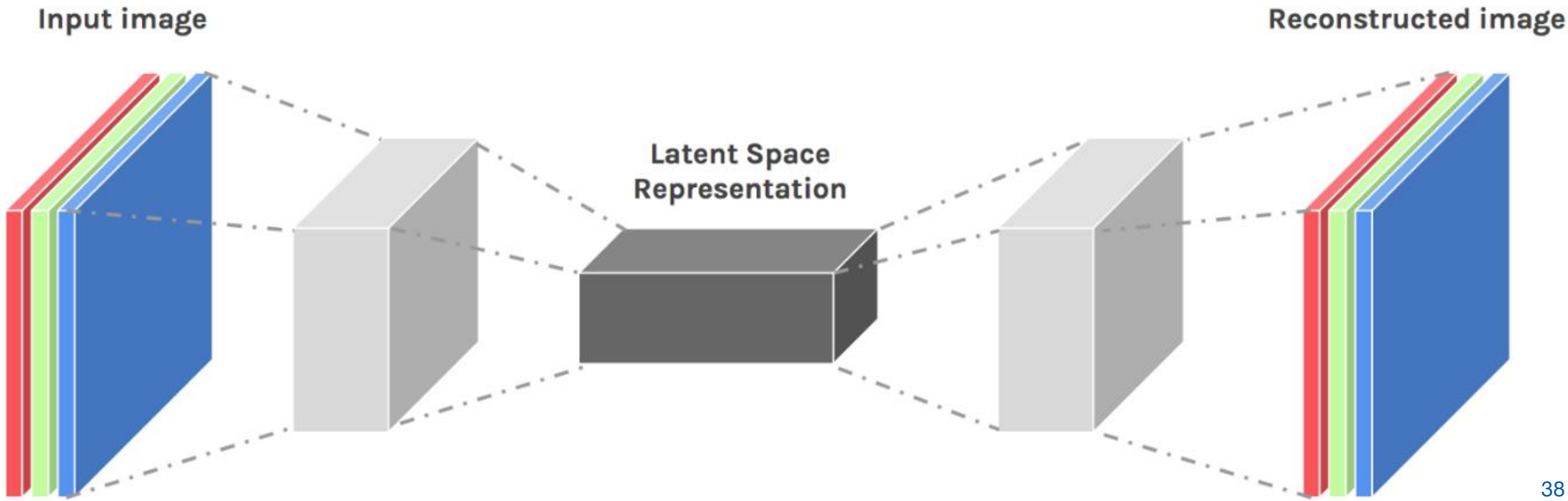
ИСП РАН

Автокодировщики
Вариационные автокодировщики
Wasserstein GAN (WGAN)
Progressive growing GAN (PROGAN)
Conditional GAN (CGAN)
Pix2pix
Cycle GAN
Stack GAN
Super resolution GAN (SRGAN)

Автокодировщики

ИСП РАН

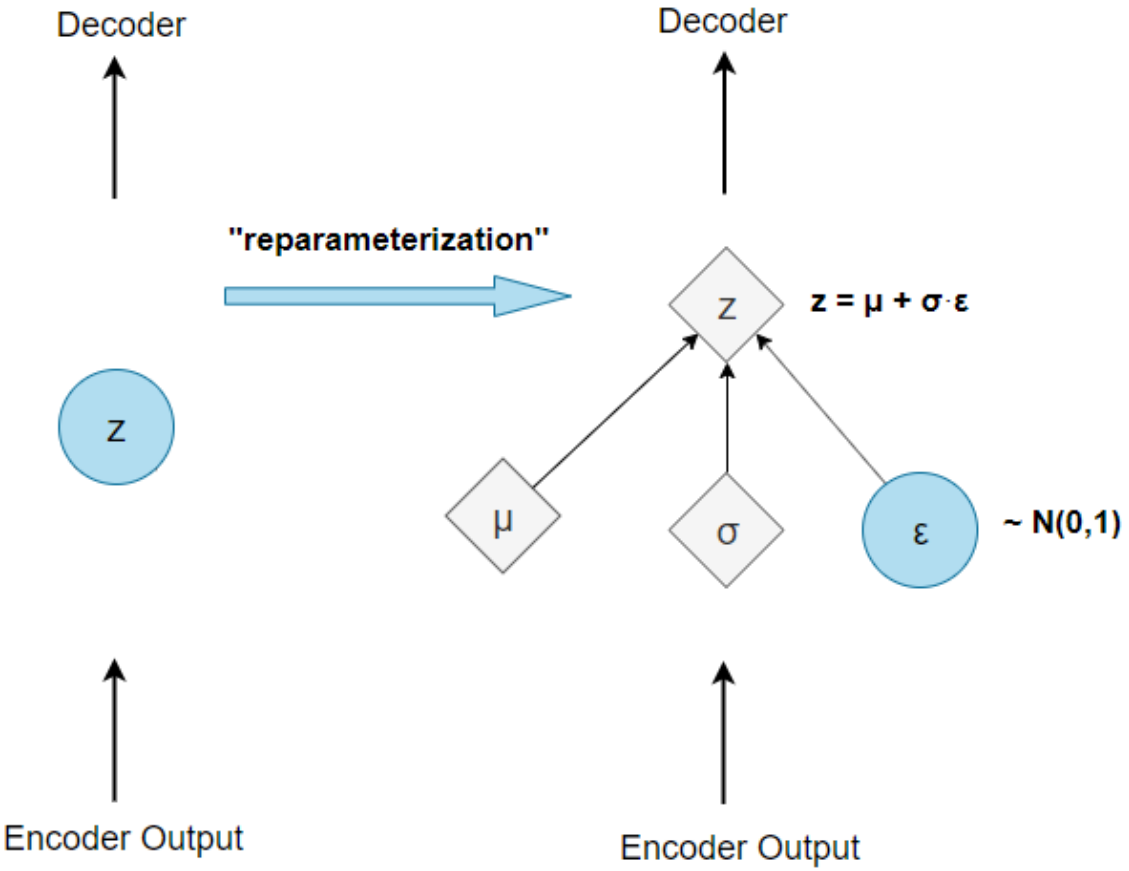
Автокодировщик (AE) – архитектура нейронной сети, состоящая из кодировщика и декодировщика. Принцип работы заключается в сжатии входного изображения в вектор малой размерности и восстановление исходного изображения из полученного вектора.

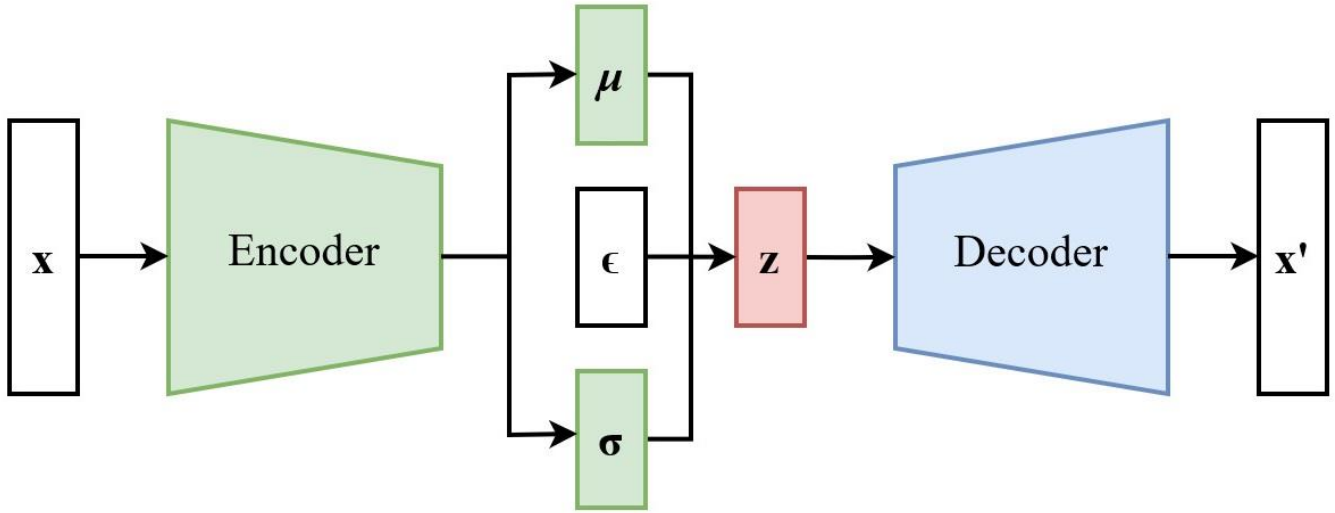


Вариационный автокодировщик (VAE)

ИСП РАН

Улучшенная версия обычного автокодировщика.
Между выходом кодировщика и входом декодировщика добавляется слой репараметризации, который изменяет нормальное распределение, превращая его в распределение скрытого пространства.

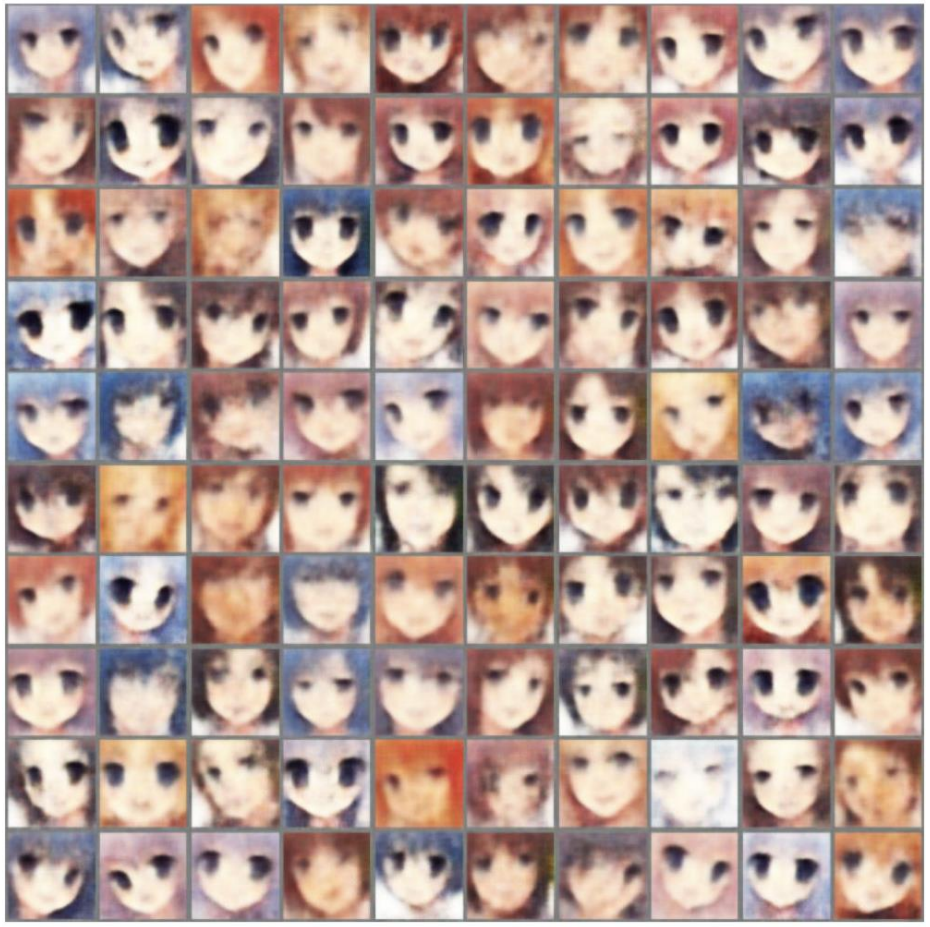


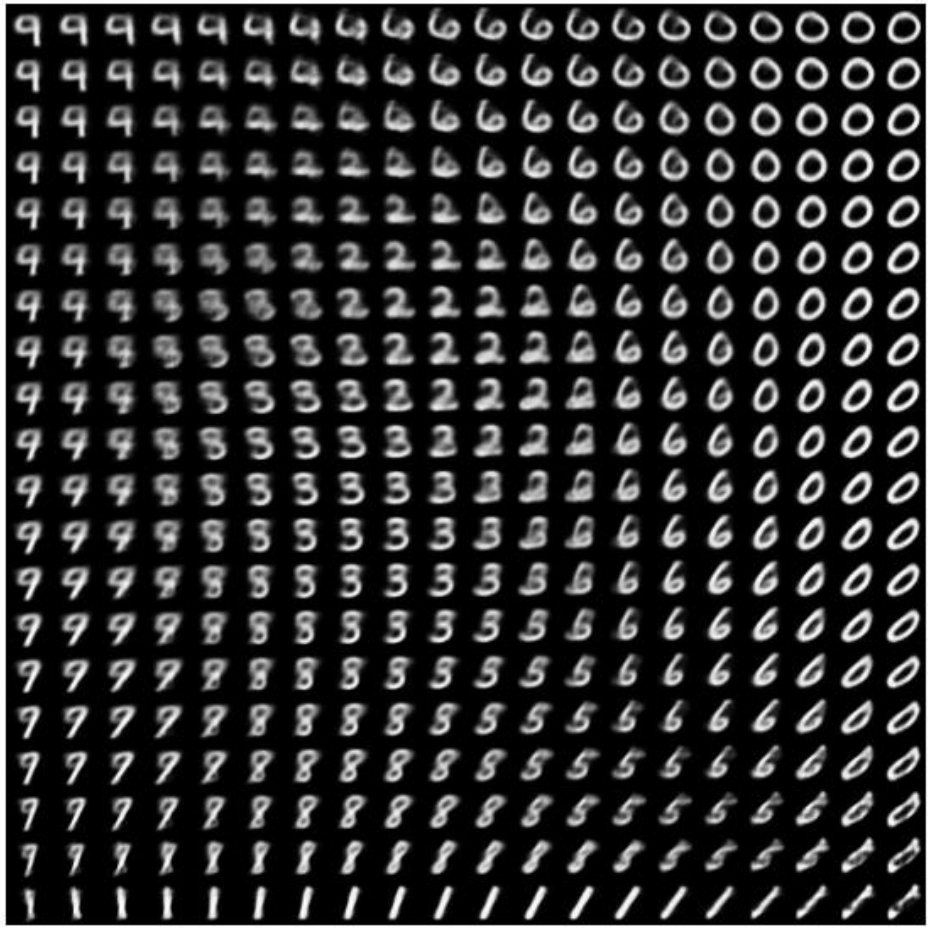


Проблемы автокодировщика

ИСП РАН

Изображения, получаемые на выходе вариационных автокодировщиков, зачастую сильно размыты и слабо отличаются от тренировочных примеров.





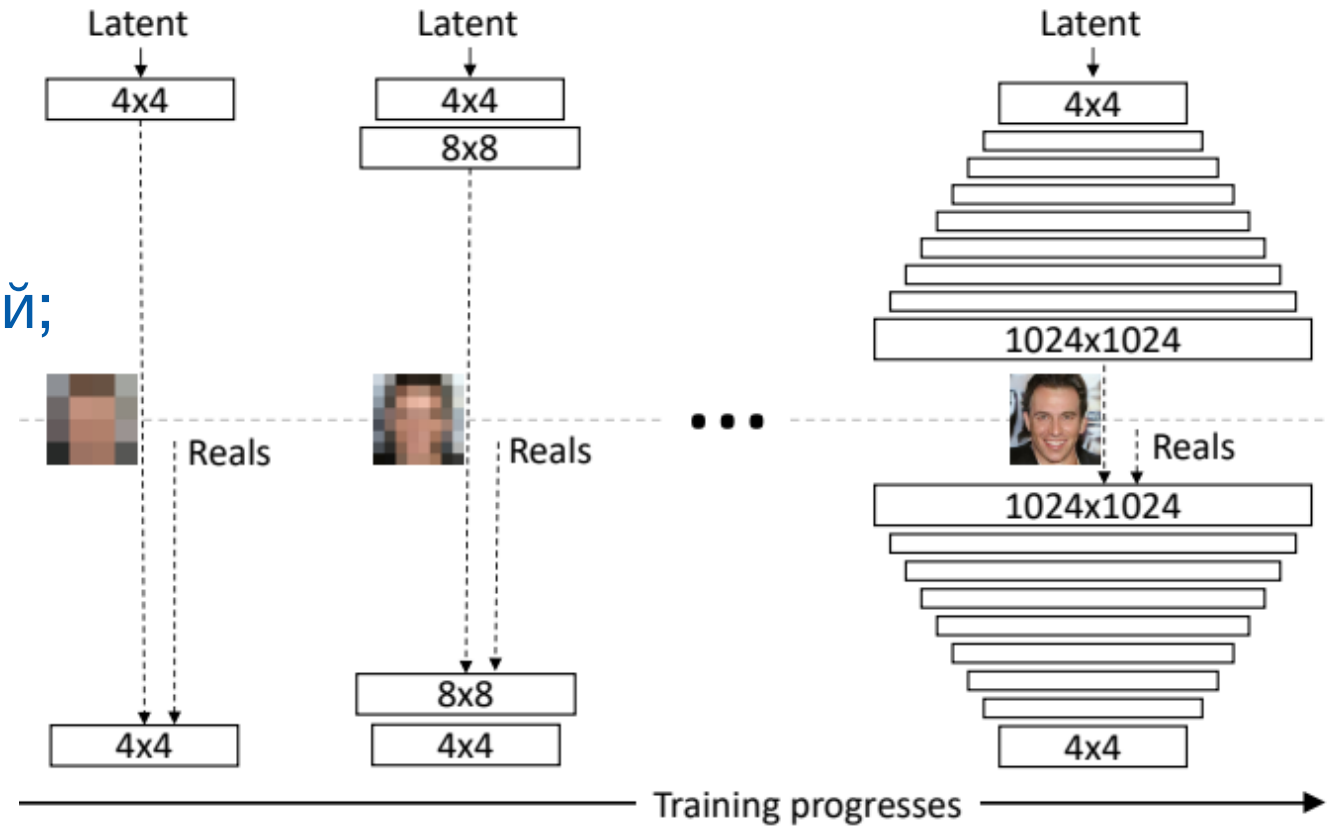
Progressive growing GAN

ИСП РАН

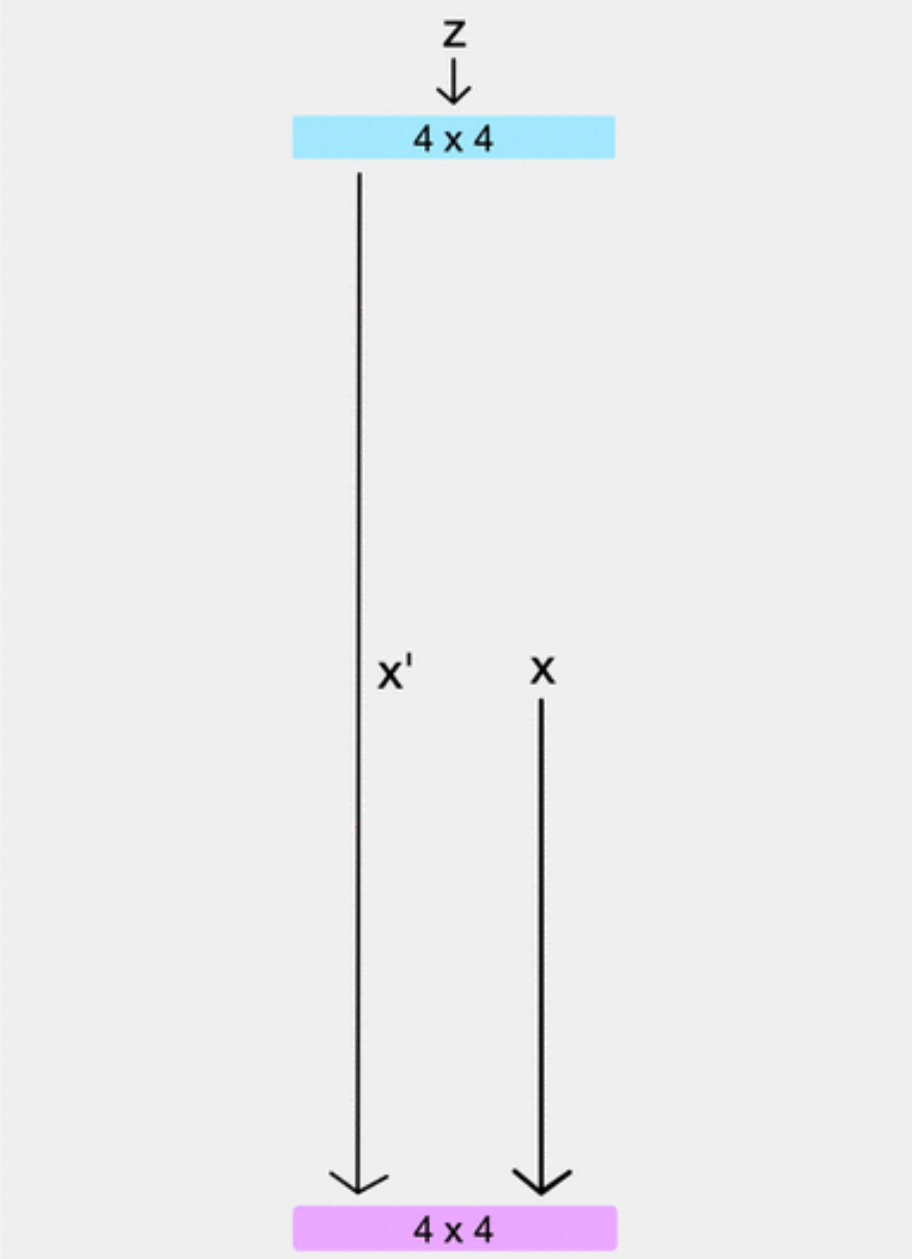
Прогрессивная генеративно состязательная сеть: во время обучения обе нейросети (генератор и дискриминатор) развиваются посредством добавления новых слоёв.

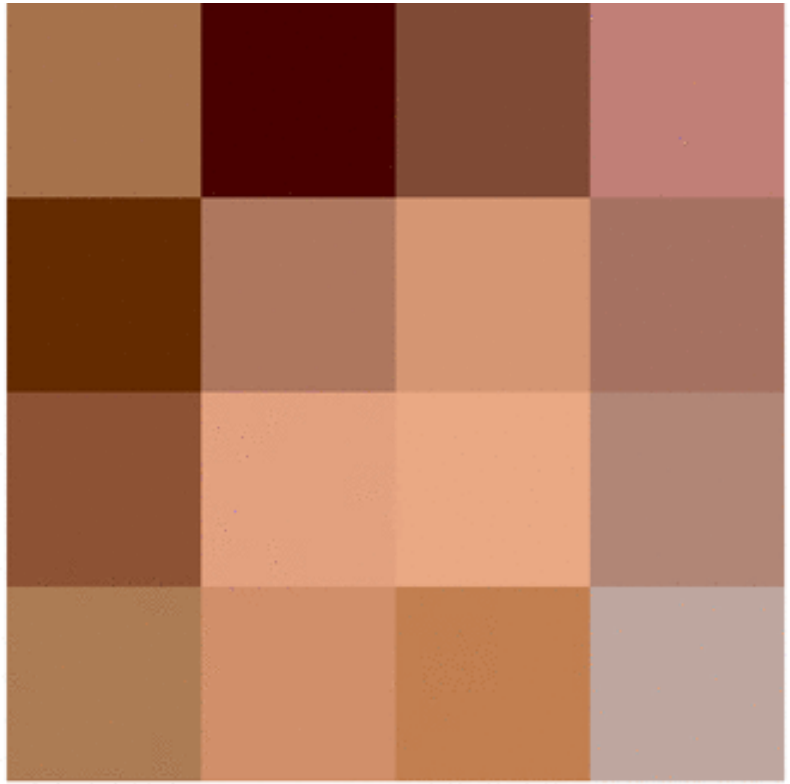
Основные плюсы:

- более реалистичные изображения;
- изображения в высоком качестве;
- большая вариативность изображений;
- легче и быстрее обучается;



Progressive growing GAN (gif)





Training time: 0 days

4x4 resolution

Generator

Discriminator

z = random code

x = real image

x' = generated image

Progressive growing GAN (примеры)

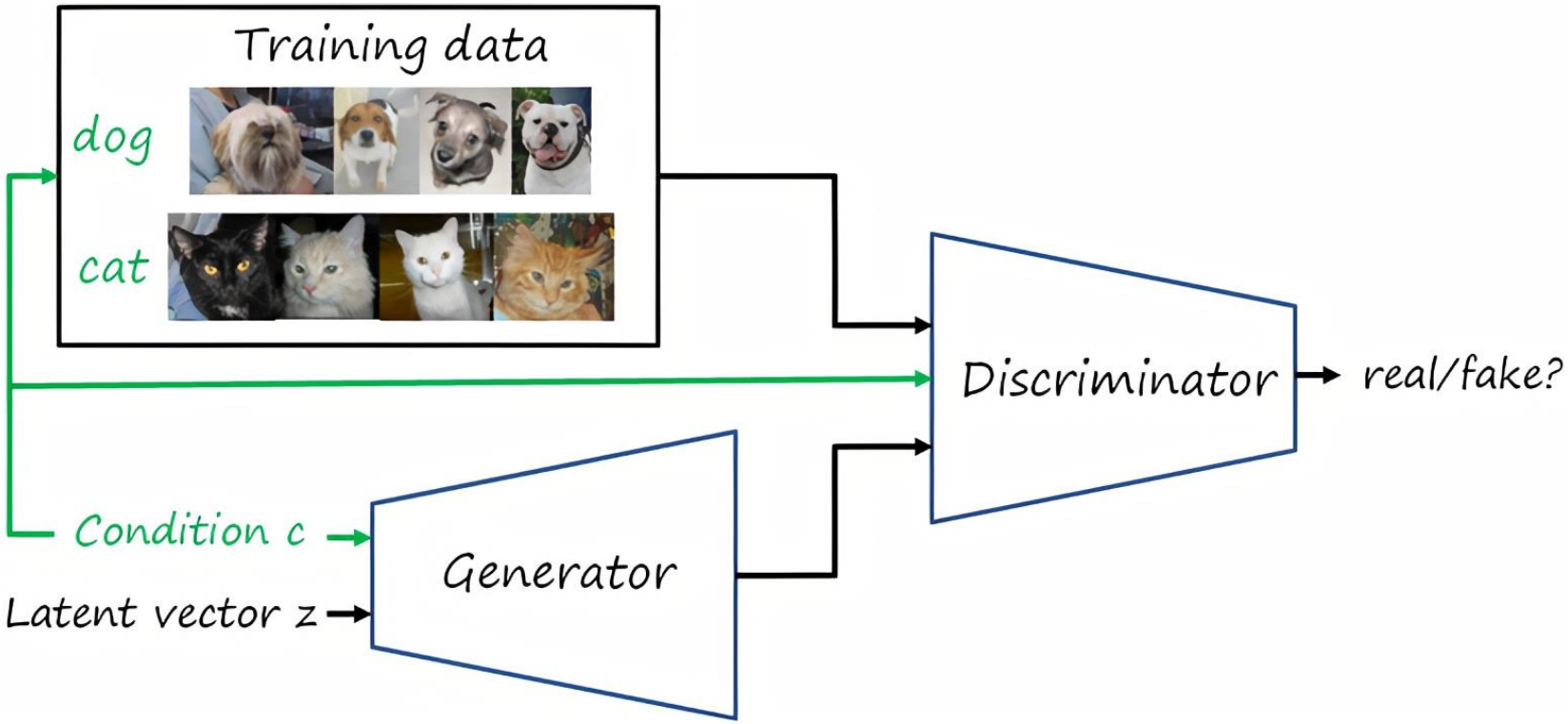
ИСП РАН

Conditional GAN

ИСП РАН

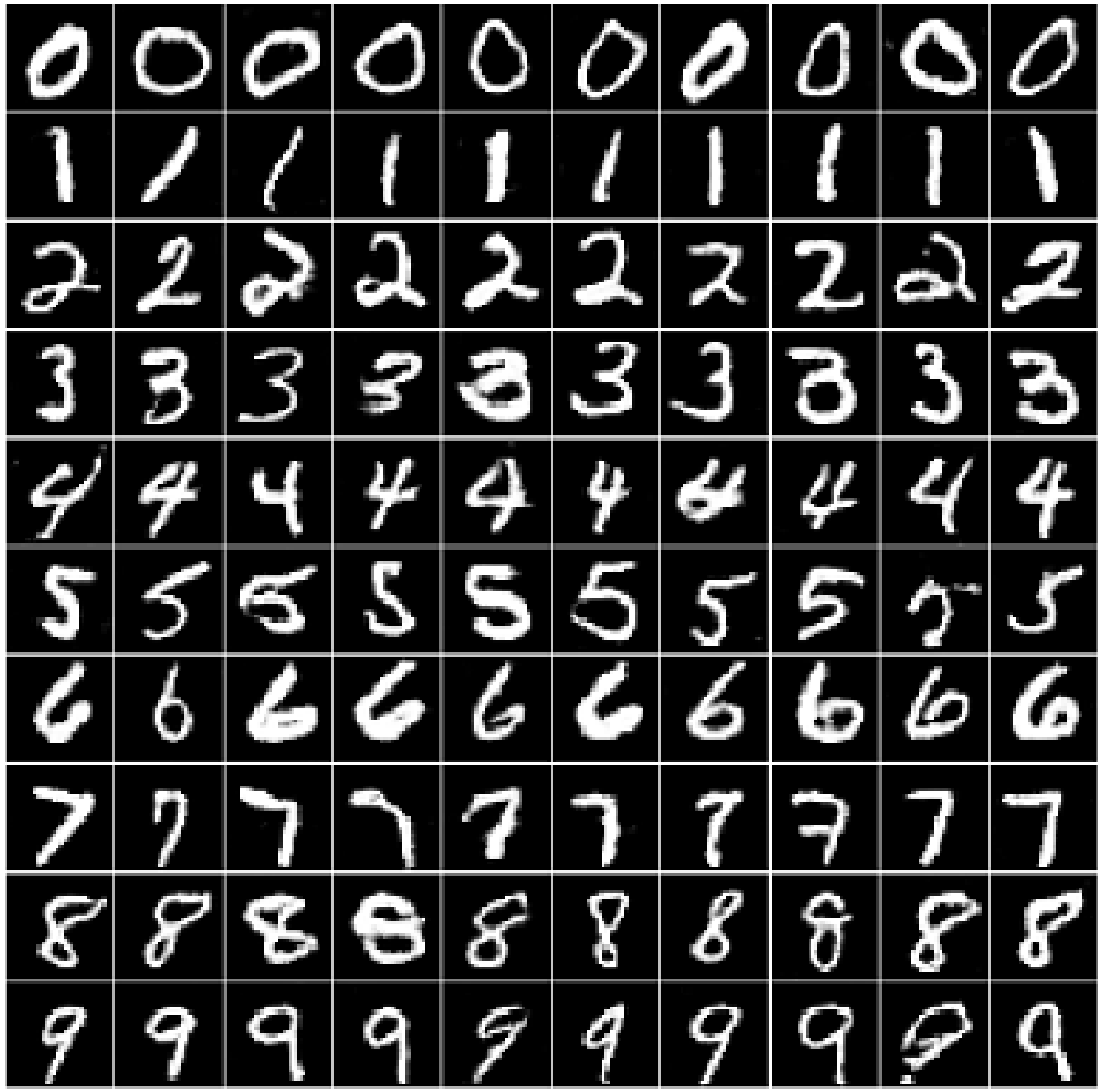
Условная генеративно-состязательная сеть – это обычная GAN, в которой помимо самих изображений используется информация об их классах.

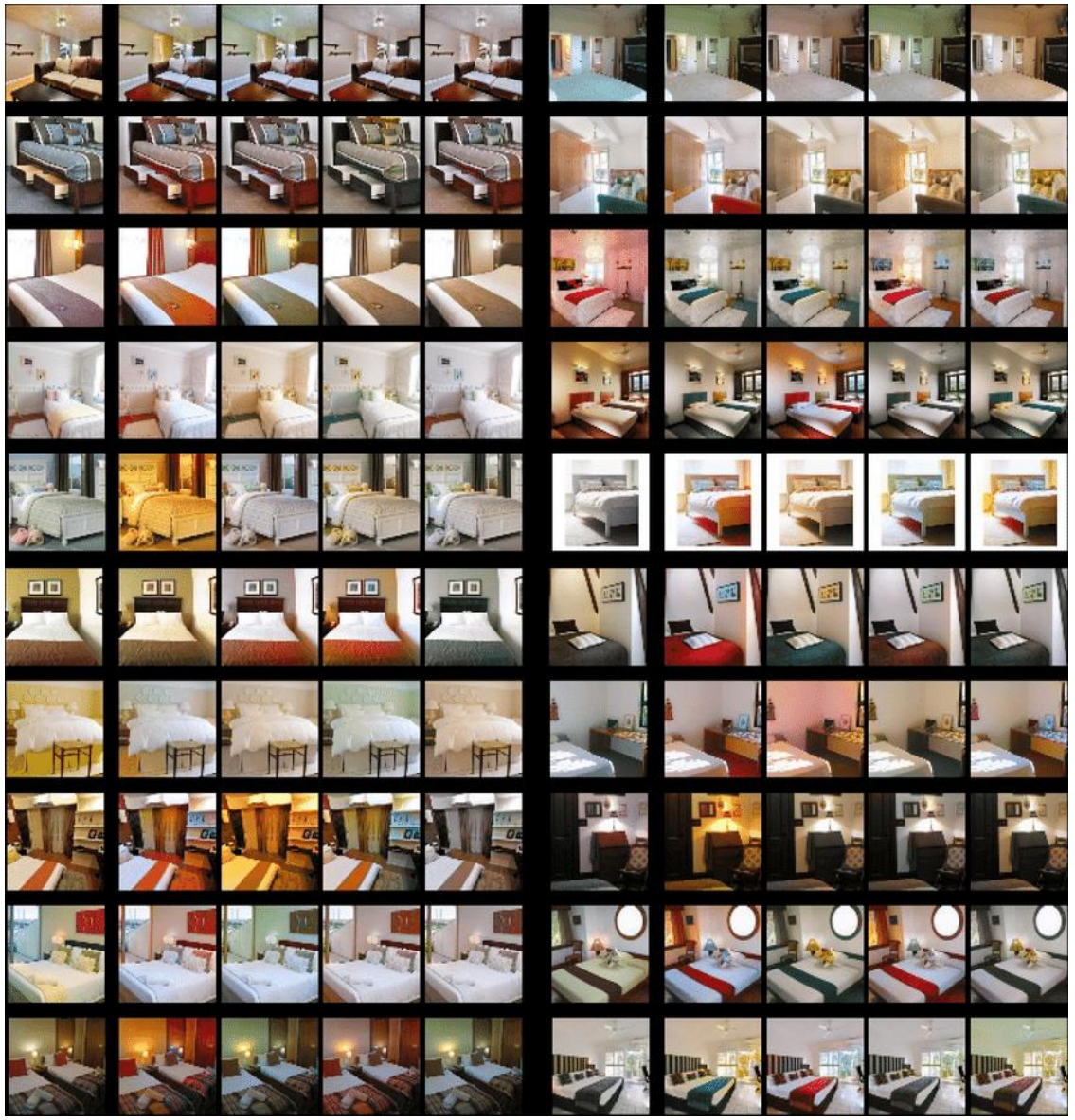
Подавая на вход генератора метку нужного класса, можно контролировать получаемые изображения.



Conditional GAN (примеры)

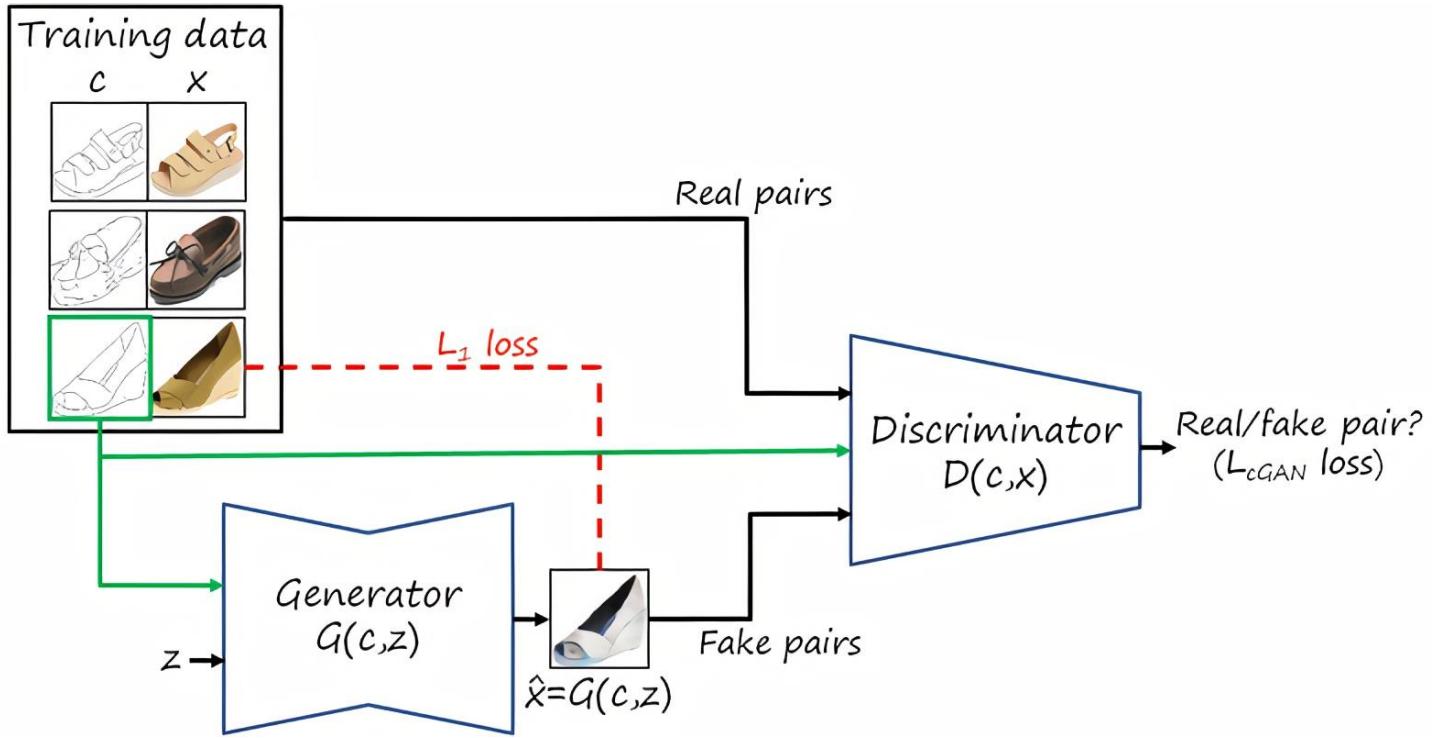
ИСП РАН



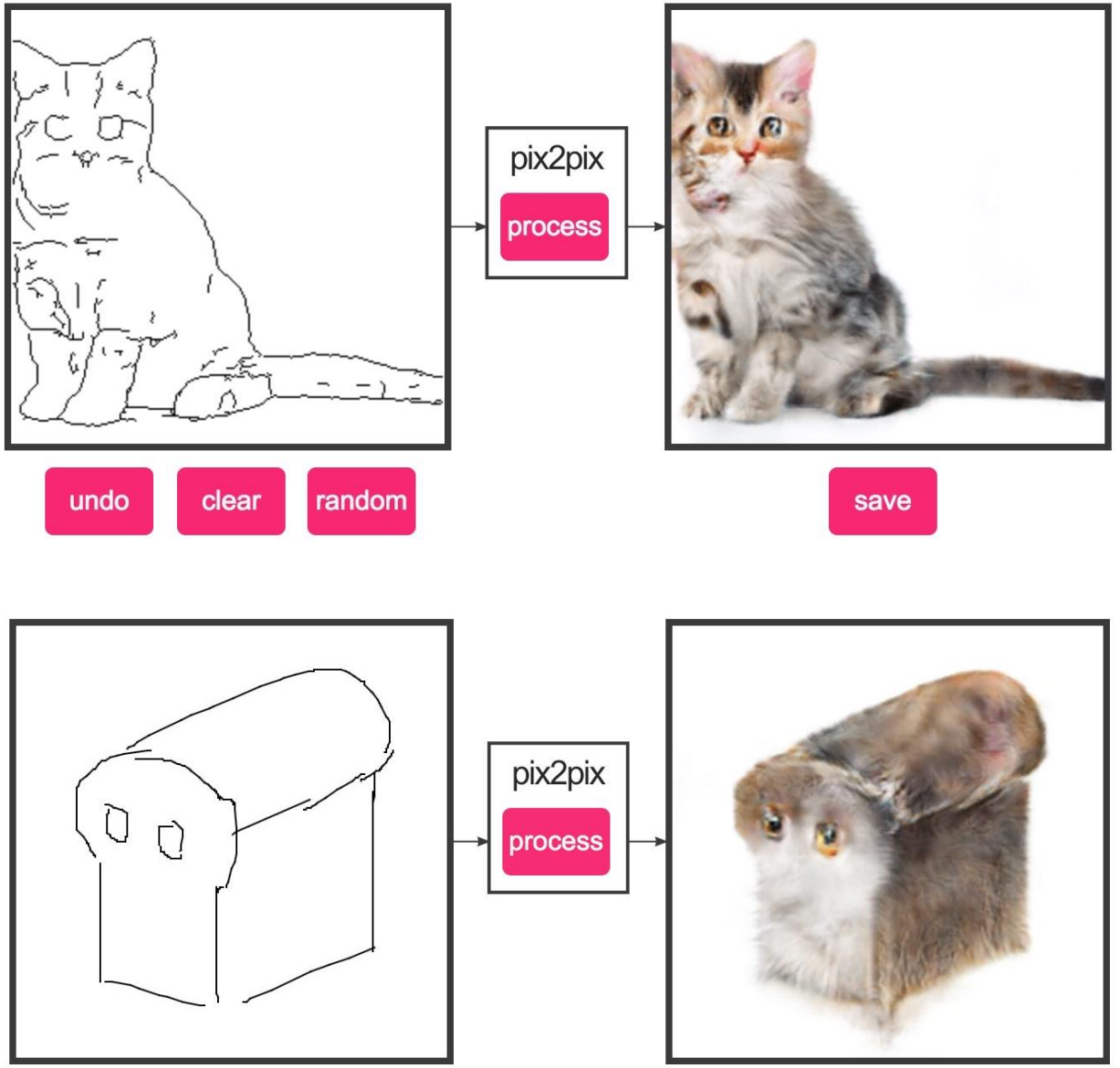


Генератору на вход дается изображение. На ее основе генератор должен сгенерировать картинку на выход.

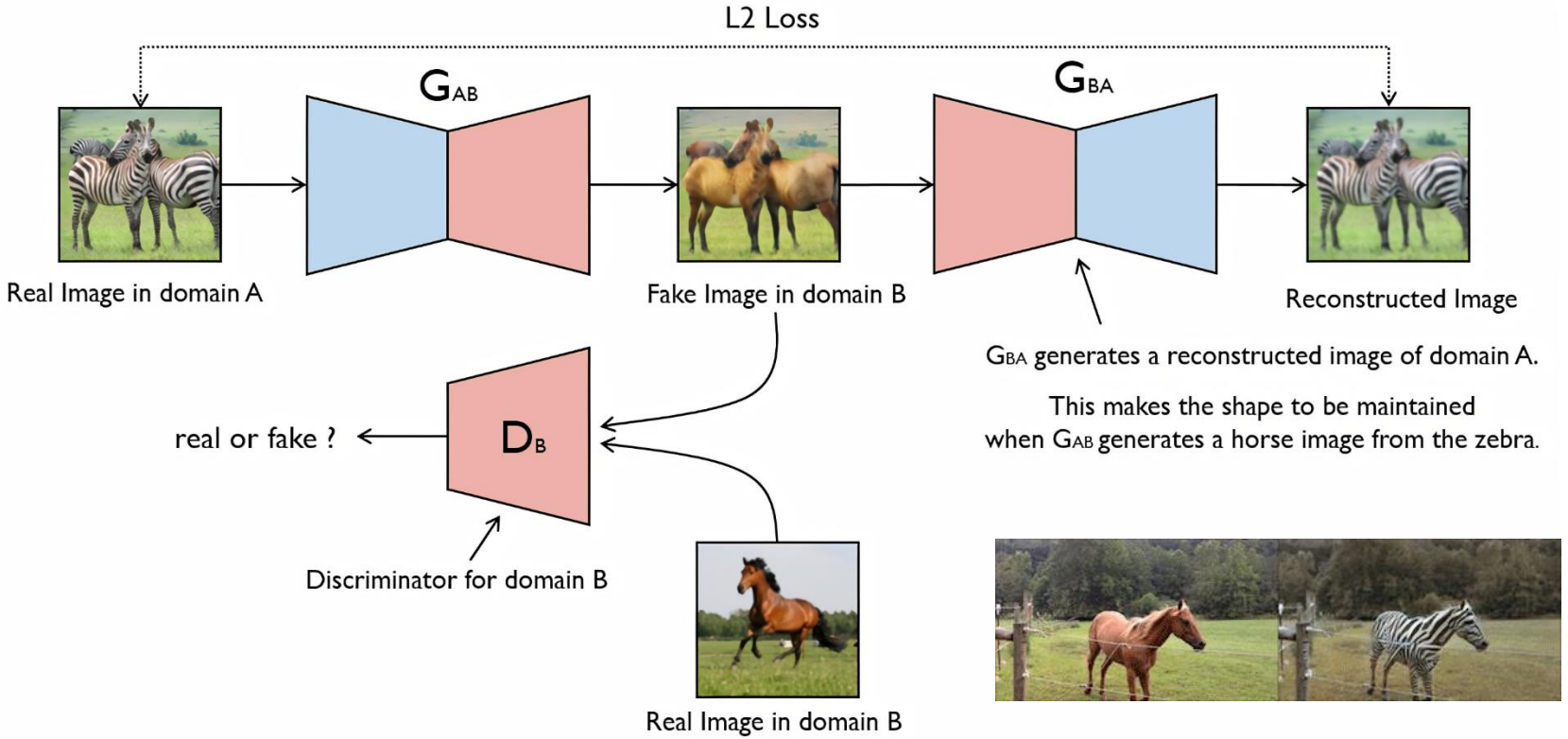
Дискриминатор получает то же изображение и результат работы генератора и определяет, является ли сгенерированная картинка настоящей.



Pix2pix (примеры)

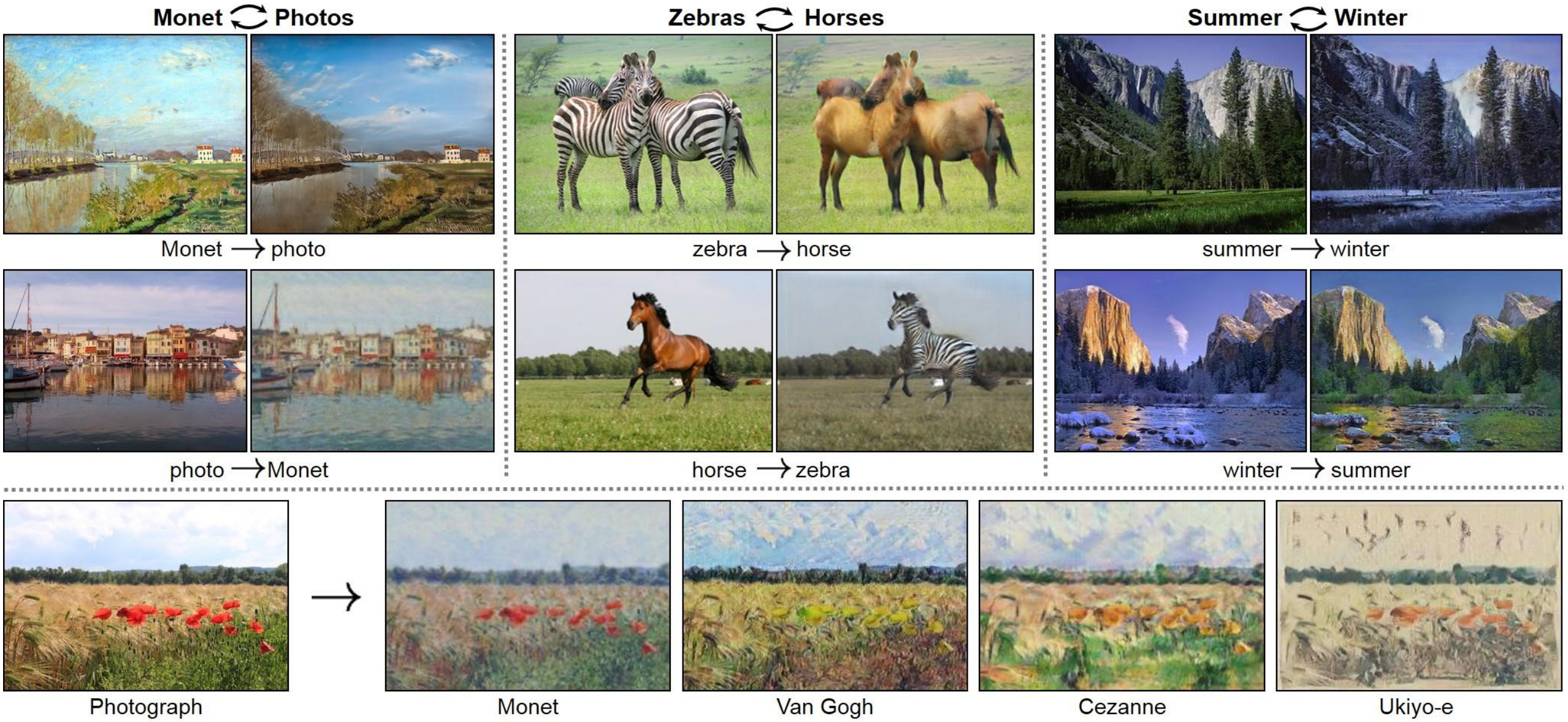


Содержит два генератора-автоэнкодера.
Первый создаёт изображение в другом домене.
Второй из полученного генерирует исходное изображение.
Дискриминатор проверяет, похоже ли изображение, созданное первым генератором на настоящее, или нет.



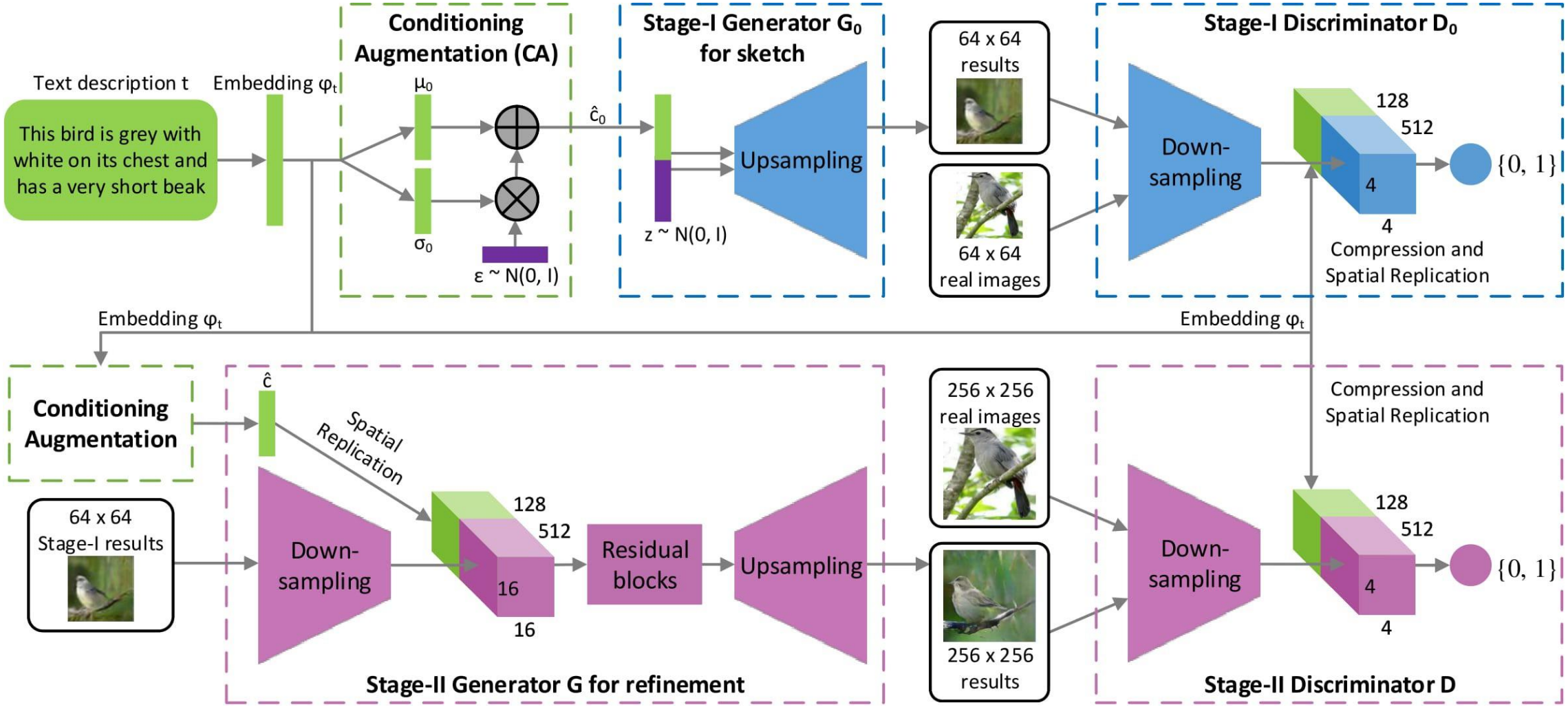
Cycle GAN (примеры)

ИСП РАН



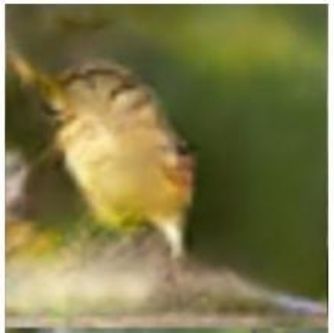
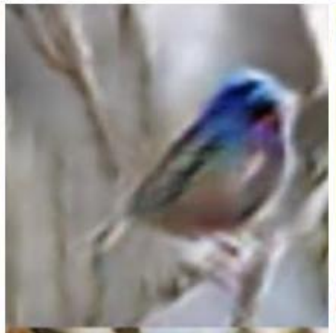
ИСП РАН

Позволяет создавать фотореалистичные изображения по текстовому описанию.

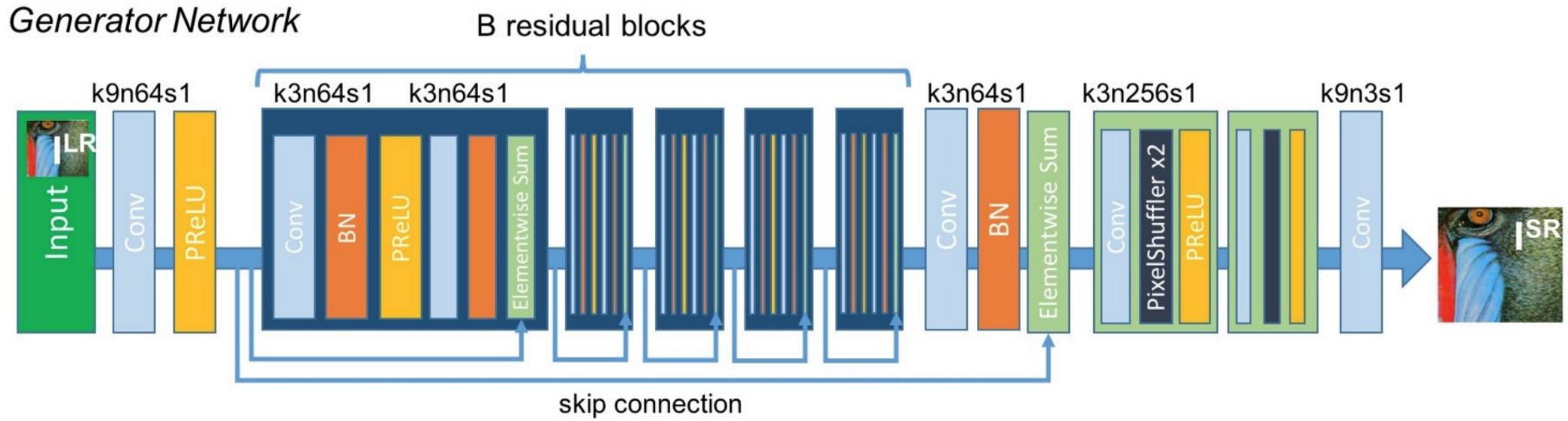


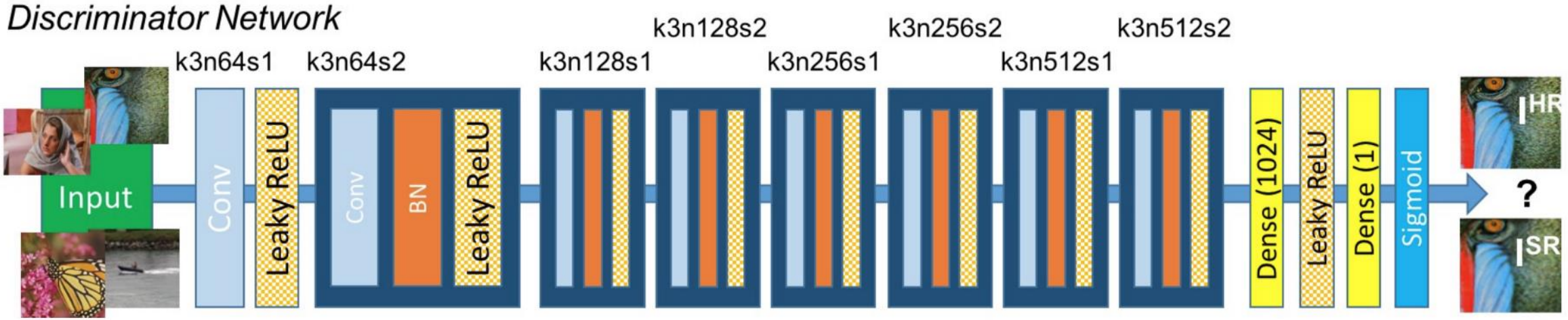
Stack GAN (примеры)

ИСП РАН

Text description	This bird is blue with white and has a very short beak	This bird has wings that are brown and has a yellow belly	A white bird with a black crown and yellow beak	This bird is white, black, and brown in color, with a brown beak	The bird has small beak, with reddish brown crown and gray belly	This is a small, black bird with a white breast and white on the wingbars.	This bird is white black and yellow in color, with a short black beak
Stage-I images							
Stage-II images							

Super resolution GAN





Super resolution GAN (пример)

ИСП РАН

bicubic
(21.59dB/0.6423)

SRResNet
(23.53dB/0.7832)

SRGAN
(21.15dB/0.6868)

original

Полезные ссылки

ИСП РАН

- GAN — What is Generative Adversarial Networks GAN
- Generative Adversarial Network (GAN) using Keras
- Deep Convolutional Generative Adversarial Network
- How to Train a GAN? Tips and tricks to make GANs work
- Генеративно-состязательная нейросеть. Руководство для новичков
- GAN loss functions и GAN problems
- Pix2pix: Как работает генератор кошечек
- Introduction to CGAN
- Implementing StackGAN using Keras
- GAN — Wasserstein GAN & WGAN-GP
- GAN — Super Resolution GAN (SRGAN)

Вопросы?