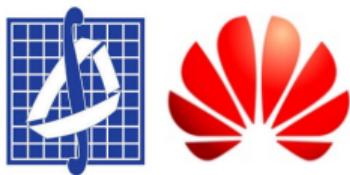


Практические аспекты и тренды задачи улучшения качества изображений

к.ф.-м.н. Иванов И.Е.

кафедра Математической Теории Интеллектуальных Систем

02 ноября 2022 г.



Иванов Илья Евгеньевич

- Закончил кафедру МаТИС и защитил кандидатскую диссертацию под руководством д.ф.-м.н. Д.Н. Бабина
- Закончил школу анализа данных Яндекса
- Сотрудник лаборатории ПТК, м.н.с ^а
- Руководитель группы CameraAI в московском исследовательском центре Huawei
- Координатор образовательной программы SHARE



^а<http://intsys.msu.ru/staff/ivanov/>



Задача улучшения качества изображения

Мотивация

На высококонкурентном рынке смартфонов камера является одним из важнейших конкурентных преимуществ является качество производимых изображений и видео, поэтому компании вкладывают огромные средства в эти исследования.



Задача улучшения качества изображения

Мотивация

На высококонкурентном рынке смартфонов камера является одним из важнейших конкурентных преимуществ является качество производимых изображений и видео, поэтому компании вкладывают огромные средства в эти исследования.

Цель команды CameraAI

Сделать качество работы камеры телефона смартфона на уровне профессиональной камеры



Задача улучшения качества изображения

Мотивация

На высококонкурентном рынке смартфонов камера является одним из важнейших конкурентных преимуществ является качество производимых изображений и видео, поэтому компании вкладывают огромные средства в эти исследования.

Цель команды CameraAI

Сделать качество работы камеры телефона смартфона на уровне профессиональной камеры

Формальная постановки задачи

Дано: изображение, к которому было применена некоторая деградация.

Задача: восстановление исходного изображения

Примеры возможных задач



Базовые деградации

- Шум
- Блюр
- Артефакты сжатия
- Уменьшение пространственной размерности путем интерполяции
- ...



Виды деградаций

Базовые деградации

- Шум
- Блюр
- Артефакты сжатия
- Уменьшение пространственной размерности путем интерполяции
- ...

Как правило реальные деградации являются комбинацией базовых с различными параметрами.



Базовые деградации

- Шум
- Блюр
- Артефакты сжатия
- Уменьшение пространственной размерности путем интерполяции
- ...

Как правило реальные деградации являются комбинацией базовых с различными параметрами.

Различают задачи с известной функцией деградации и неизвестной. Такое разделение рождает принципиально разные классы задач.



- Сверхразрешение с известной функцией деградации (BI, BD, DN)
 - История развития задачи
 - Основные тренды
 - Текущие SoTA решения
- Сверхразрешение для реальных изображений
- Основные тренды
- Проблемы текущих решений

Постановка задачи

Задача сверхразрешения

Увеличение разрешения изображения: $LR \rightarrow HR$

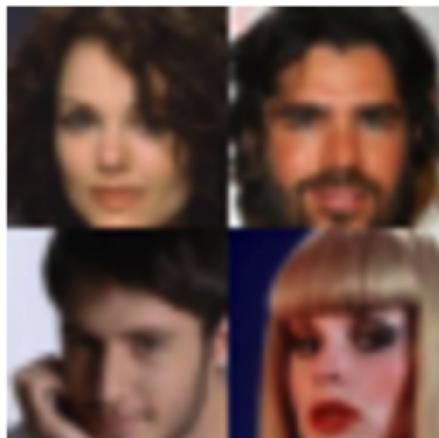


Постановка задачи

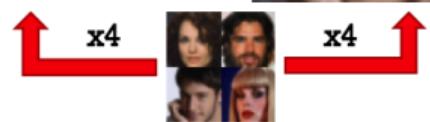
Задача сверхразрешения

Увеличение разрешения изображения: $LR \rightarrow HR$

Бикубическая интерполяция: PSNR = 24.32



Нейросверх-разрешение: PSNR = 27.20



Малоразмерный вход



Постановка задачи

- Задача сверхразрешения — некорректно поставленная задача



Постановка задачи

- Задача сверхразрешения — некорректно поставленная задача
- Для приложений важно не восстановить оригинальное изображение (в большинстве случаев оригинала вообще не существует), а получить хорошее изображение с точки зрения человека

Постановка задачи

- Задача сверхразрешения — некорректно поставленная задача
- Для приложений важно не восстановить оригинальное изображение (в большинстве случаев оригинала вообще не существует), а получить хорошее изображение с точки зрения человека
- Отсюда возникает трудность, как сравнивать различные модели

Приложения задачи сверхразрешения

- Улучшение качества изображения

Приложения задачи сверхразрешения

- Улучшение качества изображения
- Чем лучше софтверные алгоритмы улучшения качества изображений, тем более дешевые сенсоры можно использовать

Приложения задачи сверхразрешения

- Улучшение качества изображения
- Чем лучше софтверные алгоритмы улучшения качества изображений, тем более дешевые сенсоры можно использовать
- Цифровой zoom

Приложения задачи сверхразрешения

- Улучшение качества изображения
- Чем лучше софтверные алгоритмы улучшения качества изображений, тем более дешевые сенсоры можно использовать
- Цифровой zoom
- Сжатие для передачи данных в мессенджерах, TV

Метрики качества: MSE, PSNR

Mean Squared Error

$$MSE(HR, I) = \frac{1}{hw} \sum_i \sum_j (HR[i, j] - I[i, j])^2$$



Mean Squared Error

$$MSE(HR, I) = \frac{1}{hw} \sum_i \sum_j (HR[i, j] - I[i, j])^2$$

- Чем меньше, тем лучше
- Является метрикой



Метрики качества: MSE, PSNR

Mean Squared Error

$$MSE(HR, I) = \frac{1}{hw} \sum_i \sum_j (HR[i, j] - I[i, j])^2$$

- Чем меньше, тем лучше
- Является метрикой

Peak signal-to-noise ratio

$$PSNR = 10 \cdot \log_{10} \left(\frac{\max^2(I)}{MSE} \right) = 20 \cdot \log_{10} \max(I) - 10 \cdot \log_{10} MSE$$

Метрики качества: MSE, PSNR

Mean Squared Error

$$MSE(HR, I) = \frac{1}{hw} \sum_i \sum_j (HR[i, j] - I[i, j])^2$$

- Чем меньше, тем лучше
- Является метрикой

Peak signal-to-noise ratio

$$PSNR = 10 \cdot \log_{10}\left(\frac{\max^2(I)}{MSE}\right) = 20 \cdot \log_{10} \max(I) - 10 \cdot \log_{10} MSE$$

- Чем больше, тем лучше

Метрики качества: MSE, PSNR

Mean Squared Error

$$MSE(HR, I) = \frac{1}{hw} \sum_i \sum_j (HR[i, j] - I[i, j])^2$$

- Чем меньше, тем лучше
- Является метрикой

Peak signal-to-noise ratio

$$PSNR = 10 \cdot \log_{10}\left(\frac{\max^2(I)}{MSE}\right) = 20 \cdot \log_{10} \max(I) - 10 \cdot \log_{10} MSE$$

- Чем больше, тем лучше
- Измеряется в децибелах (dB)

Метрики качества: MSE, PSNR

Mean Squared Error

$$MSE(HR, I) = \frac{1}{hw} \sum_i \sum_j (HR[i, j] - I[i, j])^2$$

- Чем меньше, тем лучше
- Является метрикой

Peak signal-to-noise ratio

$$PSNR = 10 \cdot \log_{10}\left(\frac{\max^2(I)}{MSE}\right) = 20 \cdot \log_{10} \max(I) - 10 \cdot \log_{10} MSE$$

- Чем больше, тем лучше
- Измеряется в децибелах (dB)
- Метрикой не является

Достоинства

- Легко считать
- Понятная интерпретация
- Можно в прямую оптимизировать



Достоинства и недостатки MSE, PSNR

Достоинства

- Легко считать
- Понятная интерпретация
- Можно в прямую оптимизировать

Недостатки

- Попиксельная ошибка не учитывает структуру, которая есть в изображении
- Среднеквадратичная ошибка на пикселях не отражает человеческого восприятия



Вопрос: на каком изображении PSNR выше?



(a)



(b)



(c)



(d)



(e)



(f)

Метрики качества: SSIM¹

Идея

Найти и оценить структурную разницу сигналов

¹https://en.wikipedia.org/wiki/Structural_similarity



Метрики качества: SSIM¹

Идея

Найти и оценить структурную разницу сигналов

Structural similarity

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)},$$

где

- μ_x, μ_y — среднее x и y соответственно
- σ_x^2, σ_y^2 — дисперсия x и y соответственно
- σ_{xy} — ковариация x и y
- $c_1 = (k_1 L)^2, c_2 = (k_2 L)^2$, где L — динамический диапазон пикселей, k_1 и k_2 — параметры (по умолчанию $k_1 = 0.01, k_2 = 0.03$)

¹https://en.wikipedia.org/wiki/Structural_similarity

Рассмотрим другое представление SSIM:

$$SSIM(x, y) = [l(x, y)^\alpha c(x, y)^\beta s(x, y)^\gamma],$$

где

- Яркость (luminance) $l(x, y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}$
- Контрастность (contrast) $c(x, y) = \frac{2\sigma_x\sigma_y + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}$
- Структурность (structure) $s(x, y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3}$, где $c_3 = \frac{c_2}{2}$
- $\alpha = \beta = \gamma = 1$

Вопрос

Является ли SSIM метрикой?



SSIM

1.000



(a)

0.9168



(b)

0.9900



(c)



(d)



(e)



(f)

0.6949

0.7052

0.7748

Метрики качества: MOS

Mean opinion score

Средния оценка специально обученных оценщиков изображений



Mean opinion score

Средния оценка специально обученных оценщиков изображений

Недостатки

- Субъективный критерий качества
- Невоспроизводимо
- Невозможно тестирование автоматически
- Тестирование стоит денег



Метрики качества: MOS

Mean opinion score

Средния оценка специально обученных оценщиков изображений

Недостатки

- Субъективный критерий качества
- Невоспроизводимо
- Невозможно тестирование автоматически
- Тестирование стоит денег

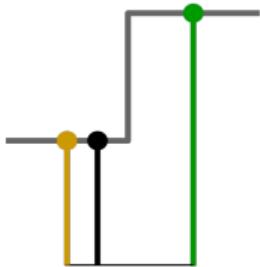
Достоинства

- В целом MOS отражает именно то, чего мы ожидаем от сверхразрешения

Методы сверхразрешения

- Интерполяция
- SRCNN
- VDSR
- ESPCN
- SRGAN

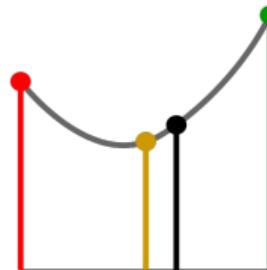
Интерполяция



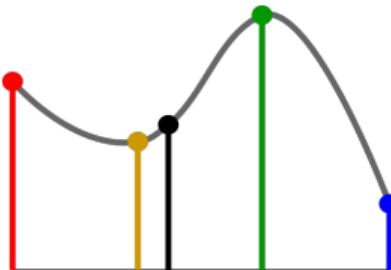
Одномерная по ближайшему соседу



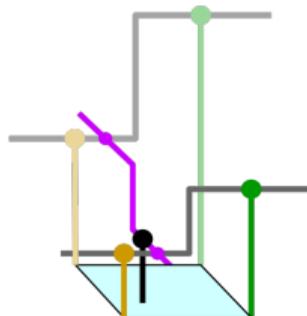
Линейная



Квадратичная



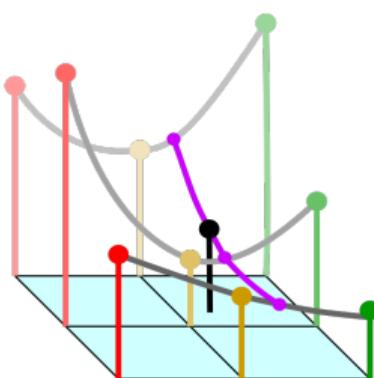
Кубическая



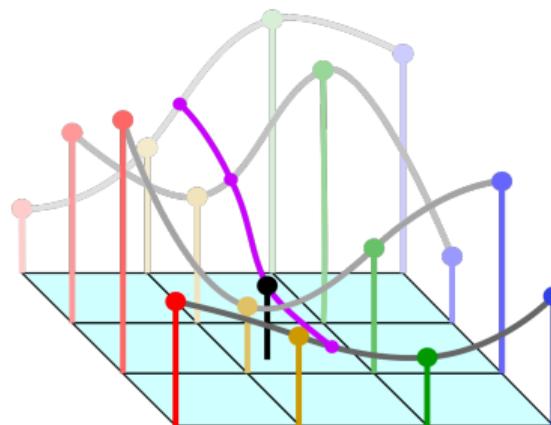
Двумерная по ближайшему соседу



Билинейная



Биквадратная



Бикубическая

Бикубическая интерполяция ²

$$p(x, y) = \sum_{i=0}^3 \sum_{j=0}^3 a_{ij} x^i y^j$$

- Обобщение кубической интерполяции на случай функции двух переменных
- Для вычисления функции требуется 16 точек
- Даёт более качественные изображения по сравнению с билинейной интерполяцией и методом ближайшего соседа

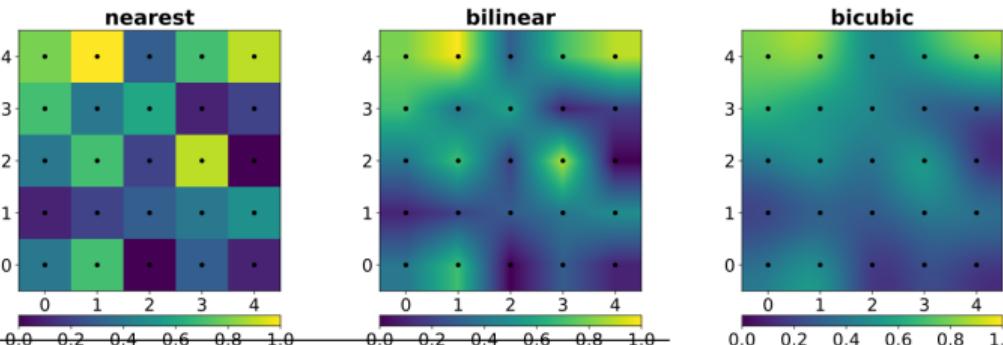
²images from https://en.wikipedia.org/wiki/Bicubic_interpolation



Бикубическая интерполяция²

$$p(x, y) = \sum_{i=0}^3 \sum_{j=0}^3 a_{ij} x^i y^j$$

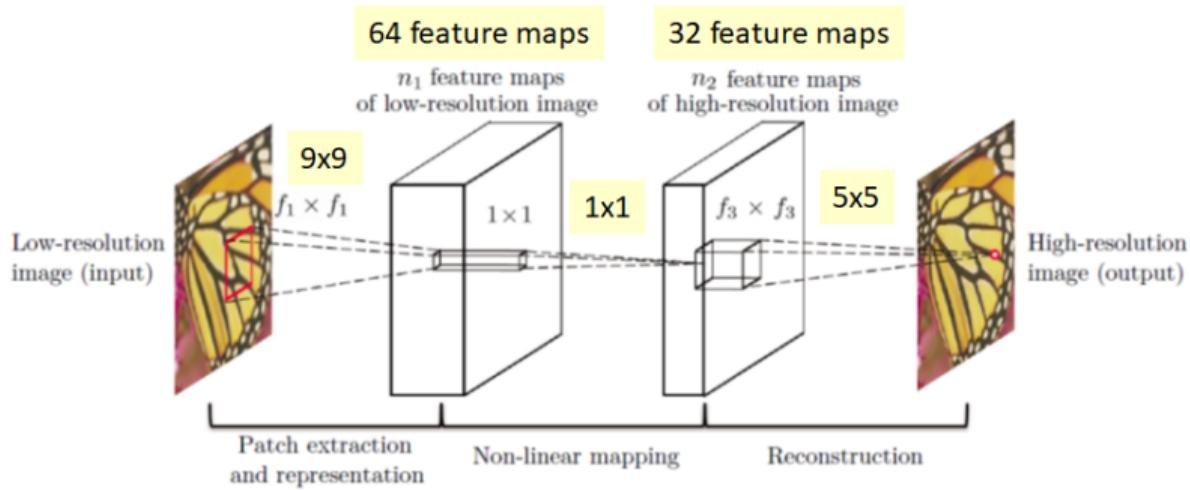
- Обобщение кубической интерполяции на случай функции двух переменных
- Для вычисления функции требуется 16 точек
- Даёт более качественные изображения по сравнению с билинейной интерполяцией и методом ближайшего соседа



²images from https://en.wikipedia.org/wiki/Bicubic_interpolation



- Качество нужного разрешения достигается путём применения бикубической интерполяции к изображению низкого разрешения
 - Далее улучшение качества изображения достигается применением трёхслойной свёрточной нейронной сети



Детали обучения SRCNN

- В качестве лосс-функции используется MSE

⁴<https://en.wikipedia.org/wiki/YCbCr>



- В качестве лосс-функции используется MSE
- Для разных слоёв используется разный шаг градиентного спуска

⁴<https://en.wikipedia.org/wiki/YCbCr>



- В качестве лосс-функции используется MSE
- Для разных слоёв используется разный шаг градиентного спуска
- Для получения лучших результатов используется Y канал из цветового пространства YCbCr⁴. К оставшимся каналам применяется бикубическая интерполяция

⁴<https://en.wikipedia.org/wiki/YCbCr>



- В качестве лосс-функции используется MSE
- Для разных слоёв используется разный шаг градиентного спуска
- Для получения лучших результатов используется Y канал из цветового пространства YCbCr⁴. К оставшимся каналам применяется бикубическая интерполяция
- Для подготовки LR изображения применялся гауссовский фильтр для размытия, после чего выкидывались лишние пиксели

⁴<https://en.wikipedia.org/wiki/YCbCr>



SRCNN: Результаты

Eval. Mat	Scale	Bicubic	SC [50]	NE+LLE [4]	KK [25]	ANR [41]	A+ [41]	SRCNN
PSNR	2	33.66	-	35.77	36.20	35.83	36.54	36.66
	3	30.39	31.42	31.84	32.28	31.92	32.59	32.75
	4	28.42	-	29.61	30.03	29.69	30.28	30.49
SSIM	2	0.9299	-	0.9490	0.9511	0.9499	0.9544	0.9542
	3	0.8682	0.8821	0.8956	0.9033	0.8968	0.9088	0.9090
	4	0.8104	-	0.8402	0.8541	0.8419	0.8603	0.8628
IFC	2	6.10	-	7.84	6.87	8.09	8.48	8.05
	3	3.52	3.16	4.40	4.14	4.52	4.84	4.58
	4	2.35	-	2.94	2.81	3.02	3.26	3.01
NQM	2	36.73	-	42.90	39.49	43.28	44.58	41.13
	3	27.54	27.29	32.77	32.10	33.10	34.48	33.21
	4	21.42	-	25.56	24.99	25.72	26.97	25.96
WPSNR	2	50.06	-	58.45	57.15	58.61	60.06	59.49
	3	41.65	43.64	45.81	46.22	46.02	47.17	47.10
	4	37.21	-	39.85	40.40	40.01	41.03	41.13
MSSSIM	2	0.9915	-	0.9953	0.9953	0.9954	0.9960	0.9959
	3	0.9754	0.9797	0.9841	0.9853	0.9844	0.9867	0.9866
	4	0.9516	-	0.9666	0.9695	0.9672	0.9720	0.9725



SRCNN: Визуальные результаты



Original / PSNR



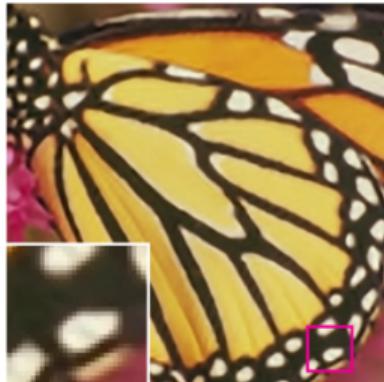
Bicubic / 24.04 dB



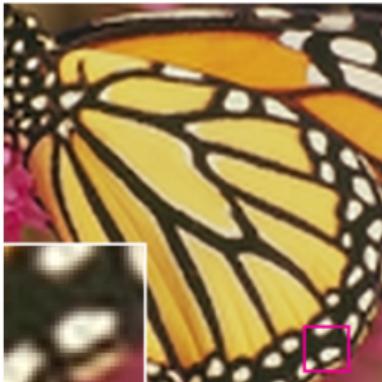
SC / 25.58 dB



NE+LLE / 25.75 dB



KK / 27.31 dB



ANR / 25.90 dB



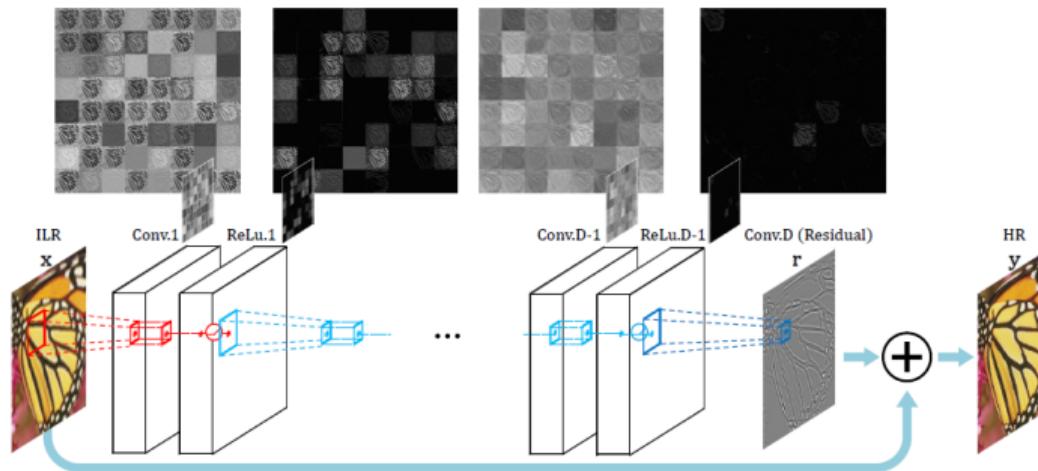
A+ / 27.24 dB



SRCNN / 27.95 dB

VDSR⁵

- Авторы были вдохновлены работой VGG и сделали глубокую сеть для задачи сверхразрешения



⁵Kim, Jiwon, Jung Kwon Lee, and Kyoung Mu Lee. "Accurate image super-resolution using very deep convolutional networks." 2016.

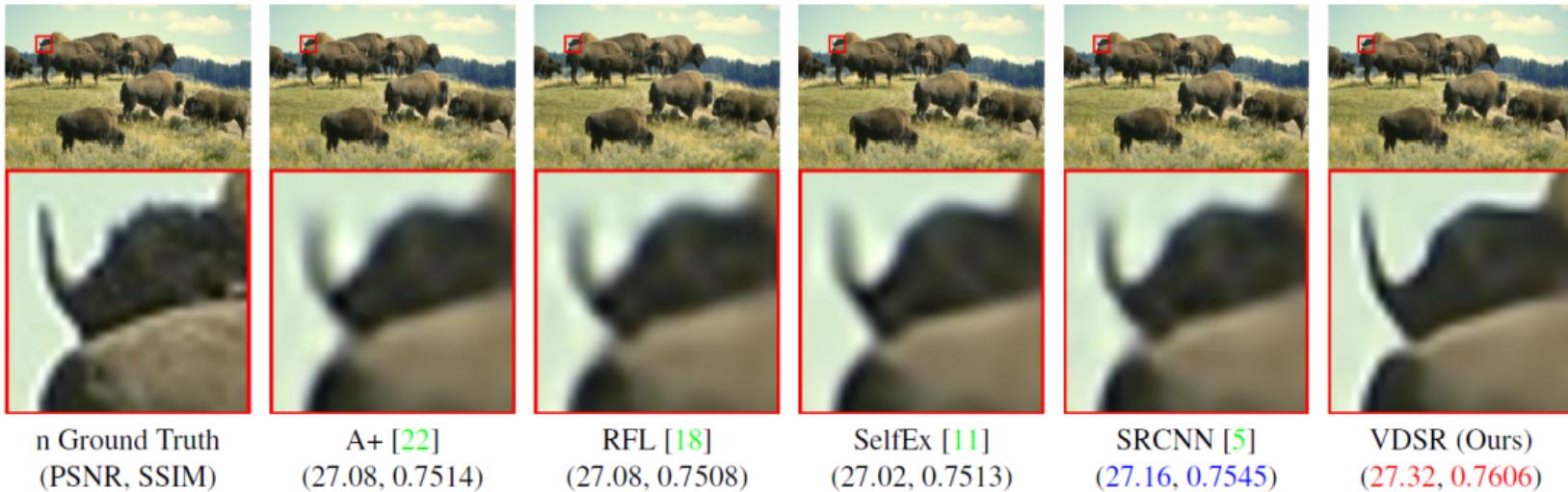


VDSR: Результаты

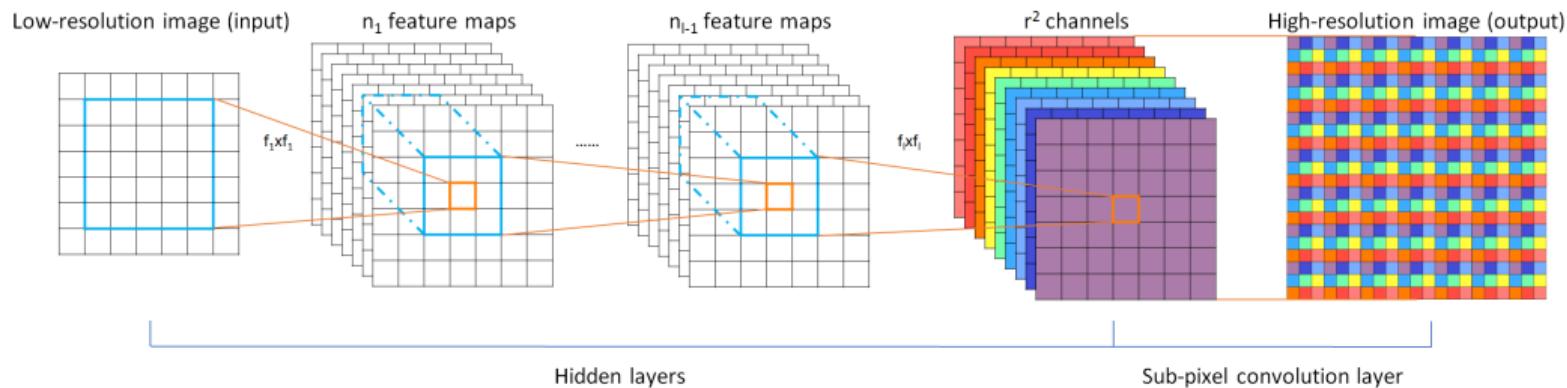
Dataset	Scale	Bicubic PSNR/SSIM/time	A+ [22] PSNR/SSIM/time	RFL [18] PSNR/SSIM/time	SelfEx [11] PSNR/SSIM/time	SRCNN [5] PSNR/SSIM/time	VDSR (Ours) PSNR/SSIM/time
Set5	×2	33.66/0.9299/0.00	36.54/ 0.9544 / 0.58	36.54/0.9537/0.63	36.49/0.9537/45.78	36.66 /0.9542/2.19	37.53 / 0.9587 / 0.13
	×3	30.39/0.8682/0.00	32.58/0.9088/ 0.32	32.43/0.9057/0.49	32.58/ 0.9093 /33.44	32.75 /0.9090/2.23	33.66 / 0.9213 / 0.13
	×4	28.42/0.8104/0.00	30.28/0.8603/ 0.24	30.14/0.8548/0.38	30.31/0.8619/29.18	30.48 / 0.8628 /2.19	31.35 / 0.8838 / 0.12
Set14	×2	30.24/0.8688/0.00	32.28/0.9056/ 0.86	32.26/0.9040/1.13	32.22/0.9034/105.00	32.42 / 0.9063 /4.32	33.03 / 0.9124 / 0.25
	×3	27.55/0.7742/0.00	29.13/0.8188/ 0.56	29.05/0.8164/0.85	29.16/0.8196/74.69	29.28 / 0.8209 /4.40	29.77 / 0.8314 / 0.26
	×4	26.00/0.7027/0.00	27.32/0.7491/ 0.38	27.24/0.7451/0.65	27.40/ 0.7518 /65.08	27.49 /0.7503/4.39	28.01 / 0.7674 / 0.25
B100	×2	29.56/0.8431/0.00	31.21/0.8863/ 0.59	31.16/0.8840/0.80	31.18/0.8855/60.09	31.36 / 0.8879 /2.51	31.90 / 0.8960 / 0.16
	×3	27.21/0.7385/0.00	28.29/0.7835/ 0.33	28.22/0.7806/0.62	28.29/0.7840/40.01	28.41 / 0.7863 /2.58	28.82 / 0.7976 / 0.21
	×4	25.96/0.6675/0.00	26.82/0.7087/ 0.26	26.75/0.7054/0.48	26.84/ 0.7106 /35.87	26.90 /0.7101/2.51	27.29 / 0.7251 / 0.21
Urban100	×2	26.88/0.8403/0.00	29.20/0.8938/ 2.96	29.11/0.8904/3.62	29.54 / 0.8967 /663.98	29.50/0.8946/22.12	30.76 / 0.9140 / 0.98
	×3	24.46/0.7349/0.00	26.03/0.7973/ 1.67	25.86/0.7900/2.48	26.44 / 0.8088 /473.60	26.24/0.7989/19.35	27.14 / 0.8279 / 1.08
	×4	23.14/0.6577/0.00	24.32/0.7183/ 1.21	24.19/0.7096/1.88	24.79 / 0.7374 /394.40	24.52/0.7221/18.46	25.18 / 0.7524 / 1.06



VDSR: Визуальные результаты



- Для увеличения размера изображения используется специальный субпиксельный слой



⁶Shi, Wenzhe, et al. "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network." 2016



ESPCN: Результаты

Dataset	Scale	SRCNN (91)	ESPCN (91 <i>relu</i>)	ESPCN (91)	SRCNN (ImageNet)	ESPCN (ImageNet <i>relu</i>)
Set5	3	32.39	32.39	32.55	32.52	33.00
Set14	3	29.00	28.97	29.08	29.14	29.42
BSD300	3	28.21	28.20	28.26	28.29	28.52
BSD500	3	28.28	28.27	28.34	28.37	28.62
SuperTexture	3	26.37	26.38	26.42	26.41	26.69
Average	3	27.76	27.76	27.82	27.83	28.09



ESPCN: Результаты



(f) Comic Original



(g) Bicubic / 23.12db



(h) SRCNN [7] / 24.56db



(i) TNRD [3] / 24.68db



(j) ESPCN / **24.82db**



SRGAN⁷

- Основная идея — использовать GAN для задачи сверхразрешения
- Adversarial loss не восстанавливает изображение, а делает его более естественным
- Content loss vs Perceptual loss

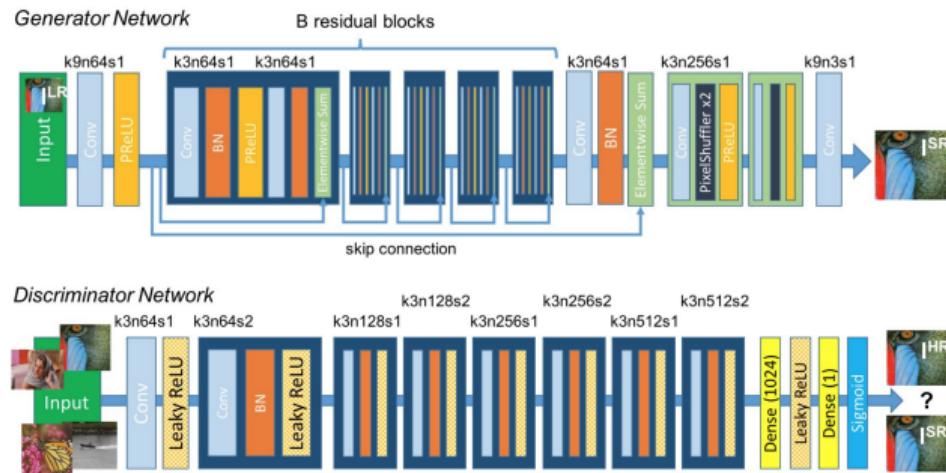
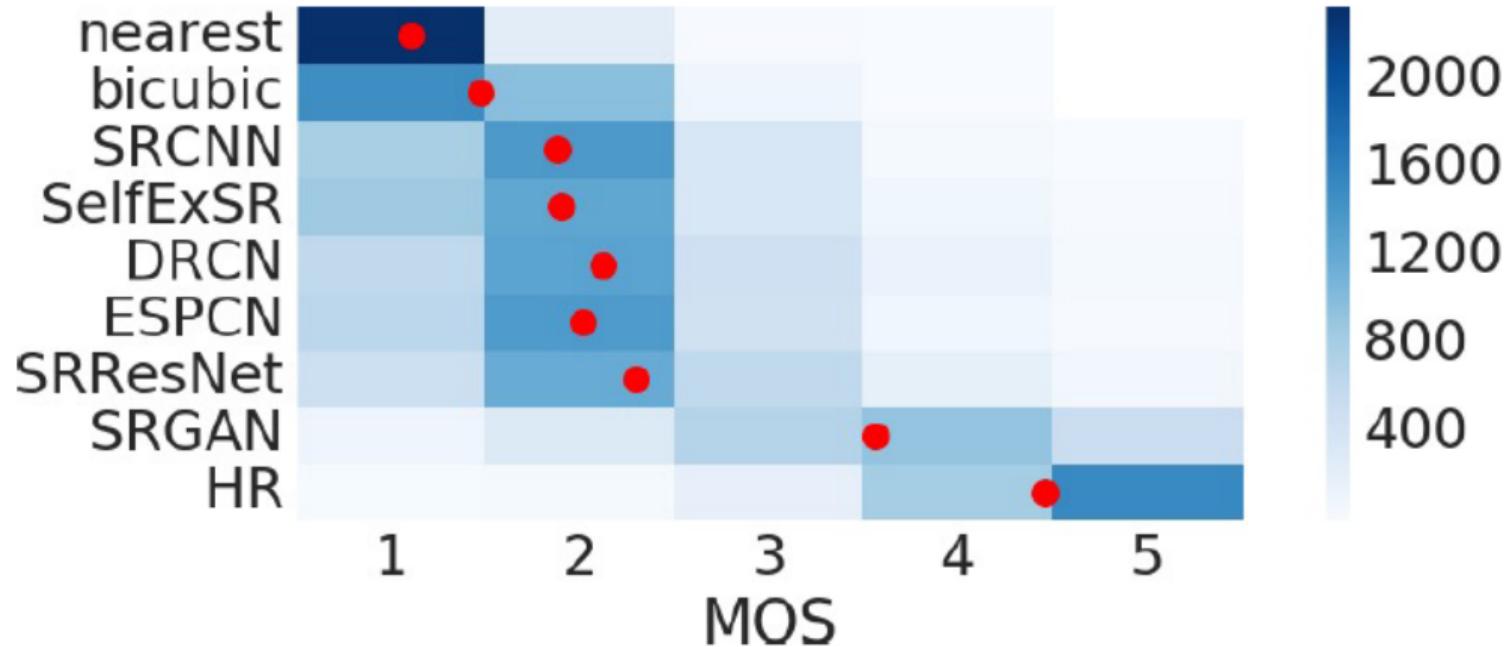


Figure 4: Architecture of Generator and Discriminator Network with corresponding kernel size (k), number of feature maps (n) and stride (s) indicated for each convolutional layer.

⁷Ledig, Christian, et al. "Photo-realistic single image super-resolution using a generative adversarial network" 
2017

SRGAN: Результаты



SRGAN: Визуальные результаты

bicubic
(21.59dB/0.6423)



SRResNet
(23.53dB/0.7832)



SRGAN
(21.15dB/0.6868)



original



Residual Training: EDSR

Идея

После выхода архитектуры ResNet стало возможно обучать более глубокие нейронные сети и для задачи сверхразрешения.

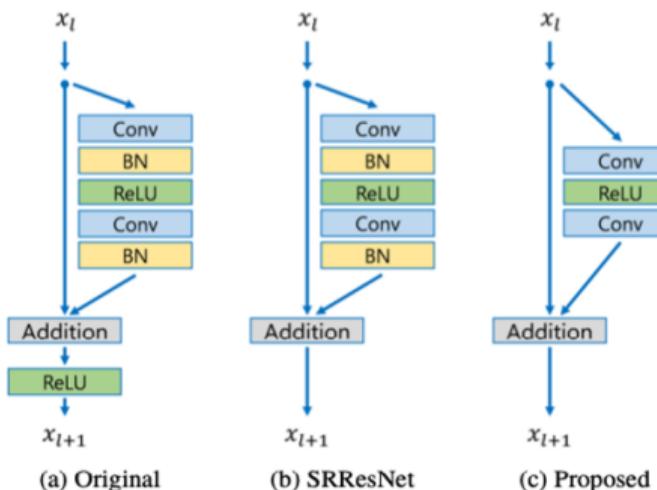


Figure 2: Comparison of residual blocks in original ResNet, SRResNet, and ours.

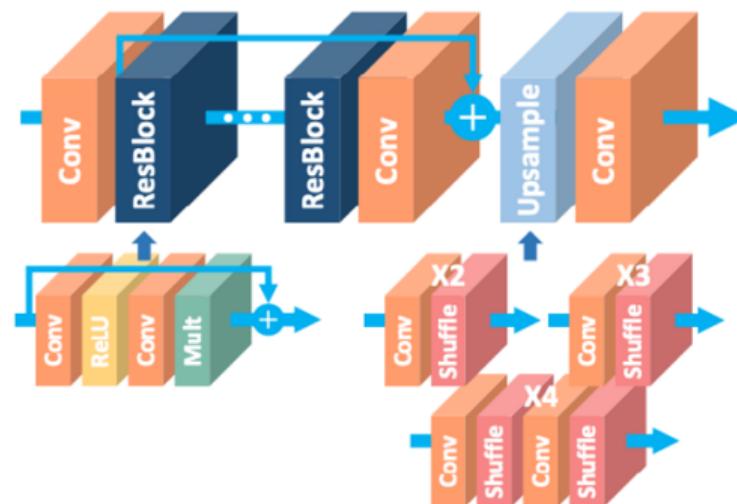


Figure 3: The architecture of the proposed single-scale SR network (EDSR).

Проблема

В большинстве современных решений по сверхразрешению добавление батч нормализации приводит к деградации качества.

Причина

Авторы обнаружили, что после слоёв нормализации дисперсия сигнала становится меньше. При уменьшении дисперсии сети становится сложнее выделять границы, что ведет к снижению качества.

Решение

Авторы предложили блок AdaDM, который увеличивает дисперсию обратно.

⁸<https://arxiv.org/pdf/2111.13905.pdf>

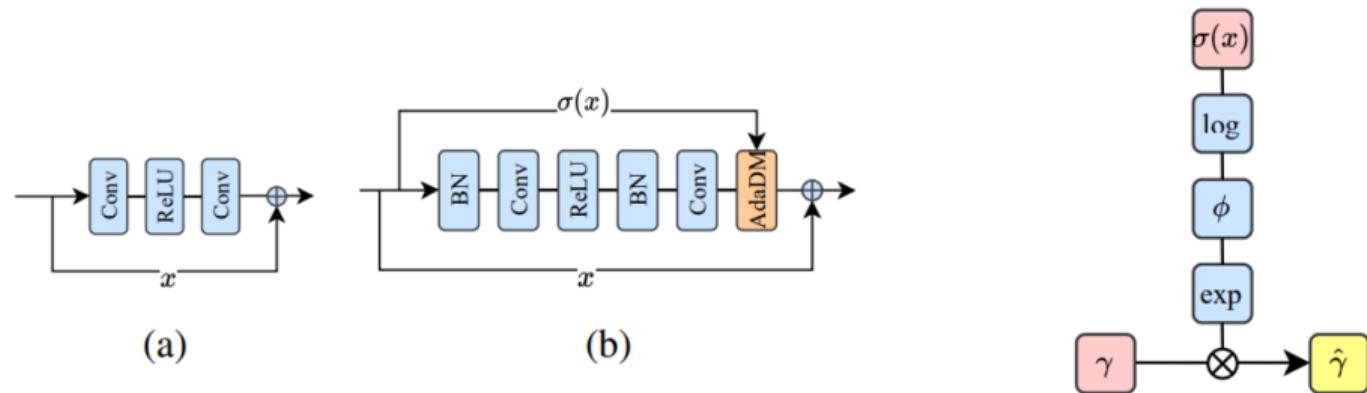
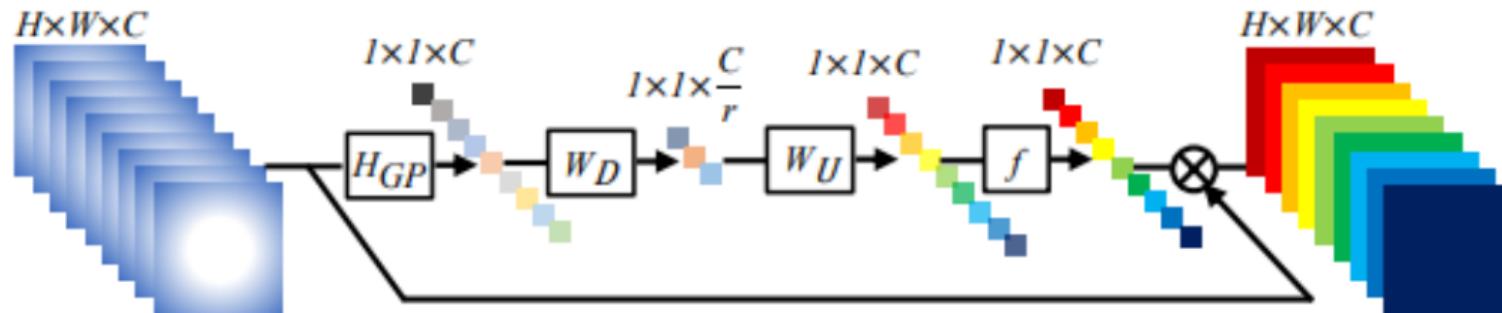


Figure 5: Schema of AdaDM.

⁹<https://arxiv.org/pdf/2111.13905.pdf>

Мотивация

Механизм внимания помогает существенно улучшить качество моделей сверхразрешения. Внимание помогает усиливать важную информацию и игнорировать нерелевантную.

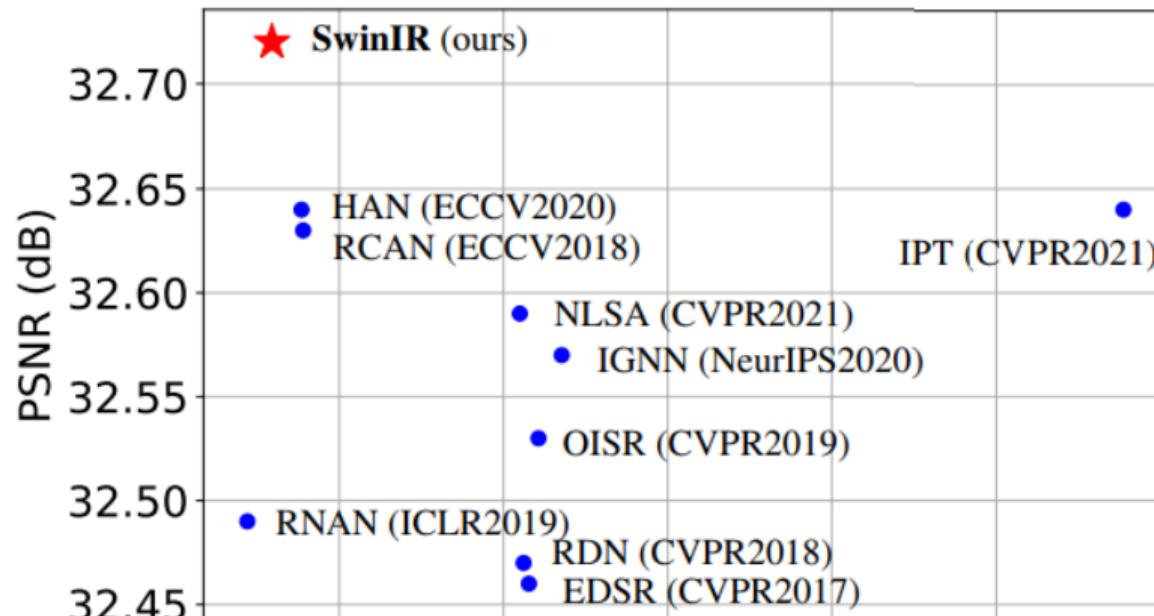


Пример показательного механизма внимания ¹⁰

¹⁰<https://arxiv.org/abs/1807.02758>

Идея

Применение self-attention механизма позволяет достичь SoTA результатов.



Результаты SwinIR

Method	Scale	Training Dataset	Set5 [3]		Set14 [87]		BSD100 [58]		Urban100 [34]		Manga109 [60]	
			PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
RCAN [95]	×2	DIV2K	38.27	0.9614	34.12	0.9216	32.41	0.9027	33.34	0.9384	39.44	0.9786
SAN [15]	×2	DIV2K	38.31	0.9620	34.07	0.9213	32.42	0.9028	33.10	0.9370	39.32	0.9792
IGNN [100]	×2	DIV2K	38.24	0.9613	34.07	0.9217	32.41	0.9025	33.23	0.9383	39.35	0.9786
HAN [63]	×2	DIV2K	38.27	0.9614	34.16	0.9217	32.41	0.9027	33.35	0.9385	39.46	0.9785
NLSA [61]	×2	DIV2K	38.34	0.9618	34.08	0.9231	32.43	0.9027	33.42	0.9394	39.59	0.9789
SwinIR (Ours)	×2	DIV2K	38.35	0.9620	34.14	0.9227	32.44	0.9030	33.40	0.9393	39.60	0.9792
SwinIR+ (Ours)	×2	DIV2K	38.38	0.9621	34.24	0.9233	32.47	0.9032	33.51	0.9401	39.70	0.9794
DBPN [31]	×2	DIV2K+Flickr2K	38.09	0.9600	33.85	0.9190	32.27	0.9000	32.55	0.9324	38.89	0.9775
IPT [9]	×2	ImageNet	38.37	-	34.43	-	32.48	-	33.76	-	-	-
SwinIR (Ours)	×2	DIV2K+Flickr2K	38.42	0.9623	34.46	0.9250	32.53	0.9041	33.81	0.9427	39.92	0.9797
SwinIR+ (Ours)	×2	DIV2K+Flickr2K	38.46	0.9624	34.61	0.9260	32.55	0.9043	33.95	0.9433	40.02	0.9800

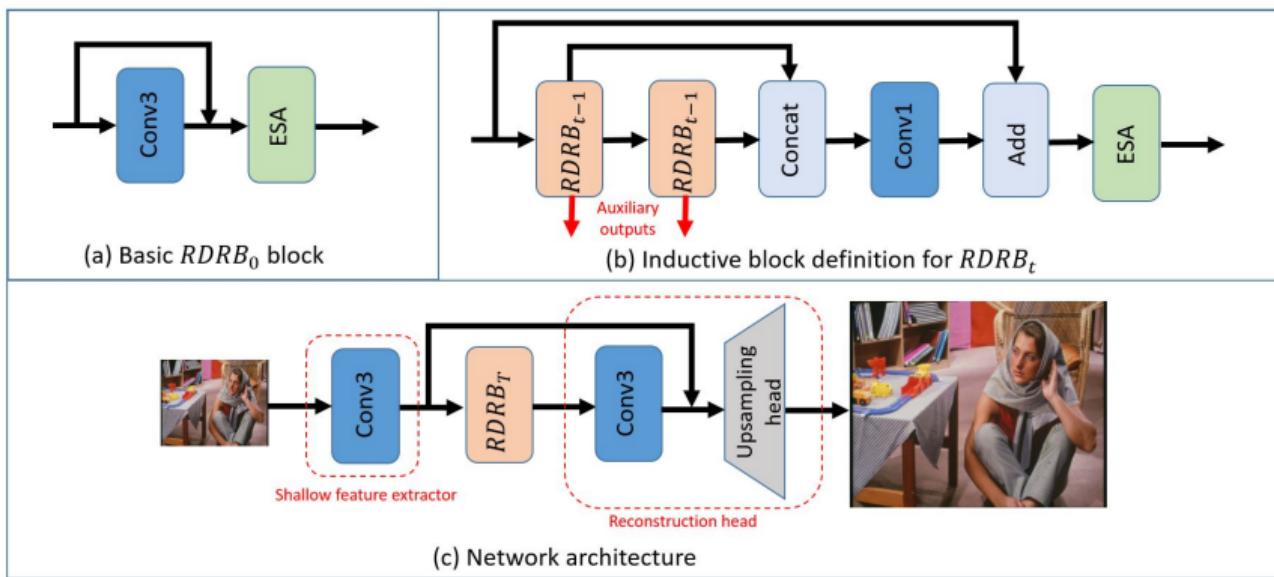


RDRN: Recursively Defined Residual Network for Image Super-Resolution

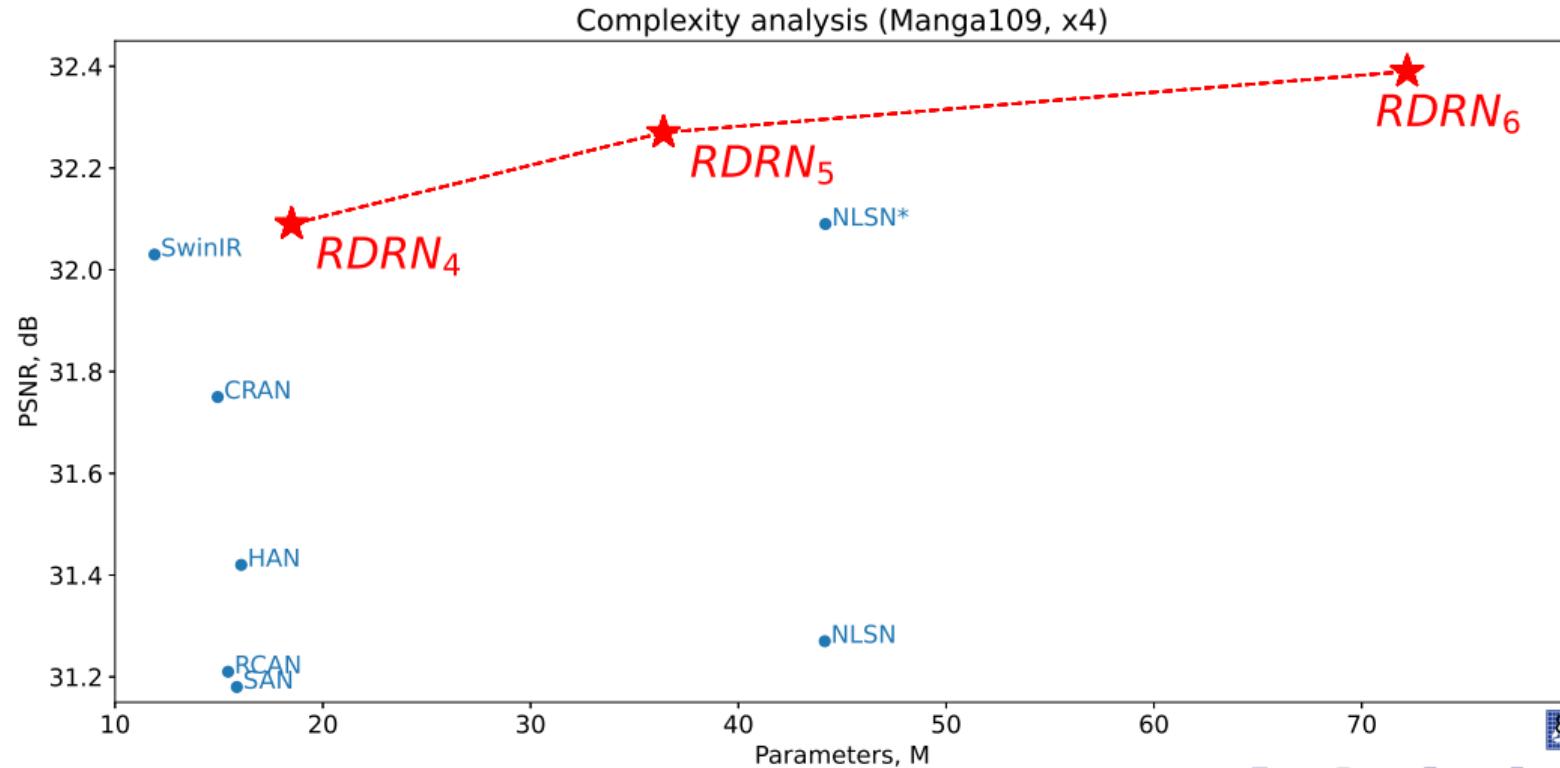
Работа моей группы, принятая на конференцию ACCV2022!

Идея

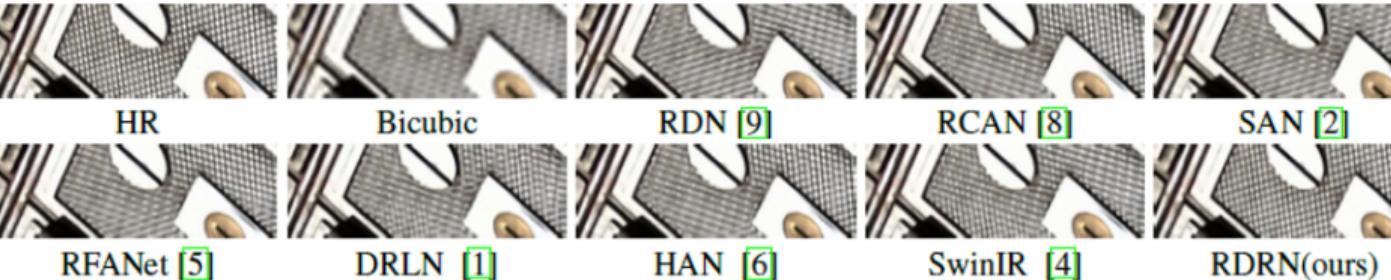
Рекурсивное определение основного блока сети позволяет улучшить качество



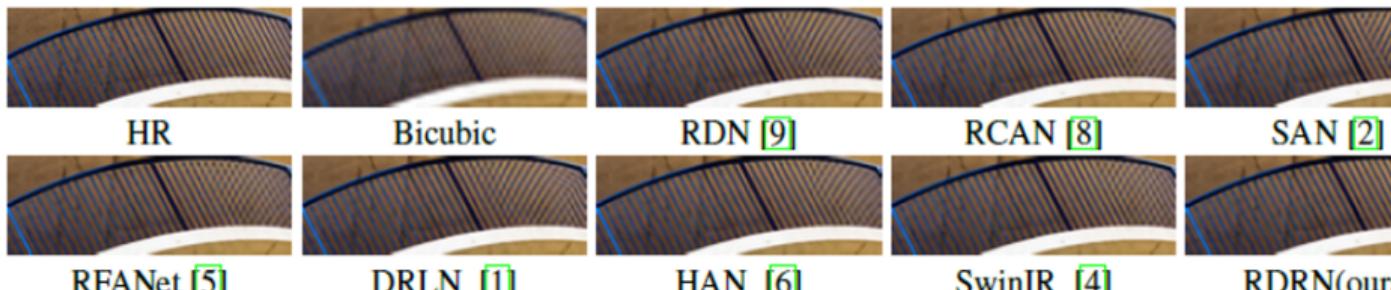
RDRN результаты



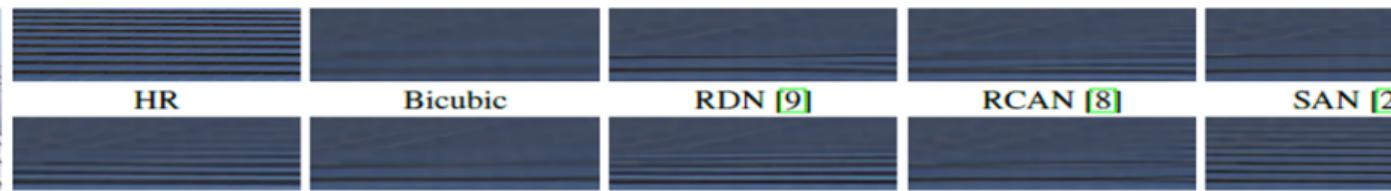
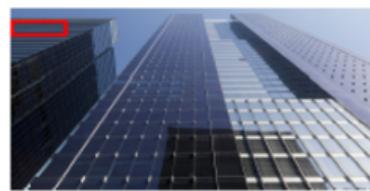
RDRN результаты



Urban100: img_005 ($\times 2$)



Urban100: img_100 ($\times 2$)



Urban100: img_033 ($\times 4$)

Проблема

- Задача сверхразрешения хорошо решается для искусственно сгенерированных датасетов
- Если же таким образом обученную модель применить к реальному изображению, чтобы увеличить его размер, то результат будет низкого качества
- довольно сложно отличить нейросетевую модель от бикубической интерполяции



Сложности задачи сверхразрешения для реальных изображений

Проблема

- Задача сверхразрешения хорошо решается для искусственно сгенерированных датасетов
- Если же таким образом обученную модель применить к реальному изображению, чтобы увеличить его размер, то результат будет низкого качества
- довольно сложно отличить нейросетевую модель от бикубической интерполяции

Возможные решения

- Собрать парные датасеты
- Моделировать функцию downsampling-a (деградации)



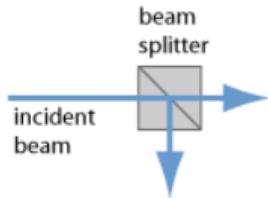
Сложности задачи сверхразрешения для реальных изображений

Проблема

- Задача сверхразрешения хорошо решается для искусственно сгенерированных датасетов
- Если же таким образом обученную модель применить к реальному изображению, чтобы увеличить его размер, то результат будет низкого качества
- довольно сложно отличить нейросетевую модель от бикубической интерполяции

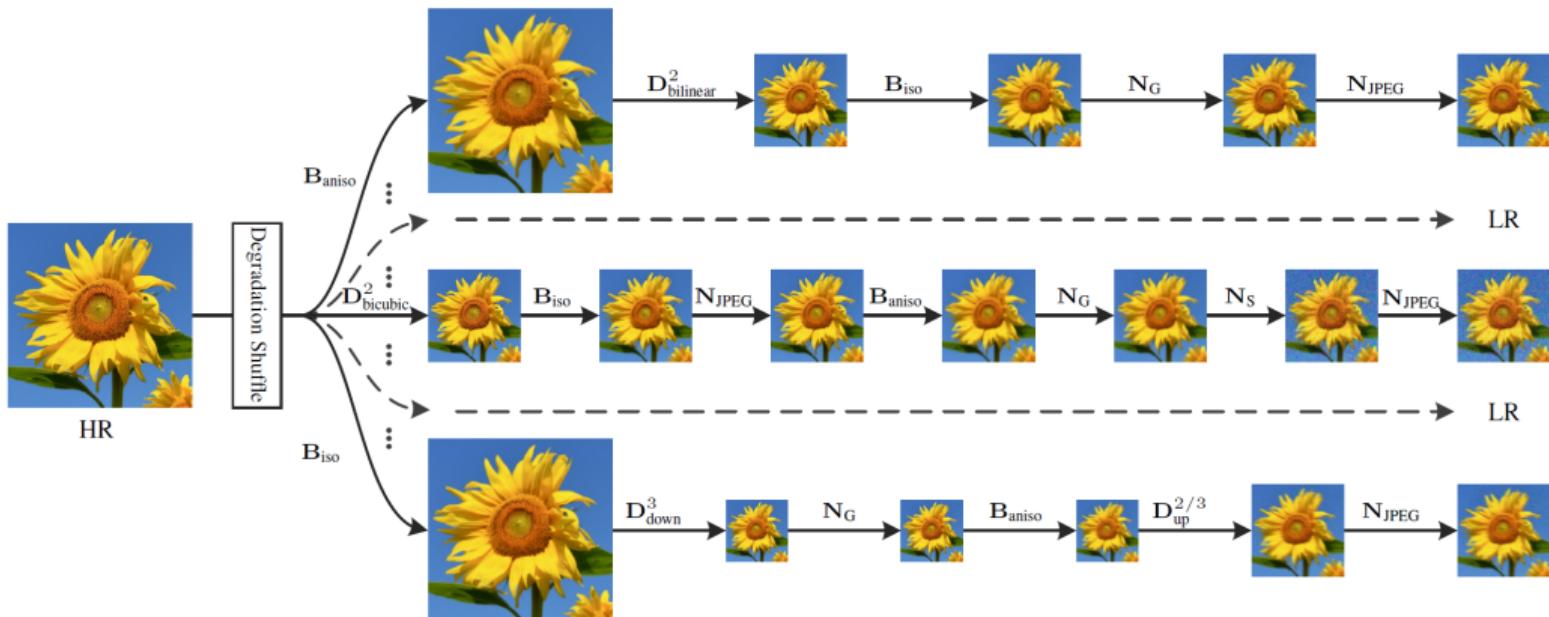
Возможные решения

- Собрать парные датасеты
- Моделировать функцию downsampling-a (деградации)



- Использование генеративных моделей
 - Генеративно-состязательные модели (GAN)
 - Диффузионные модели (Denoising Diffusion Probabilistic Models)
- Подбор деградации к данным

Провинутые деградации



Генеративные модели: StyleGAN, GFPGAN

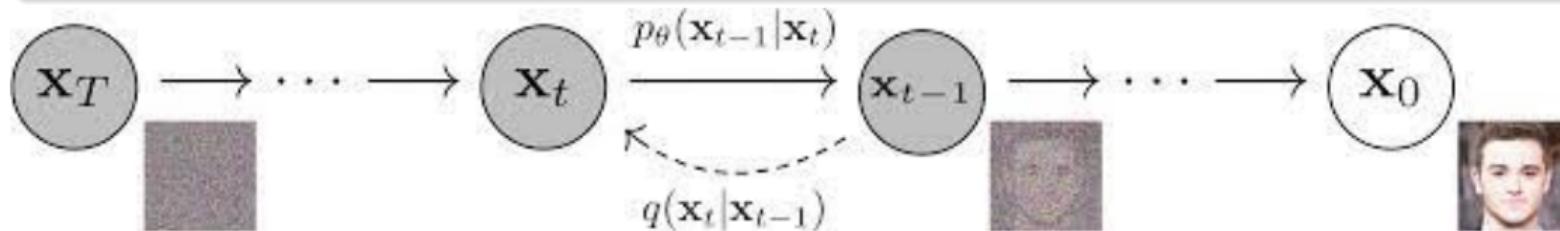


Denoising Diffusion Probabilistic Models (DDPM)

Идея

Прямой проход: добавление шума на каждой итерации до полностью зашумленной картинки.

Обратный проход: удаление шума, генерация изображения из полностью зашумленной картинки



DDPM: SR3

<https://iterative-refinement.github.io/>



Super Resolution For Real Faces

Task: Improve the baseline super-resolution model for the face super-resolution problem for real-world images

Target: Improve PQ score on test set from 6 points to 8 points (challenging target - to 9 points)

Main Research Directions:

1) GAN based models

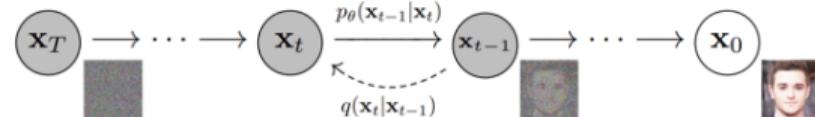
- Pros: Well studied models (since 2014)
Works good with acceptable for application inference time
There are methods to control face attributes (StyleGANs)
- Cons: Sometimes GANs produce bad looking artefacts

2) Denoising Diffusion Probabilistic Models (DDPM)

- Pros: DDPMs show outstanding performance, most promising models
- Cons: Very slow

Current Status

- Started research on GAN based face SR, editing and enhancement
 - Main our goal is to understand how to **control face attributes**
- **Proposed by CameraAI Face Recognition Guided Diffusion (FRGD) pipeline significantly outperforms baseline GFPGAN and other models for extremely small faces**



Example of DDPM inference



Real Face SR: Face Recognition Guided Diffusion (FRGD)

Main Idea: For low quality images the useful signal is weak and it's necessary to use reference for content generation

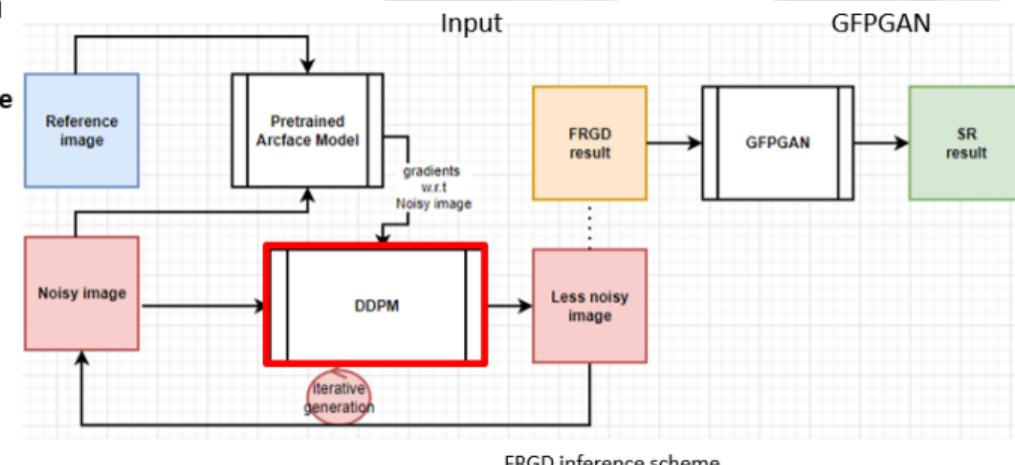
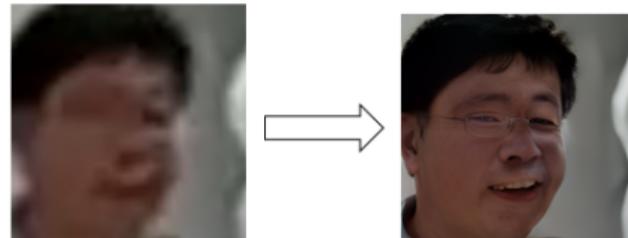
Our contributions

- We designed novel training scheme of **embedding-conditional DDPM**

- We proposed conditioning design, design of losses and supervision (masked-, noisy-, checkerboard- fusion)

- Novel pipeline and face recognition guided inference scheme for embedding-conditional DDPM
- Main FRGD properties:

- Stochasticity
- Trade-off between Preserving Content and Preserving Identity
- Changing Identity



FRGD inference scheme



FRGD Properties: Stochasticity



Input image



Reference image



GFGAN result

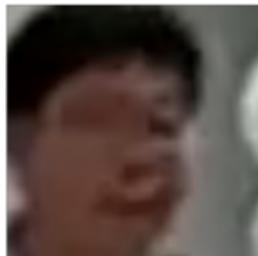


Ours

**Conclusions: For same input and reference FRGD can produce different results
It's possible to sample different images until we are happy with the result**



FRGD Properties: Trade-off between Preserving Content and Preserving Identity



Input image



Reference image



Content preserving mode, wrong ID



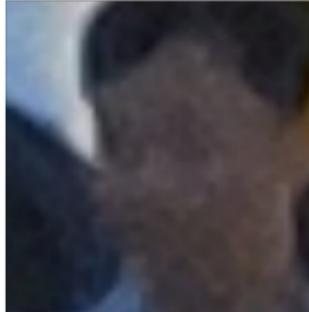
Balance of ID and content preserving



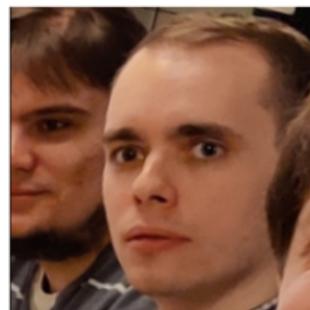
Identity preserving mode, image
doesn't correspond to input image



FRGD Properties: Changing Identity



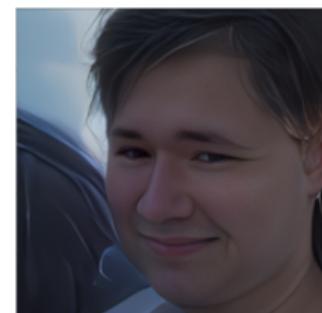
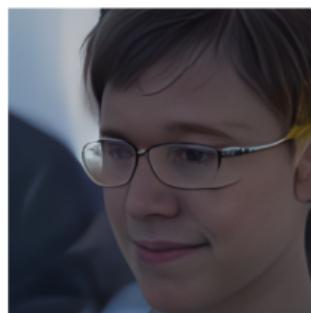
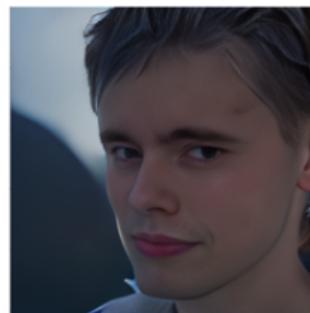
Input image



Reference image



Our results





Input



FRGD



Input



FRGD

HUAWEI TECHNOLOGIES CO., LTD.

Huawei Confidential

7



JAWEI

FRGD Conclusions

- **Advantages of proposed method**

- FRGD handles extremely hard cases (where every other single model fails)
- FRGD generates faces with high similarity to reference ID
- FRGD can be used for interactive face enhancement (for example, in image editor like Photoshop)
 - Ability to control the tradeoff between content preservation and similarity to reference ID
 - Masks allow selective image editing
 - Stochastic model allows us to sample different images until we are happy with the result

- **Points to improvement**

- Large inference time: 30-60 sec for one face
- Possibility of artifacts when input and reference are very different
- Poor performance in non-canonical cases (heavily rotated faces, face attributes like masks and glasses)
- General SR model works worse than FaceSR. We see sharp face on blurred body



Интересные концепты: бесконечный zoom



Спасибо за внимание!

