

Разработка программной системы для увеличения разрешения изображений на основе нейросетевых технологий

Толмачева Анастасия Вячеславовна

Телефон: +79517942166

Email: froshiksid@gmail.com

tg: @Malenkaya_duck

Содержание

- Постановка задачи
- Обзор аналогичных систем
- Обзор нейросетевых моделей для увеличения разрешения изображения
- Коэффициент увеличения разрешения изображения
- Обзор метрик оценки качества
- Пример работы
- Вопросы и предложения

Постановка задачи

Исходное изображение



Разрешение 500×364 px

Система



Выходное изображение



Разрешение 2000×1456 px
Увеличение разрешения в 4 раза

Обзор аналогов (общие характеристики)

№	Название	Онлайн доступ	Десктопная версия	API	Пробные фото	Цена	Целевая категория изображений	Открытый код	Базовая модель ИНС
1	Gigapixel AI	-	+	+	-	Платный	Фото	URL	CNN
2	Let's Enhance	+	-	+	+	Платный	Арт и фото	URL	CNN и GAN
3	Bigjpg	+	+	+	+	Есть платные услуги	Арт	-	CNN
4	Waifu2x	+	-	-	+	Бесплатный	Арт	URL	CNN
5	Imglarger	+	-	-	+	Платный	Фото	-	-
6	Upscayl	+	+	+	+	Бесплатный	Арт и фото	URL	GAN

Обзор аналогов (общие характеристики)

№	Название	Формат		Мах увеличение		Мах размеры входного изображения
		импорта	экспорта	разрешения	изображения	
1	Gigapixel AI	JPEG, PNG, TIFF, RAW	JPEG, PNG, TIFF, DNG	32 000 × 32 000 px	6x	- (но TIFF до 4 Гб)
2	Let's Enhance	JPEG, PNG, WebP	JPEG, PNG, WebP	23000 × 23000 px	16x	50 Мб
3	Bigjpg	JPEG, PNG, TIFF, WebP	JPEG, PNG, TIFF, WebP	Нет данных	16x	50 Мб
4	Waifu2x	JPEG, PNG, WebP, TIFF	PNG, WebP	Нет данных	4x	5 Мб
5	Imglarger	JPG, PNG, WebP	JPG, PNG, WebP	8000 × 4000 px	4x	10 Мб
6	Upscayl	JPEG, PNG, TIFF, WebP	JPEG, PNG, TIFF, WebP	Нет данных	16x	Нет данных

Обзор нейросетевых моделей для увеличения разрешения изображения (1/3)

№	Название	Статья	Реализация	Архитектура	Время обучения	Время инференса, с	Обучающая выборка	Работает с Фото +/- Детализация +/- Арты +/- Шум на фото +/-
1	SRCNN	PDF	Сайт Github	3 сверточных слоя	Нет данных	0,1 – 1	ILSVRC 2013 ImageNet : 395K фото, JPEG, 1000 категорий	+ - - -
2	FSRCNN	PDF	Сайт Github	5 сверточных слоев (нелинейное отображение разбито на три слоя)	Нет данных	0,01 – 0,1	91-image dataset : 91 фото, PNG; General-100 dataset : 100 изображений BMP	+ - - -

Обзор нейросетевых моделей для увеличения разрешения изображения (2/3)

№	Название	Статья	Реализация	Архитектура	Время обучения	Время инференса, с	Обучающая выборка	Работает с Фото +/- Детализация +/- Арты +/- Шум на фото +/-
3	ESPCN	PDF	Github	Три сверточных слоя, использует субпиксельную свертку	7 дней на графическом процессоре K2	0,2 – 0,3	ImageNet : 50K рандомно взятых фото	+ - - -
4	SRGAN	PDF	Github	Генератор использует субпиксельную свертку	Нет данных	0,012 – 0,030	ImageNet : 350K рандомно взятых фото	+ + - -

Обзор нейросетевых моделей для увеличения разрешения изображения (3/3)

№	Название	Статья	Реализация	Архитектура	Время обучения	Время инференса, с	Обучающая выборка	Работает с Фото +/- Детализация +/- Арты +/- Шум на фото +/-
5	ESRGAN	PDF	Github	Генератор содержит блоки Residual-in-Residual Dense	Нет данных	Нет данных	DIV2K : 800 изображений, PNG; Flickr2K : 2650 изображений, PNG; OST : формат PNG	+ + + -
6	Real-ESRGAN	PDF	Github	Генератор использует субпиксельную свертку	Нет данных	Нет данных	Такие же, как у ESRGAN	+ + + +

Коэффициент увеличения разрешения изображений (1/2)

№	Название	Коэффициенты	
		бесплатные	платные
1	Gigapixel AI	-	0.5x, 2x, 4x, 6x
2	Let's Enhance	1x, 2x, 4x	8x, 16x
3	Bigjpg	2x, 4x	8x, 16x
4	Waifu2x	1x, 1.6x, 2x, 4x	-
5	Imglarger	2x	4x
6	Upscayl	1x – 5x; 6x – 16x – может вызвать проблемы с производительностью	-

Коэффициент увеличения разрешения изображений (2/2)



Оригинал



Увеличение разрешения в 3 раза,
бикубическая интерполяция



Увеличение разрешения в 3 раза,
SRCNN

Обзор метрик качества (определение)

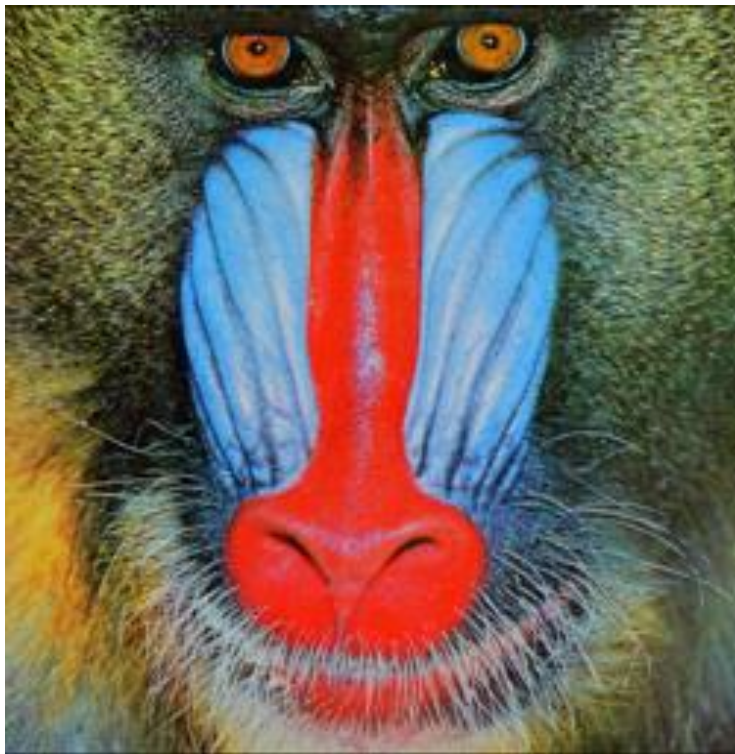
№	Название	Неформальное определение	Диапазон значений	Предпочтительное значение	Реализация
1	PSNR (Peak Signal-to-Noise Ratio)	Измеряет отношение максимальной мощности сигнала к мощности шума	Вещественное неотрицательное	Чем выше, тем лучше качество изображения	+
2	SSIM (Structural Similarity Index)	Оценивает визуальное восприятие изменений в структуре изображения. SSIM учитывает яркость, контрастность и структуру	От -1 до 1	1 – указывает на полное совпадение	+
3	NIQE (Natural Image Quality Evaluator)	Метод, который оценивает качество изображения, используя статистические признаки, и сравнивает их с эталонной моделью, построенной на естественных изображениях	От 0 до 100	0 – наилучший результат	+

Обзор метрик качества (формулы)

№	Название	Формула
1	PSNR (Peak Signal-to-Noise Ratio)	$PSNR = 10 \log_{10} \left(\frac{MAX_I^2}{MSE} \right) = 20 \log_{10} \left(\frac{MAX_I}{\sqrt{MSE}} \right),$ <p>где MAX – максимальное возможное значение пикселя изображения, MSE – среднеквадратическая ошибка</p>
2	SSIM (Structural Similarity Index)	$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)},$ <p>где μ_x, μ_y – средние значения пикселей оригинального и сгенерированного изображений, σ_x^2, σ_y^2 – дисперсии значений пикселей, σ_{xy} – ковариация между оригинальным и сгенерированным изображениями, c_1 и c_2 – малые константы, которые предотвращают деление на ноль</p>
3	NIQE (Natural Image Quality Evaluator)	$D(v_1, v_2, \Sigma_1, \Sigma_2) = \sqrt{((v_1 - v_2)^T \left(\frac{\Sigma_1 + \Sigma_2}{2} \right)^{-1} (v_1 - v_2))},$ <p>где v_1, v_2 и Σ_1, Σ_2 – это векторные средние и ковариационные матрицы для естественной модели многомерного гауссовского распределения и модели многомерного гауссовского распределения искаженного изображения</p>

Пример работы

Оригинал
(133 КБ)

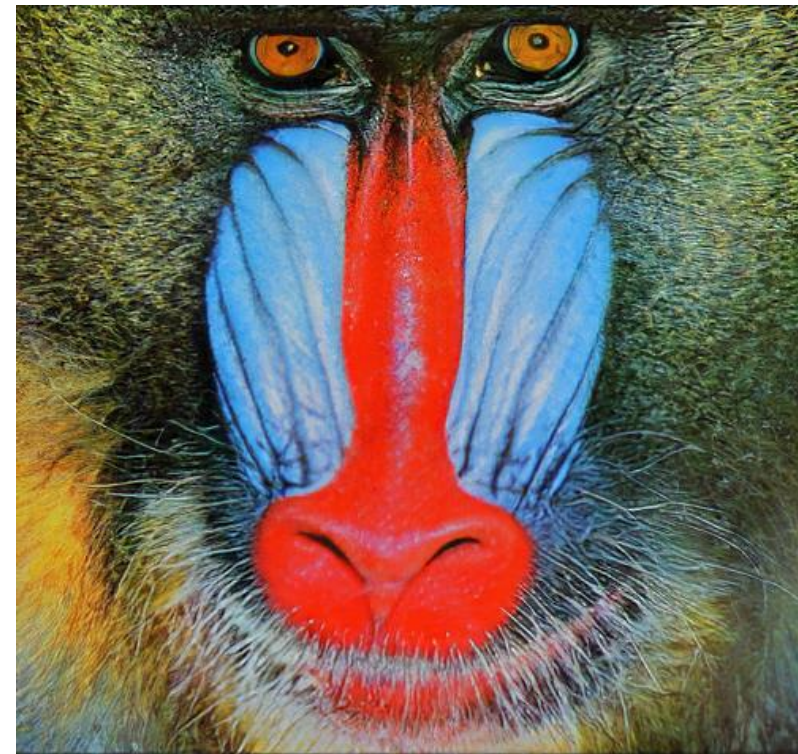


ESPCN, увеличение разрешения в
2 раза (446 КБ)



PSNR (дБ): 29.45
SSIM: 0.7850
NIQE: 4.4710

ESRGAN, увеличение разрешения
в 4 раза (555 КБ)



PSNR (дБ): 28.83
SSIM: 0.5974
NIQE: 2.3841 13/14

Вопросы и предложения

Функциональные требования:

1. Типовой цикл работы с системой:

1. Загрузка изображения,
2. Выбор моделей,
3. Выбор увеличения,
4. Генерация,
5. Подсчет метрик качества и их вывод,
6. Сохранение результата

2. Какая обучающая выборка и откуда будет передаваться? Предложение: применять для каждой модели соответствующую выборку, которую использовали разработчики.

3. Формат входных и выходных изображений: JPEG, PNG, TIFF.

Нефункциональные требования:

1. Интерфейс: командная строка или графическая оболочка?
2. На какой платформе предполагаются обучение и инференс?