

УВЕЛИЧЕНИЕ РАЗРЕШЕНИЯ ИЗОБРАЖЕНИЙ НА ОСНОВЕ НЕЙРОННЫХ СЕТЕЙ

Гарашук Н.В., студент гр.150501

*Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь*

Перцев Д.Ю. – кандидат технических наук

Аннотация. В статье рассматриваются особенности и структура современной нейросетевой архитектуры SRGAN, предназначенной для увеличения разрешения изображений. Подробно описаны ключевые компоненты модели, включая генератор и дискриминатор, а также функции потерь, используемые в процессе обучения.

Ключевые слова. SRGAN, генеративно-сопоставительные сети, увеличение разрешения изображений, бикубическая интерполяция, ImageNet, PSNR, SSIM.

Введение. Современный мир характеризуется стремительным развитием цифровых технологий, где визуальный контент играет ключевую роль в различных сферах жизни: от дизайна до научных исследований и медицинской визуализации. Высокое качество изображений – ключевое требование в современном мире, однако значительная часть существующих визуальных материалов имеет низкое разрешение, что существенно ограничивает их практическое применение. Традиционные методы увеличения разрешения, например, бикубическая интерполяция, часто не обеспечивают достаточной детализации и естественности результата, что создает необходимость в разработке более совершенных подходов.

Нейронные сети, в частности генеративно-сопоставительные архитектуры (Generative Adversarial Network, GAN), открывают новые возможности в области обработки изображений. Модель SRGAN (Super-Resolution Generative Adversarial Network) демонстрирует значительный потенциал для повышения разрешения с сохранением текстуры и деталей, что делает её перспективным инструментом для решения проблемы.

Архитектура генеративно-сопоставительной нейронной сети SRGAN. SRGAN [1] – это архитектура, основанная на генеративно-сопоставительных сетях, предназначенная для увеличения разрешения изображений с сохранением высокого качества. Она состоит из двух основных компонентов: генератора и детектора, которые обучаются в сопоставительной манере.

Генератор в SRGAN отвечает за преобразование изображения в низком разрешении (Low Resolution, LR) в высокое (High Resolution, HR). Его архитектура включает следующие ключевые элементы:

1. Входной слой принимает LR-изображение и применяет свертку с ядром 9×9 для извлечения низкоуровневых признаков.
2. Сверточные блоки (16 элементов) образуют основной компонент генератора. Каждый блок содержит сверточные слои с ядром 3×3 .
3. Пакетная нормализация (Batch Normalization, BN) используется в каждом сверточном блоке для стабилизации обучения.
4. Функция активации PReLU (Parametric Rectified Linear Unit), параметрический выпрямленный линейный блок с параметром $\alpha=0.2$ применяется после сверточных слоев.
5. Skip-соединения добавлены между блоками для предотвращения затухания градиентов.
6. Слои повышающей субдискретизации увеличивают разрешение изображения с помощью PixelShuffle (перестановка пикселей) масштаб $\times 2$.
7. Выходной слой выполняет финальную свертку 9×9 и использует tanh-активацию (гиперболический тангенс) для ограничения значений пикселей в $[-1, 1]$.

Детектор в SRGAN играет роль классификатора, который отличает сгенерированные изображения (SR, Super Resolution) от изображений с HR-изображением. Его архитектура включает следующие элементы:

1. Входной слой принимает HR-изображения или SR-изображение и применяет свертку с ядром 3×3 для извлечения признаков.
2. Сверточные блоки дискриминатора содержат сверточные слои с ядром 3×3 , слои BN и функцию активации LeakyReLU (Rectified Linear Unit с утечкой) с коэффициентом наклона $\alpha=0.2$ для отрицательных значений.
3. Полносвязные слои дискриминатора выполняют классификацию изображений на «реальные» или «сгенерированные» с использованием сигмоидальной функции активации (Sigmoid), дающей вероятность в диапазоне $[0, 1]$.

Функция потерь дискриминатора использует бинарную перекрестную энтропию (Binary Cross-Entropy, BCE) для обучения модели отличать настоящие изображения от сгенерированных.

Обучение SRGAN основано на комбинации следующих функций потерь:

1. Воспринимаемая потеря (Perceptual loss). Эта потеря измеряет разницу между сгенерированным и реальным изображением в пространстве признаков, извлеченных из предобученной сети VGG-19. Она фокусируется на восстановлении высокоуровневых деталей. Данная функция потерь описывается формулой:

$$L_{content} = \sum_{i \in S} (\phi_i[I^{SR}] - \phi_i[I^{HR}])^2, \quad (1)$$

где $\phi_i[I^{SR}]$ – признаки, извлеченные из i -го слоя сети VGG для сгенерированного изображения I^{SR} ; $\phi_i[I^{HR}]$ – признаки, извлеченные из i -го слоя сети VGG для сгенерированного изображения I^{HR} ; S – набор слоев VGG (в текущей реализации 8, 17, 26, 35 слои).

2. Состязательная потеря (Adversarial loss). Она побуждает генератор создавать изображения, которые детектор не сможет отличить от реальных. Используется для обучения генератора в рамках «состязания».

Состязательная потеря для генератора описывается формулой:

$$L_{adv} = -E_{I^{SR}}[\log D(I^{SR})], \quad (2)$$

где $D(I^{SR})$ – вероятность, что детектор классифицирует сгенерированное изображение I^{SR} как реальное.

Состязательная потеря для детектора описывается формулой:

$$L_{disc} = -E_{I^{HR}}[\log(D(I^{HR}))] - E_{I^{SR}}[\log(1 - D(I^{SR}))], \quad (3)$$

где $D(I^{HR})$ – вероятность, что детектор классифицирует реальное изображение I^{HR} как реальное;

$D(I^{SR})$ – вероятность, что детектор классифицирует сгенерированное изображение I^{SR} как реальное.

3. Комбинированная потеря. Для обучения генератора используется комбинированная функция потерь, состоящая из состязательной и перцептивной потерь, и описывается формулой:

$$L_{gen} = L_{content} + \lambda \cdot L_{adv}, \quad (4)$$

где $L_{content}$ – перцептивная потеря;

L_{adv} – состязательная потеря;

λ – коэффициент, определяющий вес состязательной потери (в работе принят равным 10^{-3}).

Полная структура нейронной сети представлена на рисунк 1.

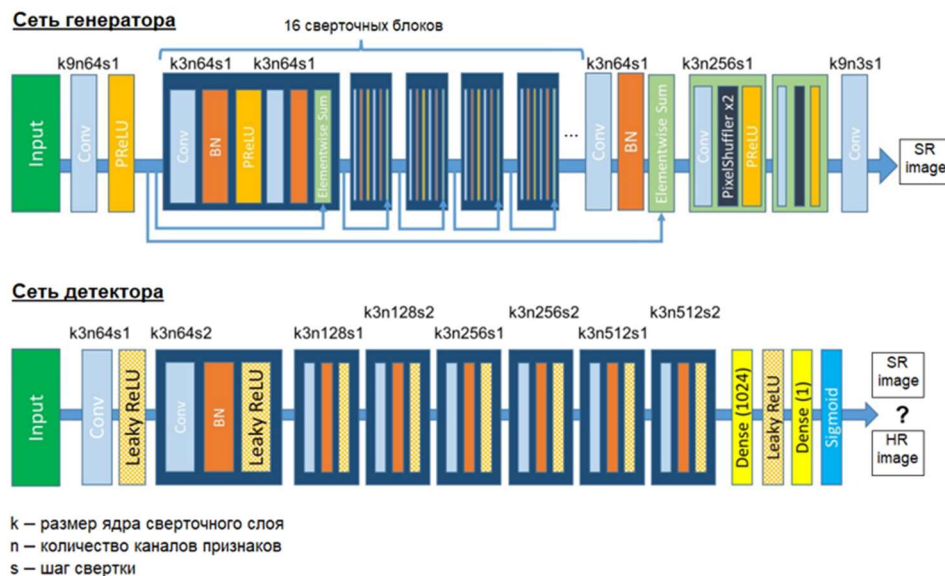


Рисунок 1 – Полная структура нейронной сети SRGAN

Описание тестовых данных. Для обучения, валидации и тестирования выбран набор данных ImageNet [2], представляющий собой большую базу изображений, содержащую более 14 миллионов изображений в разных классах (всего 1000 классов по состоянию на 2025 год). В работе были выделены следующие подмножества:

- обучающая выборка: 100 тысяч изображений высокого разрешения (к высокому разрешению отнесены изображения с разрешением 512×512 пикселей и выше);
- валидационная выборка: 10 тысяч изображений для оценки обучения модели во время обучения;
- тестовая выборка: 1000 изображений, не участвовавших в обучении, для итоговой оценки модели.

Для обучения модели выполняется предварительная обработка данных. Сначала из исходного изображения случайным образом вырезается квадратная область размером 96×96 пикселей, значения интенсивности которой нормализуются в диапазоне от -1 до 1. Эта область рассматривается как HR-изображение. Затем HR-изображение масштабируется в 4 раза с помощью бикубической интерполяции до размера 24×24 пикселя, формируя LR-изображение. Для расширения вариативности обучающей выборки к полученным HR и LR изображениям применяются аугментационные преобразования, включающие случайные горизонтальные отражения, зеркальные преобразования и повороты.

Для оценки результатов на тестовых данных использовались следующие метрики:

1. Пиковое отношение сигнала к шуму (Peak Signal-To-Noise Ratio, PSNR) [3]:

$$PSNR = 10 * \log_{10} \left(\frac{MAX^2}{MSE} \right), \quad (5)$$

где MAX – максимальное значение пикселя (например, 255 для 8-битных изображений);

MSE – средняя квадратичная ошибка (Mean Squared Error, MSE) между сгенерированным и эталонным изображением.

2. Индекс структурного сходства (Structure Similarity, SSIM) [4]:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (6)$$

где μ_x, μ_y – средние значения яркости изображений;

σ_x^2, σ_y^2 – дисперсии яркости;

σ_{xy} – ковариация между изображениями;

C_1, C_2 – константы для стабилизации.

Результаты работы. Тестовое изображение было обработано нейросетью SRGAN и алгоритмом билинейной интерполяции, что показано на рисунке 2. SRGAN эффективно восстанавливает мелкие детали (например, текстуру шерсти или элементы ландшафта), которые при билинейной интерполяции размываются.

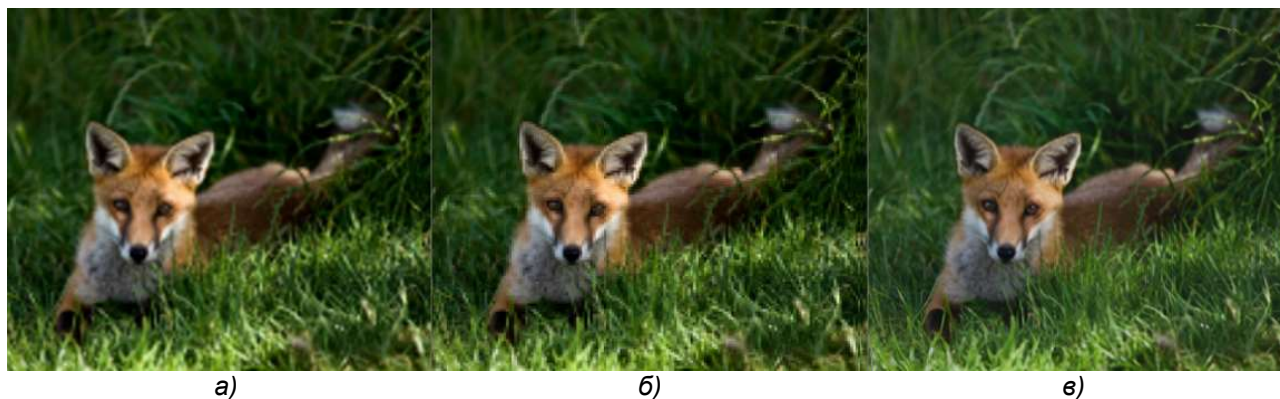


Рисунок 2 – Примеры увеличения разрешения: а – исходное изображение; б – алгоритм билинейной интерполяции; в – обработка нейросетью SRGAN

В таблице 1 приведены метрики обучения SRGAN и метрики билинейной интерполяции.

Таблица 1. Метрики сравнения алгоритмов увеличения разрешения изображения

Метрика	SRGAN	Билинейная интерполяция
PSNR, дБ	26.55	22.31
SSIM	0.79	0.64

По метрике PSNR (26.55 дБ против 22.31 дБ) SRGAN показывает улучшение на 4.24 дБ, что свидетельствует о более точном восстановлении пиксельных значений и меньшем уровне шумов. Для задач, требующих высокой точности (например, медицинская визуализация или спутниковые снимки), такое превышение является критически важным. Метрика SSIM (0.79 против 0.64) подтверждает, что SRGAN лучше сохраняет структурную целостность и естественность текстуры изображения. Значение выше 0.75 (для SRGAN) указывает на высокую визуальную корректность, тогда как результат билинейной интерполяции (0.64) попадает в зону «умеренного качества».

Ограничения при использовании SRGAN:

- SRGAN требует больше вычислительных ресурсов для обучения и использования;
- на некоторых типах изображений (например, с резкими геометрическими паттернами) возможны искажения.

SRGAN предпочтителен для задач, где важна детализация, в то время как билинейная интерполяция может использоваться для быстрой обработки, когда качество не приоритетно.

Закключение. Основной проблемой при проектировании систем увеличения разрешения изображений является необходимость баланса между качеством восстановления и вычислительной сложностью. Традиционные методы, такие как бикубическая интерполяция, хотя и менее ресурсоемки, уступают по качеству нейронным сетям, таким как SRGAN. Последние демонстрируют высокую эффективность в восстановлении деталей и текстуры изображений, что подтверждается метриками PSNR и SSIM. Однако использование SRGAN требует значительных вычислительных ресурсов, что может ограничивать их применение в реальных условиях. Таким образом, дальнейшие исследования должны быть направлены на оптимизацию архитектуры нейронных сетей для снижения их вычислительной сложности без потери качества результатов.

Список использованных источников:

1. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network [Электронный ресурс] – Электронные данные. – Режим доступа: <https://arxiv.org/pdf/1609.04802> – Дата доступа: 22.02.2025
2. ImageNet [Электронный ресурс] – Электронные данные. – Режим доступа: <https://paperswithcode.com/dataset/imagenet> – Дата доступа: 22.02.2025
3. Illustration-of-the-PSNR-measure [Электронный ресурс] – Электронные данные. – Режим доступа: https://www.researchgate.net/figure/Illustration-of-the-PSNR-measure_fig6_268807675 – Дата доступа: 22.02.2025
4. Measure the Similarity Between Images [Электронный ресурс] – Электронные данные. – Режим доступа: <https://medium.com/@datamonsters/a-quick-overview-of-methods-to-measure-the-similarity-between-images-f907166694ee> – Дата доступа: 22.02.2025

IMAGE RESOLUTION ENHANCEMENT BASED ON NEURAL NETWORKS

Garashchuk N.V.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

Pertsau D.Y. – PhD of Technical Sciences

Annotation. The article examines the features and structure of the modern neural network architecture SRGAN, designed for image resolution enhancement. The key components of the model, including the generator and discriminator, as well as the loss functions used during training, are described in detail.

Keywords: SRGAN, generative-adversarial networks, image resolution enhancement, bicubic interpolation, ImageNet, PSNR, SSIM.