

# Проведение экспериментов с рядом сверточных нейронных сетей по оценке техник Super-Resolution

А.В. Строкова<sup>1</sup>

<sup>1</sup> Университет ИТМО, Кронверский проспект, дом 49, литер А, Санкт-Петербург, 197101, Россия

## Аннотация

Работа посвящена исследованию моделей сверточных нейронных сетей SRCNN, FSRCNN и SubPixel CNN, позволяющих решать задачу Super-Resolution, с целью выявления оптимальных параметров на этапе обучения. Для оценки результатов использовалась стандартная метрика PNSR, в качестве датасета взят BSD. При анализе функций активации и оптимизаторов из библиотеки PyTorch выявлены лучшие сочетания, дающие более высокую метрику качества. Используя веса трех лучших получившихся комбинаций, была проведена сравнительная характеристика изображений, отнесенных к классам текст, лицо, ночь, природа и сгенерированное изображение. Сделан вывод о пригодности той или иной архитектуры к увеличению изображения определенного класса. Полученные результаты могут использоваться в дальнейших исследованиях, направленных на улучшение производительности и точности методов Super-Resolution.

## Ключевые слова

Сверхвысокое разрешение изображений, сверточная нейронная сеть, пиковое отношение сигнала к шуму, функция активации, оптимизатор, SRCNN, FSRCNN, SubPixel CNN, PyTorch.

# Experiments with a Number of Convolutional Neural Networks for Evaluating the Super-Resolution Technique

A.V. Stroikova<sup>1</sup>

<sup>1</sup> ITMO University, Kronversky Pr., 49 bldg., St. Petersburg, 197101, Russia

## Abstract

The work is devoted to the study of the models of convolutional neural networks SRCNN, FSRCNN, and SubPixel CNN, which solve the Super-Resolution task, for finding the optimal parameters during the training stage. To evaluate the results, the standard PNSR metric was used, BSD was taken as the dataset. During analyzing activation functions and optimizers from the PyTorch library, the best combinations, that give a higher quality metric, were identified. Using the weights of the three best combinations obtained, a comparative characterization of the images belonging to the classes of text, face, night, nature, and the generated image was given. The conclusion is made about the evaluation of a particular architecture to obtain an image of a certain class. The finishing results can be used in further studies for improving the productivity and accuracy of Super-Resolution methods.

## Keywords

Image Super-Resolution, convolutional neural network, peak signal-to-noise ratio, activation function, optimizer, SRCNN, FSRCNN, SubPixel CNN, PyTorch.

ГрафиКон 2023: 33-я Международная конференция по компьютерной графике и машинному зрению, 19-21 сентября 2023 г., Институт проблем управления им. В.А. Трапезникова Российской академии наук, г. Москва, Россия

EMAIL: s.anastasia.stroikova@gmail.com (А.В. Строкова)

ORCID: 0000-0002-6850-610X (А.В. Строкова)



© 2023 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

## 1. Введение

В настоящее время качественное и четкое изображение на любом носителе является одним из главных факторов успеха в любой сфере, начиная от рекламы и маркетинга, заканчивая областью компьютерного зрения. Однако не всегда можно получить фотографии нужного размера и качества. Например, нейронная сеть Fusion Brain от Сбера (Kandinsky 2.2) генерирует изображение, максимальный размер которого составляет 1024 пикселя в длину и ширину. Этого не всегда хватает, чтобы покрыть потребность пользователя-дизайнера. В нашем примере изображение потеряет свою привлекательность при попытке растяжения.

Для решения этой проблемы можно использовать несколько подходов. Первый из них – генерировать изображения частями, склеивая их в сторонней программе или напрямую в Kandinsky 2.2. Но этот путь долгий и не применим в ситуациях, когда качественное изображение нужно получать в режиме реального времени. Второй подход – векторизировать изображение с помощью сторонних программ, а затем увеличить до нужных размеров. Несмотря на достаточно быструю обработку, реалистичность картинки теряется, придавая ей мультипликационный вид. Третий подход – использование нейронных сетей для увеличения изображений в два, четыре, восемь или шестнадцать раз. Такая техника называется Super-Resolution. О ней и пойдет речь в данной статье.

Целью данной статьи является проведение экспериментов с рядом нейронных сетей по оценке техник Super-Resolution. В работе будет рассмотрено несколько известных моделей сверточных нейронных сетей и сравнено их качество при применении к изображениям, разделенных на определенные классы: природный пейзаж, человеческое лицо, текст, ночной снимок, сгенерированное изображение.

Полученные результаты могут быть полезными для дальнейшего совершенствования методов сверхвысокого разрешения, а также помогут выбрать наиболее подходящую модель для конкретного применения.

Для решения задачи Super-Resolution чаще всего используются следующие виды нейронных сетей: сверточные нейронные сети (например, у С. Dong [1], J. S. Ren [2], K. Zhang [3]), генеративно-состязательные нейронные сети (например, у С. Ledig [4] и X. Wang [5]) и прочие более глубокие и тяжелые архитектуры (например, у В. Lim [6], J. Kim [7], М. Haris [8]). В рамках экспериментов были рассмотрены архитектуры сверточных нейронных сетей Super-Resolution CNN [9], Fast Super-Resolution CNN [10] и SubPixel CNN [11].

Для сравнения работы нейронных сетей использовался параметр пикового отношения сигнала к шуму – PNSR:

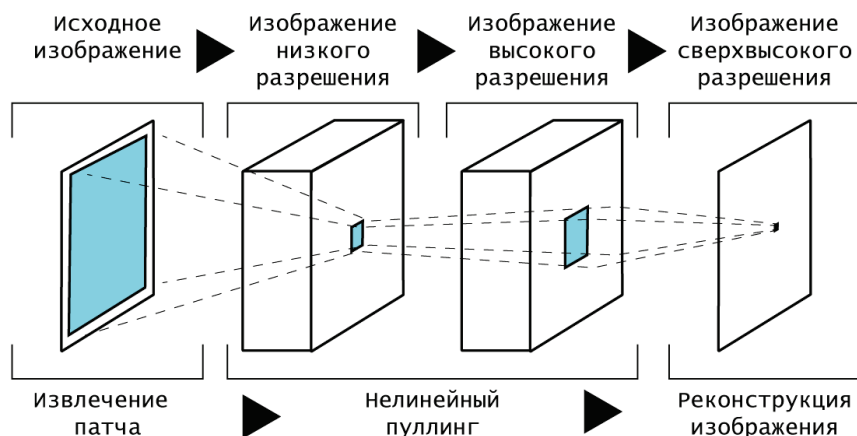
$$PNSR = 20 * \log_{10} \left( \frac{MAX}{\sqrt{MSE}} \right), \quad (1)$$

где  $MAX$  – максимальное значение диапазона пикселей,  $MSE$  – среднеквадратическая ошибка.

В качестве датасета был выбран BSD (Berkeley Segmentation Dataset). В нём содержится около пятисот изображений, которые представляют собой реальные фотографии разных объектов, таких как пейзажи, люди, здания и прочие предметы. Изображения имеют различное разрешение и размер, что создает разнообразие для обучения нейронных сетей. Каждое изображение в датасете сопровождается маской сегментации, которая описывает группы пикселей с определенными характеристиками, такими как цвет, текстура или яркость.

## 2. Описание используемые нейронных сетей

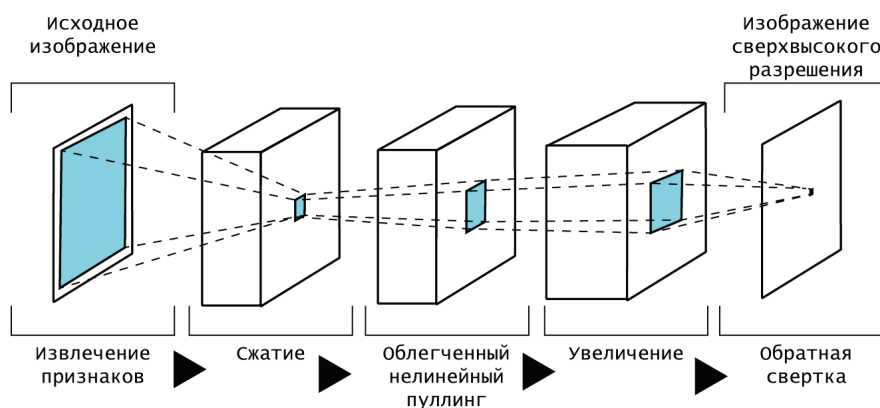
Первая из рассматриваемых нейронных сетей – Super-Resolution Convolutional Neural Network (далее - SRCNN) [9]. Её архитектура представлена на рисунке 1.



**Рисунок 1** – Архитектура нейронной сети SRCNN

Основная идея данной архитектуры заключается в том, что алгоритм заведомо уменьшает качество входного изображения и с помощью математических операций сравнивает расположение пикселей на фотографиях. Основываясь на полученных данных, нейронная сеть визуально улучшает изображения, увеличивая его при этом в  $N$  раз [9]. Процесс работы алгоритма выглядит следующим образом. Исходная фотография разбивается на патчи, откуда извлекаются признаки, присущие данному расположению объектов. Каждый пиксель преобразуется в вектор из значений цвета (например, по палитре RGB). Далее изображение уменьшается, чтобы искусственно создать низкое разрешение (таких преобразований может быть несколько, например, уменьшение в 2 и в 4 раза с помощью бикубической интерполяции). Сравнивая расположение пикселей на изображениях с разными разрешениями, нейронная сеть восстанавливает фотографию в увеличенном размере.

Вторая из рассматриваемых нейронных сетей – Fast Super-Resolution Convolutional Neural Network (далее FSRCNN) [10]. Её архитектура представлена на рисунке 2.

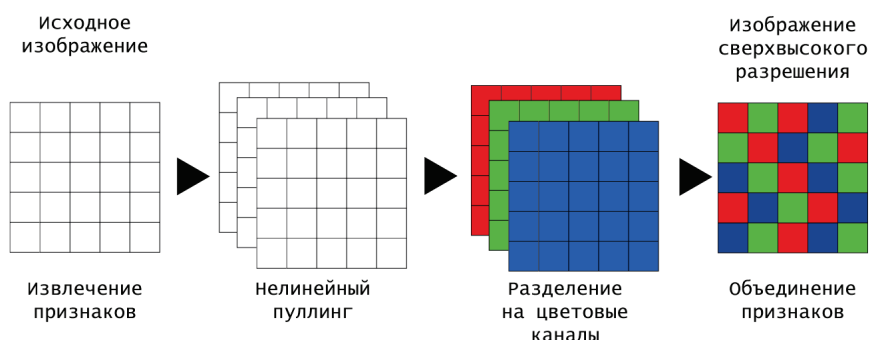


**Рисунок 2** – Архитектура нейронной сети FSRCNN

Если SRCNN использует классическую архитектуру, состоящую из нескольких сверточных слоев и слоев пулинга для извлечения свойств изображений в разных масштабах, то FSRCNN использует более легковесную архитектуру, состоящую из меньшего количества слоев. Слои пулинга в данном случае игнорируются или используются с меньшим коэффициентом. Это позволяет значительно ускорить обучение и производительность модели. Однако выигрыш в скорости может привести к некоторой потере качества изображения, в сравнении с предыдущей моделью.

Главное отличие состоит в том, что изображение, которое подается на вход, не уменьшается с помощью бикубической интерполяции или других алгоритмов. Признаки извлекаются из первоначального, а затем уже происходит сжатие изображения, чтобы сделать нелинейный пуллинг. После пиксельного анализа изображение увеличивается и происходит обратная свертка, после чего пользователь уже видит изображение сверхвысокого разрешения.

Третья из рассматриваемых нейронных сетей – Sub-Pixel Convolutional Neural Network (далее SubPixel CNN) [11]. Её архитектура представлена на рисунке 3.



**Рисунок 3** – Архитектура нейронной сети Sub-Pixel CNN

Особенность SubPixel CNN заключается в том, что в данной модели каждый цвет обрабатывается на своем сверточном слое, который обрабатывается отдельно от других. На последнем же шаге все признаки по пикселю собираются, в результате чего на выходе можно получить изображение со сверхвысоким разрешением.

### 3. Экспериментальная проверка

В рамках экспериментов настраивались параметры, используемые на стадии обучения нейронной сети, а именно функции активации, оптимизаторы из библиотеки PyTorch и количество эпох обучения. На первом этапе нейронные сети SRCNN, FSRCNN, SubPixel CNN обучались с использованием двадцати шести функций активации. Затем результаты сортировались по метрике PNSR от большего к меньшему. До второго этапа допускались модели, которые заняли первые пять строчек рейтинга. Вне конкурса бралась функция активации ReLU [12], как одна из самых часто встречающихся на GitHub при реализации техники Super-Resolution на сегодняшний день. В рамках второго этапа для каждой из шести выбранных моделей было проведено повторное обучение, но уже с использованием шести оптимизаторов из библиотеки PyTorch. Результаты также были отсортированы по метрике PNSR от большего к меньшему. Три модели, находящиеся первыми в списке, оценивались с точки зрения количества эпох обучения.

В таблице 1 приведены результаты экспериментов с функциями активации, которые получили наибольшее значение PNSR. Как уже было отмечено ранее, использовались двадцать шесть функций активации, заданных в библиотеке PyTorch, а именно: CELU, ELU, GELU, Hardshrink, HardSigmoid, Hardswish, Hardtanh, LeakyReLU, LogSigmoid, LogSoftmax, Mish, PReLU, ReLU, ReLU6, SELU, Sigmoid, SiLU, Softmax, Softmax2d, Softmin, Softplus, Softsign, Softshrink, Tanh, Tanhshrink, из которых было выбрано 5 лучших и ReLU. Значения метрики PNSR в приведенной таблице отсортированы от большего к меньшему.

**Таблица 1** – Выбор функции активации для сверточных нейронных сетей Super-Resolution

SRCNN		FSRCNN		SubPixel CNN	
Функция активации	PNSR	Функция активации	PNSR	Функция активации	PNSR
ReLU	6.5896	GELU	6.3954	Hardswish	6.5107
ReLU6	6.5498	Hardshrink	6.3912	Mish	6.439
Softmax2d	6.4547	LeakyReLU	6.3782	SiLU	6.4261
GELU	6.4427	RReLU	6.378	Tanhshrink	6.3523
LeakyReLU	6.4282	LogSigmoid	6.3768	Softmax2d	6.3416
Hardswish	6.4058	ReLU	6.3667	ReLU	5.409

В таблице 2 представлены результаты экспериментов над оптимизаторами в связке с функциями активации, отобранных на предыдущем этапе. В качестве оптимизаторов

использовались: Adadelta, Adagrad, Adam, Adamax, AdamW, ASGD, NAdam, RAdam, RMSprop, Rprop, SGD. В таблице показана три лучших сочетания, исходя из метрики PNSR.

**Таблица 2** – Выбор оптимизатора для сверточных нейронных сетей Super-Resolution

SRCNN		FSRCNN		SubPixel CNN	
Функция активации и оптимизатор	PNSR	Функция активации и оптимизатор	PNSR	Функция активации и оптимизатор	PNSR
LeakyReLU/Adamax	24.2647	LogSigmoid/Radam	24.2293	ReLU/Adagrad	24.1752
Hardswish/Adamax	24.1443	LeakyReLU/Adamax	24.1902	Mish/Adam	24.1174
ReLU6/Adamax	24.1236	RReLU/Adamax	24.0555	Hardswish/Adamax	24.1650

При исследовании зависимости качества изображения от количества эпох обучения была замечена линейная зависимость: чем больше эпох обучения, тем выше метрика PNSR. В таблице 3 представлены итоговые метрики качества для выбранных архитектур сверточных нейронных сетей, которые в дальнейшем будут использоваться в разделе «Демонстрация работы». Все перечисленные ниже нейронные сети проходили тысячу эпох обучения.

**Таблица 3** – Исследование влияния количества эпох обучения на выбранные параметры сверточных нейронных сетей Super-Resolution

Нейронная сеть	Функция активации	Оптимизатор	PNSR для 1000 эпох обучения
SRCNN	LeakyReLU	Adamax	24.6837
FSRCNN	LogSigmoid	Radam	24.8743
SubPixel CNN	Hardswish	Adamax	24.7264

Таким образом, для дальнейшей работы со сверточными нейронными сетями, были отобраны оптимальные параметры с точки зрения метрики PNSR. Для SRCNN это функция активации LeakyReLU и оптимизатор Adamax, для FSRCNN – LogSigmoid и Radam, для SubPixel CNN – Hardswish и Adamax. Именно эти веса, полученные при 1000 эпохах обучения, будут использоваться в дальнейшем.

## 4. Демонстрация работы

Для сравнения отобранных нейронных сетей с наилучшими параметрами было взято несколько изображений следующих классов: текст, лицо человека, ночь, природный пейзаж, сгенерированное графическое изображение. Так, текст представляет собой контрастное написание символов, чаще всего на белом фоне. Для лица человека характерно наличие глаз, носа и рта, расположенных в определенных местах. При этом некоторые части могут быть скрыты, например, маской или головным убором. Для ночных фотографий характерно обилие темных цветов, нередко наличие яркого акцентного пятна, например, уличного фонаря. Природный пейзаж содержит обилие мелких деталей, таких как листва, трава, горная порода. Обилие синих и зеленых оттенков. Сгенерированные графические изображения (Kandinsky, Midjourney и прочие) отличаются мультипликационным видом.

Получившиеся фрагменты изображений можно увидеть на рисунке 4. Исходные изображения имели форму квадрата. Их размер был уменьшен до 300 пикселей в длину и ширину, а затем восстановлен с помощью нейронных сетей Super-Resolution с тремя разными архитектурами.

Сделаем вывод по получившимся изображениям. Во-первых, модели, которые были построены в рамках исследования выполняют свою функцию, что видно на рисунке 5. Действительно, исходное изображение было увеличено в несколько раз.

Во-вторых, можно заметить, что исходное изображение имеет более крупную пиксельную структуру, чем обработанное нейронной сетью Super-Resolution.

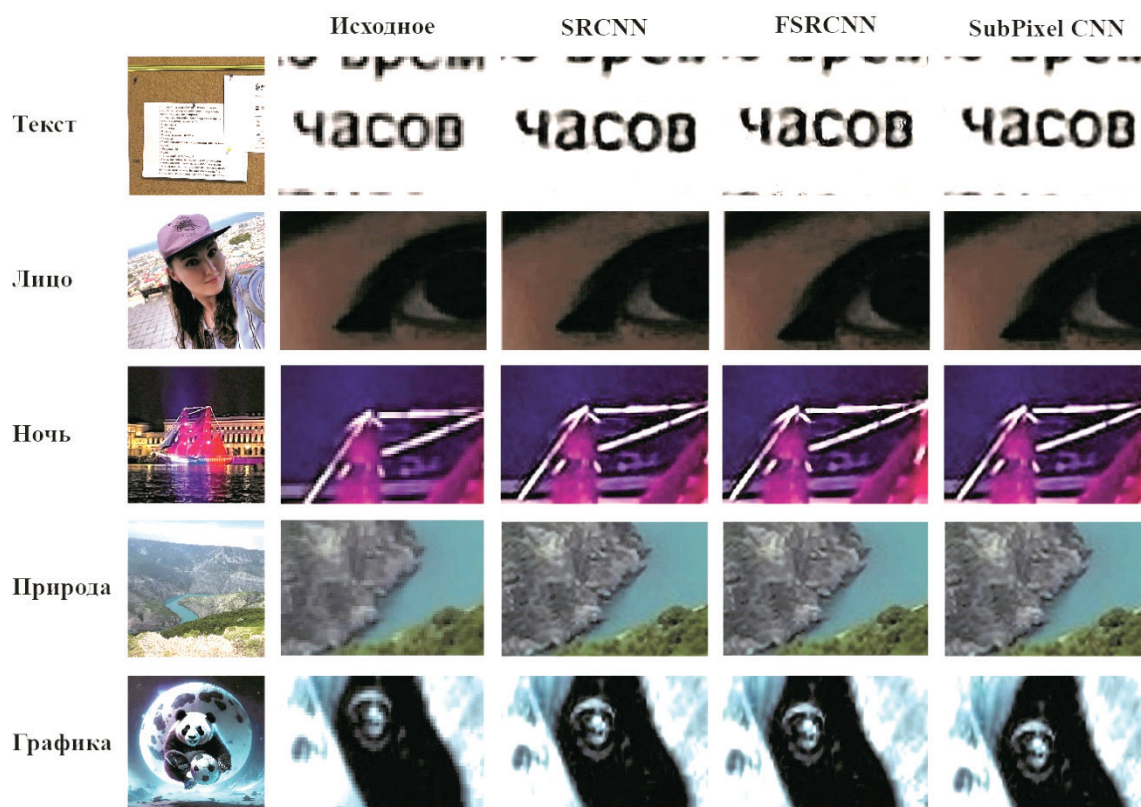


Рисунок 4 – Фрагменты изображений разных классов, обработанных сверточными нейронными сетями

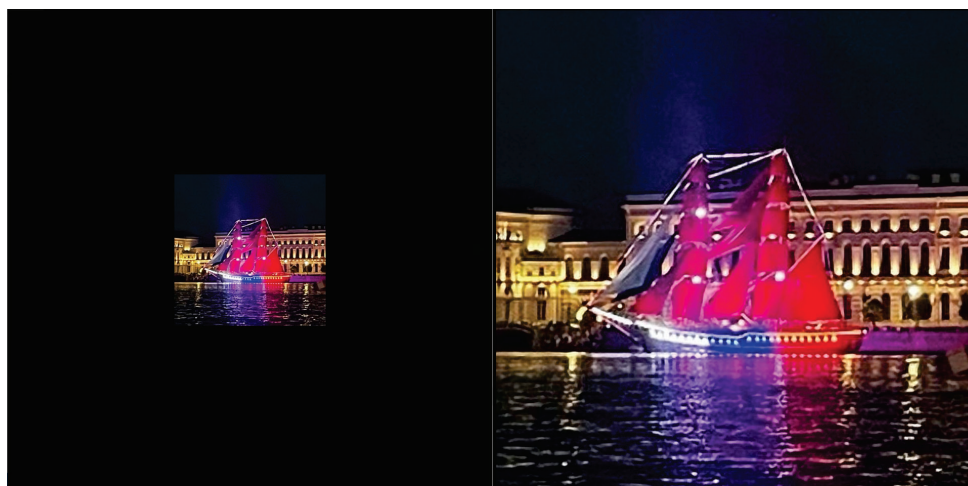


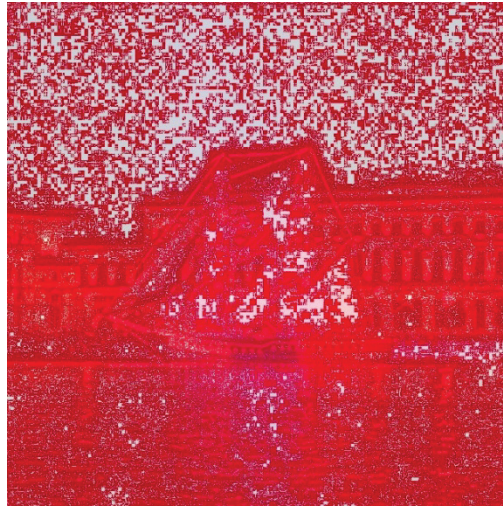
Рисунок 5 – Сравнение размеров исходного изображения (слева) с одним из увеличенных с помощью Super-Resolution (справа). Скриншот экрана

При подробном рассмотрении текста можно заметить, что вокруг букв в FSRCNN присутствует небольшая тень, что делает начертание более размытым. На версии SubPixel CNN присутствуют микроскопические черные вкрапления на белом фоне, которые встречаются чаще, чем в случае SRCNN. Однако и то, и другое позволяет прочитать текст на увеличенной фотографии.

Когда мы говорим о человеческом лице, то видим, что SRCNN делает зону вокруг зрачка вытянутой, что придает неестественность глазу. В случае SubPixel CNN теряется граница между зрачком и макияжем.

Остальные категории (ночь, природа и сгенерированное изображение) обрабатываются без существенной разницы для человеческого глаза, отличаются только насыщенность цветочных каналов RGB.

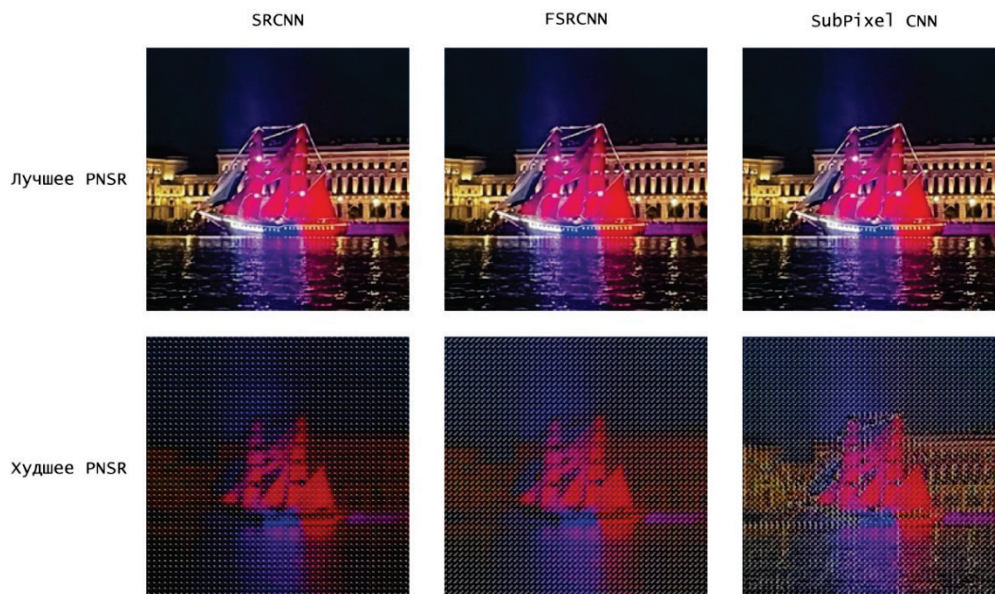
На рисунке 6 красным цветом выделены пиксели, цвет которых различен между выходными фотографиями корабля с красными парусами (например, между SRCNN и FSRCNN). Не смотря на то, что цвет пикселей выходных изображений несколько отличается друг от друга, все три нейронные сети показали хорошие результаты. При первоначальном визуальном анализе нельзя однозначно выявить победителя. Это доказывает то, что параметры при обучении нейронной сети (в нашем случае это функция активации и оптимизатор) намного важнее, чем архитектура (SRCNN, FSRCNN, SubPixel CNN), если не брать в расчет скорость обработки и количество потребляемой памяти при получении изображения Super-Resolution.



**Рисунок 6** – Попиксельное сравнение цветов на увеличенных изображениях

В-третьих, было доказано, что качество работы различных моделей Super-Resolution не зависит от объектов, которые находятся на фото. Если нейронная сеть дает хороший результат, то он будет виден и на текстовых изображениях, и на природных пейзажах.

На рисунке 7 можно увидеть, на сколько сильно поднялось качество выходных изображений, если взять веса с самым низким положительным PNSR и с самым высоким для каждой из трех архитектур. Самые низкие результаты для SRCNN показала функция активации LogSigmoid, для FSRCNN – LogSoftmax, для Sub-Pixel CNN – Hardtanh. Для демонстрации разницы брались веса, которые были получены без оптимизаторов.



**Рисунок 7** – Сравнение изображений с наименьшим положительным и наибольшим значением PNSR

Таким образом, можно сделать вывод, что любая из архитектур CNN при подборе оптимальных параметров на стадии обучения модели при близком значении PNSR (не более двух десятых) выдает практически неразличимые для человеческого глаза результаты. Нейронная сеть Super-Resolution будет увеличивать изображения с одинаковым качеством без привязки к тому, что именно изображено на входной фотографии.

## 5. Выводы

В ходе выполнения работы были последовательно рассмотрены три модели нейронных сетей CNN для решения задачи Super-Resolution (SRCNN, FSRCNN, Sub-Pixel CNN). Проведенные эксперименты с нейронными сетями позволили сделать следующие выводы:

- оптимальные функции активации для моделей следующие: SRCNN – LeakyReLU, FSRCNN – LogSigmoid, Sub-Pixel CNN – ReLU;
- оптимальные оптимизаторы для моделей следующие: SRCNN – Adamax, FSRCNN – Radam, Sub-Pixel CNN – Adagard;
- с увеличением количества эпох метрика качества увеличивается, однако уже примерно после 100-ой эпохи каждый новый виток дает все меньший прирост к PNSR;
- обученная модель будет увеличивать изображения вне зависимости от того, что на них изображено (например, будет одинаково работать и с текстовыми фотографиями, и с природными);
- параметры при обучении сверточной нейронной сети важнее, чем её архитектура, если не брать в расчет скорость увеличения изображения и объем потребляемой памяти.

## 6. СПИСОК ИСТОЧНИКОВ

- [1] Dong, C., Loy, C. C., He, K., Tang, X. (2014). Learning a deep convolutional network for image super-resolution. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13* (pp. 184-199). Springer International Publishing.
- [2] Ren, J. S., Xu, L., Yan, Q., Sun, W. (2015). Shepard convolutional neural networks. *Advances in neural information processing systems*, 28.
- [3] Zhang, K., Zuo, W., Gu, S., Zhang, L. (2017). Learning deep CNN denoiser prior for image restoration. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3929-3938).
- [4] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4681-4690).
- [5] Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Change Loy, C. (2018). Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops* (pp. 0-0).
- [6] Lim, B., Son, S., Kim, H., Nah, S., & Mu Lee, K. (2017). Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 136-144).
- [7] Kim, J., Lee, J. K., & Lee, K. M. (2016). Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1646-1654).
- [8] Haris, M., Shakhnarovich, G., & Ukita, N. (2018). Deep back-projection networks for super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1664-1673).
- [9] Dong, C., Loy, C. C., He, K., & Tang, X. (2015). Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2), 295-307.
- [10] Dong, C., Loy, C. C., & Tang, X. (2016). Accelerating the super-resolution convolutional neural network. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The*



- Netherlands, October 11-14, 2016, Proceedings, Part II 14 (pp. 391-407). Springer International Publishing.
- [11] Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., Wang, Z. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1874-1883).
- [12] Ahn, H., Chung, B., & Yim, C. (2019, January). Super-resolution convolutional neural networks using modified and bilateral ReLU. In 2019 International Conference on Electronics, Information, and Communication (ICEIC) (pp. 1-4). IEEE.