

ТЕХНИЧЕСКИЕ НАУКИ



УДК 631.35.05, 631.359

Принципы использования сверточных нейронных сетей для сжатия изображений с потерями

Р.И. Князев, А.А. Скляренко

Донской государственный технический университет, г. Ростов-на-Дону, Российская Федерация

Аннотация

В данной работе рассматривается метод сжатия изображений с использованием нейросетевой автоэнкодерной модели. Предложенный подход позволяет достигать высокого качества восстановления изображения при значительном уменьшении объема данных. В ходе экспериментов было получено значение среднеквадратичной ошибки (MSE) 0,00118177 и пиковое отношение сигнал/шум (PSNR) 30,75 dB, что превосходит классическое JPEG-сжатие, у которого PSNR для изображений в среднем составляет около 28 dB. Кроме того, благодаря использованию квантования и представлению данных в формате uint8 удалось сократить размер хранимых латентных представлений в четыре раза по сравнению с форматом float32. Это обеспечило средний вес одного монохромного изображения разрешения 64×64 в пределах от 0,6 КБ до 2,1 КБ, в то время как размер исходного изображения составляет 4 КБ. Предложенный метод может быть применим для эффективного хранения изображений с минимальными потерями качества и является перспективным для дальнейшей оптимизации сжатия.

Ключевые слова: сверточные нейронные сети, сжатие изображений с потерями, MNIST, энкодер-декодер, PSNR, MSE

Для цитирования: Князев Р.И., Скляренко А.А. Принципы использования сверточных нейронных сетей для сжатия изображений с потерями. *Молодой исследователь Дона*. 2025;10(5):41–45.

Principles of Using Convolutional Neural Networks for Lossy Image Compression

Ruslan I. Knyazev, Anna A. Sklyarenko

Don State Technical University, Rostov-on-Don, Russian Federation

Abstract

The paper studies an image compression technique by means of an autoencoder type neural network. The proposed approach enables achieving high-quality image restoration upon significant reduction of data volume. During the experiments the following results were obtained — a mean square error (MSE) value of 0.00118177 and a peak signal-to-noise ratio (PSNR) of 30.75 dB, which surpasses classical JPEG compression implying an average PSNR of approximately 28 dB. Furthermore, due to quantization and using uint8 format for representing data, the size of stored latent representations was reduced by four times compared to float32 format. As a result, an average file size of a single 64x64 monochrome image ranged from 0.6 KB to 2.1 KB, whereas the original image size was 4 KB. The proposed technique can be applied to ensure efficient image storage with minimal quality loss and is quite promising for further optimization of compression.

Keywords: convolutional neural networks, lossy image compression, MNIST, encoder-decoder, PSNR, MSE

For Citation: Knyazev RI, Sklyarenko AA. Principles of Using Convolutional Neural Networks for Lossy Image Compression. *Young Researcher of Don*. 2025;10(5):41–45.

Введение. Сжатие изображений занимает ключевую позицию в современных технологиях обработки данных, особенно в условиях ограниченных вычислительных ресурсов и хранения. Традиционные методы, такие как JPEG, применяют дискретное косинусное преобразование (DCT) и энтропийное кодирование, что позволяет эффективно уменьшать объем данных [1]. Однако, помимо JPEG, существуют и другие классические подходы, например, JPEG 2000, который основан на вейвлет-преобразовании. Он позволяет достичь лучшего качества при высокой степени сжатия за счет многомасштабного анализа изображения. Еще одним распространенным методом

является WebP, разработанный Google, использующий предсказательное кодирование, аналогичное применяемому в видеокодеках, а также комбинацию DCT и вейвлетов. Существуют и алгоритмы без потерь, такие как PNG, который использует сжатие на основе LZ77 и кодирование Хаффмана. Несмотря на эффективность этих методов, они часто страдают от артефактов или значительного объема занимаемого дискового пространства. В данной работе рассматривается применение нейросетевого автоэнкодера для сжатия изображений, что обеспечивает лучшее соотношение между качеством восстановления и степенью сжатия.

Основная часть. Целью исследования является повышение эффективности сжатия изображений через использование моделей глубокого обучения, в частности, автоэнкодеров. Основная задача заключается в уменьшении размера изображения при минимальной потере качества и улучшении показателей PSNR [2] по сравнению с традиционными методами.

В работе Ballé и др. «End-to-end Optimized Image Compression» [3] предложено сжатие изображений с использованием вариационных автоэнкодеров (VAE), которые обучаются с учетом распределения латентных переменных. В отличие от традиционных методов, основанных на дискретных преобразованиях, таких как дискретное косинусное преобразование (DCT) в JPEG, этот подход позволяет автоэнкодеру самостоятельно находить оптимальное представление данных в скрытом пространстве. В итоге удается достичь высокой степени сжатия при сохранении деталей изображения, поскольку латентное пространство адаптируется к свойствам изображений, минимизируя потери информации.

В свою очередь, в работе Jiang и др. «An End-to-End Compression Framework Based on Convolutional Neural Networks» [4] авторы предложили применять сверточные нейронные сети (CNNs) для кодирования изображений в компактное представление. В отличие от классических методов сжатия, основанных на жестко заданных алгоритмах, сверточные сети обучаются на реальных данных и могут адаптивно выделять ключевые признаки изображения. Это позволяет им кодировать значимую информацию более эффективно, исключая ненужные детали, которые минимально влияют на визуальное восприятие. Такой подход обеспечивает более высокое качество восстановления по сравнению с JPEG, особенно на сложных изображениях с текстурами и градиентами, где традиционные методы могут создавать артефакты.

В качестве модели сжатия использован сверточный автоэнкодер, состоящий из двух ключевых компонентов: энкодера и декодера. Энкодер принимает изображение в качестве входных данных и постепенно преобразует его в компактное латентное представление, а декодер выполняет обратное преобразование, восстанавливая изображение из этого представления.

Энкодер включает несколько сверточных слоев, каждый из которых выполняет операцию свертки, выявляя ключевые признаки изображения и уменьшая пространственное разрешение за счет использования шагов и слоев предвыборки. На каждом этапе размерность изображения уменьшается, но глубина (число каналов) возрастает, что позволяет нейросети выделять более сложные и абстрактные признаки.

Ключевой особенностью модели является латентное представление – компактный вектор, содержащий сжатую информацию об изображении. Оно формируется на последнем слое энкодера и имеет размерность 32, 24 или 12 в зависимости от желаемой степени сжатия. Латентное представление можно рассматривать как числовую запись наиболее значимых признаков изображения, позволяющую его реконструировать. В этом пространстве модели проще работать с изображениями, так как оно более компактное и сохраняет всю необходимую информацию для восстановления.

Декодер выполняет обратное преобразование: используя транспонированные свертки (Transposed Convolutions) или операции увеличения размера (Upsampling), он постепенно восстанавливает исходное изображение из латентного представления. На каждом этапе разрешение увеличивается, а сверточные фильтры помогают воссоздать детали изображения. В финальном слое декодера применяется свертка с функцией активации sigmoid, что обеспечивает нормализацию значений пикселей в диапазоне от 0 до 1.

Были реализованы несколько модулей, обеспечивающих гибкость и удобство использования предложенного метода сжатия изображений:

1. Модуль построения модели:

- Реализована архитектура сверточного автоэнкодера, включая энкодер и декодер.
- Поддерживается настройка параметров модели, таких как размер латентного представления и гиперпараметры обучения.
- Внедрены механизмы нормализации (Batch Normalization) и функции активации (Leaky ReLU) для повышения стабильности обучения.

2. Модуль переиспользования модели:

- Обеспечена возможность загрузки предобученной модели без необходимости повторного обучения;
- Реализованы функции для кодирования и декодирования изображений через обученный автоэнкодер;
- Поддержка работы с различными форматами входных данных (например, jpg, png, jpeg, npz).

3. Модули анализа результатов:

- Реализованы инструменты для сравнения качества сжатия с традиционными методами (JPEG, PNG, BMP);
- Встроены метрики оценки качества восстановления PSNR;
- Автоматизирован сбор и визуализация статистики.

4. Функции-хэлперы:

- Разработаны функции для ресайза изображений, позволяющие тестировать модель на разных разрешениях (например, 28×28 , 64×64 , 128×128);

– Поддержка автоматического предварительного преобразования изображений перед подачей в автоЭнкодер.

Эти модули обеспечивают удобство работы с моделью, ее переиспользование и детальный анализ полученных результатов. Благодаря такому подходу возможно быстрое переиспользование больших датасетов под необходимое разрешение.

Для экспериментов использовались монохромные изображения размером 64×64 пикселя. Чтобы сжатие было более заметным, изображения из набора Fashion MNIST (исходно 28×28) были увеличены до 64×64 . Размерность латентного представления выбрана по умолчанию равной 32, поскольку уменьшение этого параметра приводило к ухудшению качества восстановления (снижению PSNR), а увеличение — к росту размера файла без значительного улучшения качества. Однако при меньшем латентном представлении в 24 или 12 качество восстановления (PSNR) опускается до 30,75, хотя и достигается выигрыш по объему памяти.

Проведенные эксперименты показали, что предложенный метод обеспечивает $MSE = 0,0008177$ и $PSNR = 30,75$ дБ при латентном представлении в 12 ед., что уступает показателям JPEG-сжатия, у которого $PSNR \approx 35$ дБ при среднем сжатии, но становится равным при агрессивном сжатии, из-за которого качество восстановления JPEG опускается до 29,8.

Сперва были проведены эксперименты для латентного представления в 32. Результаты работы декодирования латентного представления приведены на рис. 1.

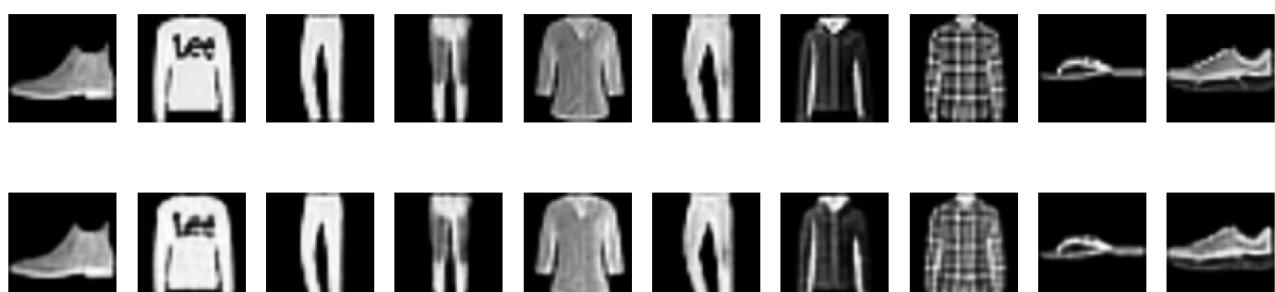


Рис. 1. Сравнение исходных изображений (сверху) с восстановленными (снизу) при латентном представлении в 32 ед.

Также удалось добиться уменьшения размера файла: средний вес закодированного изображения при латентном представлении (л. п.) в 12 единиц составляет 0,6 КБ, в то время как средний размер сжатия JPEG составляет 1 КБ. Это стало возможным благодаря дополнительным манипуляциям — квантованию латентного представления. Сравнение занимаемой памяти каждого из методов представлено на рис. 2.

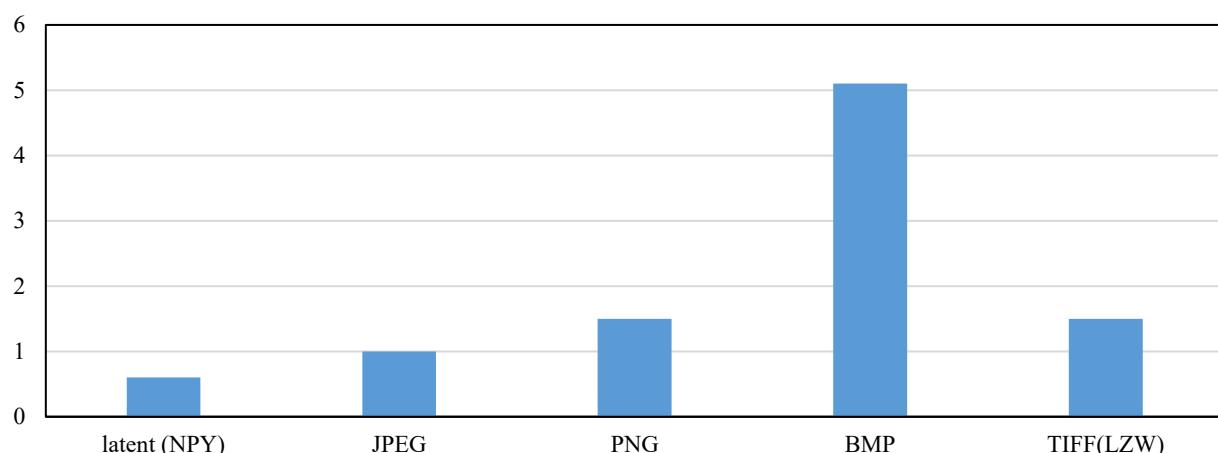


Рис. 2. Сравнение существующих методов с латентным представлением в 12 ед.

При латентном представлении в 12 единиц восстановленные изображения незначительно теряют в качестве. Значение PSNR находится на приемлемом уровне для распознавания, однако символы читаются хуже. Пример восстановления при заданном гиперпараметре представлен на рис. 3.

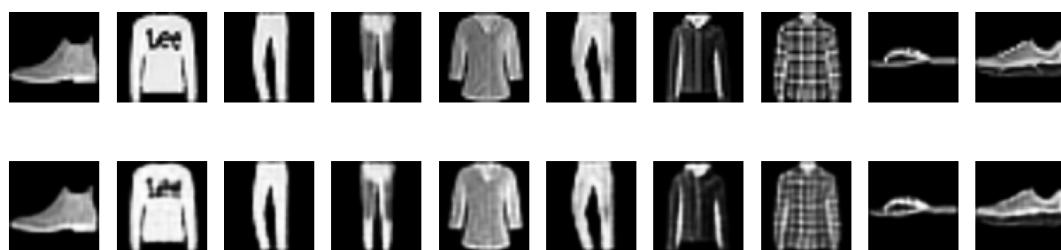


Рис. 3. Сравнение исходных изображений (сверху) с восстановленными (снизу) при латентном представлении в 12 ед.

Агрессивное сжатие JPEG может весить на 0,2 КБ меньше нашего латентного представления. Их объемы занимаемой памяти показаны на рис. 4.

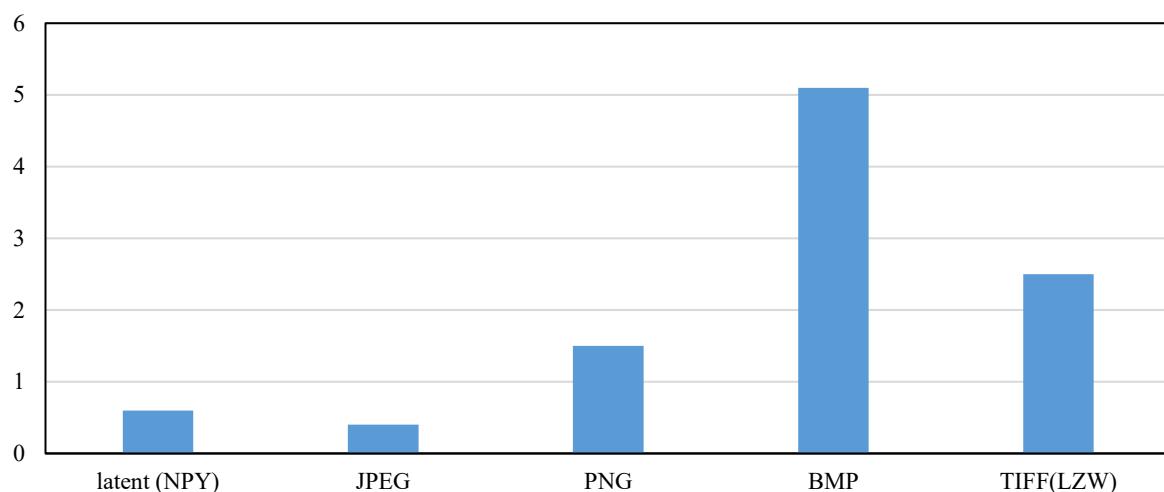


Рис. 4. Сравнение размеров файлов при агрессивном сжатии JPEG

Однако, при таком сценарии уже в оценке PSNR наш подход показывает лучший результат по сравнению с JPEG. Пример сравнения приведен на рис. 5.

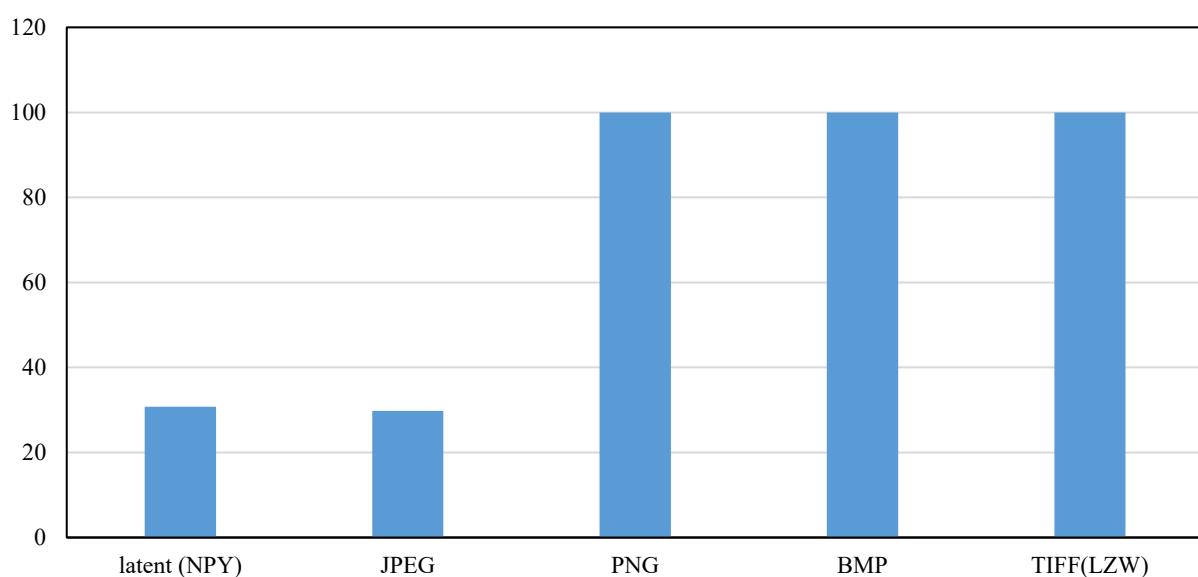


Рис. 5. Сравнение оценки качества восстановленного изображения при агрессивном сжатии JPEG и латентном представлении в 12 ед.

Заключение. Результаты указывают на перспективность нейросетевого подхода для сжатия изображений. Достигнутое качество предоставляет возможность применения предложенного метода в разнообразных областях, где важно сохранение деталей. Например:

- медицинская визуализация — сжатие рентгеновских снимков, КТ и МРТ, где критически важны мельчайшие детали изображения;
- спутниковые и аэрофотоснимки — уменьшение объема данных при передаче и хранении спутниковых изображений без потери информативности;
- архивирование научных данных — компактное хранение изображений с высоким разрешением, таких как в астрономии или микроскопии;
- удаленный доступ к графическим данным — передача сжатых изображений в облачных сервисах, например, при обработке фотографий и видео в реальном времени;
- видеоигры и 3D-графика — хранение текстур и карт нормалей с высококачественным представлением при минимальном объеме.

При этом остается потенциал для дальнейшего улучшения. В частности, возможно использование дополнительных методов сжатия к латентному представлению, таких как энтропийное кодирование (Huffman, арифметическое кодирование) или квантование с переменной точностью, что позволит значительно сократить размер файлов без заметной потери качества.

Список литературы

1. Wallace GK. The JPEG Still Picture Compression Standard. *IEEE Transactions on Consumer Electronics*. 1992;38(1):xviii–xxxiv. <https://doi.org/10.1109/30.125072>
2. Huynh-Thu Q, Ghanbari M. Scope of Validity of PSNR in Image/Video Quality Assessment. *Electronics Letters*. 2008;44:800–801 URL: <https://www.scirp.org/reference/referencespapers?referenceid=2149300> (acccesed: 02.04.2025).
3. Ballé J, Laparra V, Simoncelli EP. End-to-end Optimized Image Compression. *arXiv*. 2017;1611.01704v. <https://doi.org/10.48550/arXiv.1611.01704>
4. Tao W, Jiang F, Zhang S, Ren J, Shi W, Zuo W. An End-to-End Compression Framework Based on Convolutional Neural Networks. In: *Proceedings of 2017 Data Compression Conference (DCC)*. Snowbird, USA; 2017. P. 463–463. <https://doi.org/10.1109/DCC.2017.54>

Об авторах:

Руслан Ишханович Князев, магистрант кафедры «Программное обеспечение вычислительной техники и автоматизированных систем» Донского государственного технического университета (344003, Российская Федерация, г. Ростов-на-Дону, пл. Гагарина, 1), knyazevseven@gmail.com

Анна Анатольевна Скляренко, кандидат технических наук, доцент кафедры «Программное обеспечение вычислительной техники и автоматизированных систем» Донской государственный технический университет (344003, Российская Федерация, г. Ростов-на-Дону, пл. Гагарина, 1), Deranna1982@yandex.ru

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Все авторы прочитали и одобрили окончательный вариант рукописи.

About the Authors:

Ruslan I. Knyazev, Master's Degree Student of the Software for Computer Engineering and Automated Systems Department, Don State Technical University (1, Gagarin Sq., Rostov-on-Don, 344003, Russian Federation), knyazevseven@gmail.com

Anna A. Sklyarenko, Cand.Sci.(Engineering), Associate Professor of the Software for Computer Engineering and Automated Systems Department, Don State Technical University (1, Gagarin Sq., Rostov-on-Don, 344003, Russian Federation), Deranna1982@yandex.ru

Conflict of Interest Statement: the authors declare no conflict of interest.

All authors have read and approved the final manuscript.