

Адаптивный алгоритм сжатия на основе JPEG2000 с нейросетевым корректором четкости декодированных изображений

С.В. Сай¹, В.С. Никонов¹, Е.С. Фомина¹

¹ Тихоокеанский государственный университет,
680035, Россия, г. Хабаровск, ул. Тихоокеанская, д. 136

Аннотация

В статье приводится описание особенностей цифровой обработки изображения в процессе адаптивного сжатия на основе дискретного вейвлет-преобразования, используемого в стандарте сжатия JPEG2000. В отличие от известного алгоритма сжатия в новом алгоритме используется уменьшение (масштабирование) размеров матриц коэффициентов дискретного вейвлет-преобразования сигналов Y, U и V на первой итерации дискретного вейвлет-преобразования, после которой для дальнейшей обработки использована уменьшенная копия изображения низкочастотного диапазона. Таким образом, на этом этапе объем видеоданных сокращается в 4 раза. Для получения исходного разрешения в отличие от известных методов интерполяции использовано обратное преобразование дискретного вейвлет-преобразования с добавленными нулевыми коэффициентами в высокочастотные субдиапазоны первой итерации. Предлагаемый алгоритм позволяет сжимать файлы изображений в среднем в 40...50 раз с удовлетворительным качеством. Для восстановления высокого качества изображений разработан оригинальный нейросетевой корректор четкости, основанный на сверточной модели с обучением выборочных блоков изображений по индексу яркости в пространстве Lab. На базе архитектуры ResNet разработана собственная модель глубокой нейронной сети, основанная на объединении нескольких методик, используемых для решения задач реконструкции изображений в других архитектурах. Выбран оптимальный вариант обучения нейронной сети, позволяющий использовать обученную модель для коррекции четкости декодированных изображений до высоких объективных показателей качества с субъективными оценками «хорошо» и «отлично».

Ключевые слова: корректор четкости, анализ изображений, метрика искажений, дискретное вейвлет-преобразование, эффективность сжатия, нейросеть.

Цитирование: Сай, С.В. Адаптивный алгоритм сжатия на основе JPEG2000 с нейросетевым корректором четкости декодированных изображений / С.В. Сай, В.С. Никонов, Е.С. Фомина // Компьютерная оптика. – 2025. – Т. 49, № 6. – С. 986-993. – DOI: 10.18287/2412-6179-CO-1586.

Citation: Sai SV, Nikonov VS, Fomina ES. Adaptive compression algorithm based on JPEG2000 with neural network corrector of decoded image definition. Computer Optics 2025; 49(6): 986-993. DOI: 10.18287/2412-6179-CO-1586.

Введение

В современных интеллектуальных видеосистемах проблема сохранения качества передаваемых деталей изображения является одной из основных проблем при решении задач поиска, распознавания и идентификации объектов. Увеличение количества пикселей формата видеокadra до стандартов высокой четкости (HD, 4K и 8K) приводит к повышению разрешения изображения, но при этом для обеспечения необходимой скорости передачи требуется применение более эффективных методов сжатия. Одним из способов сокращения объема видеоданных является его масштабирование, т.е. уменьшение размера изображения с последующим восстановлением исходного разрешения.

Цифровые изображения подвержены воздействию искажений, которые появляются в процессе формирования, цифровой обработки и передачи сигналов. Основные искажения качества возникают за счет

использования сжатия с потерями, что обусловлено необходимостью передачи сигналов по каналам связи с ограниченной пропускной способностью, а также компактным хранением больших объемов видеоданных. Другой причиной искажений является масштабирование изображения в формат с более высоким разрешением.

Основными показателями качества передачи изображения является его четкость и резкость, которые зависят от разрешающей способности (разрешения) видеосистемы. В процессе цифровой обработки (сжатие, увеличение формата) «размываются» или «теряются» мелкие детали и ухудшается резкость границ крупных деталей, что сильно ухудшает визуальное качество изображения.

В целом, получаемое изображение Y может быть описано как $Y = F(X)$, где X – это оригинальное изображение с выхода высококачественной камеры и F – функция потерь. Основная проблема заключается в

том, что даже зная искажающие характеристики цифровых преобразований, мы не сможем с высокой точностью найти обратную функцию F , с помощью которой можно было бы точно восстановить оригинальное изображение без потерь.

Задачу восстановления или повышения качества изображения можно решать следующими способами: повышением качественных характеристик методов и алгоритмов сжатия; применением корректирующих фильтров; использованием технологий искусственного интеллекта (AI) на основе нейронных сетей.

В настоящее время перспективным направлением исследований является создание новых методов и алгоритмов повышения качества изображений на основе свёрточных нейронных сетей. Нейронные сети могут использоваться на этапе предварительного анализа и распознавания участков оригинального изображения, таких как границы деталей, мелкие структуры, фоновые участки. Разбиение на участки позволяет более эффективно сжимать изображение с применением методов адаптивного кодирования. На этапе восстановления (реконструкции) изображения нейросетевые методы позволяют корректировать искажения сжатия на основе обученной сети. В [1] рассмотрены нейросетевые способы коррекции искажений JPEG для низких параметров качества Q и, следовательно, при высоких коэффициентах сжатия. В частности, показано [1], что при значениях $Q < 20\%$ свёрточные нейронные сети позволяют улучшить качество изображений не более чем на 1,5 дБ по метрике PSNR.

В работе [2] предложен адаптивный алгоритм сжатия на основе JPEG2000. Способ адаптивного квантования основан на регулировке параметров шкалы квантования в каждом субдиапазоне в зависимости от классификации блоков исходного изображения. Применение адаптивного квантования позволяет повысить эффективность сжатия изображения низкой и средней детальности до 30–40% без ухудшения четкости и резкости по объективным критериям [3]. При условии обеспечения высокой четкости ($Q \geq 80\%$) коэффициент сжатия адаптивного алгоритма не превышает 10...20 раз, и при этом коррекция качества не требуется.

Повышение коэффициента компрессии возможно за счет снижения параметра качества Q алгоритма сжатия. В частности, снижение параметра Q в адаптивном алгоритме с 80% до 40% приводит к повышению коэффициента сжатия примерно в два раза. Однако при этом искажения существенны и заметны глазом. В этом случае требуется применение корректора качества.

В работе [4] приведен обзор последних достижений в области сверхвысокого разрешения (SR) изображений с использованием подходов глубокого обучения. На сайте [5] приведено описание лучших приложений повышения качества изображений с помощью SR-методов, таких как ImgLarger, Let's Enhance, Image Upscale, Bigjpg, Deep Image, PhotoLemur и др.

К особенностям приложений улучшения качества изображений с использованием AI относится то, что они в основном используются для повышения разрешения [6], при этом их применение не дает положительных результатов в задачах коррекции нелинейных искажений, возникающих за счет грубого квантования коэффициентов дискретного вейвлет-преобразования (ДВП) при высоких коэффициентах сжатия.

В статье предлагается оригинальный алгоритм сжатия и нейросетевой корректор четкости, обученный для этого алгоритма, позволяющие обеспечить высокий коэффициент сжатия и коррекцию четкости изображений до высоких объективных показателей качества с субъективными оценками «хорошо» и «отлично».

1. Алгоритм сжатия

За основу алгоритма взят адаптивный JPEG2000. В отличие от предыдущего варианта [2] в алгоритм внесены следующие изменения:

1) На входе кодера составляющие изображения YUV масштабируются в два раза, т.е. выполняется переход из модели 4:4:4 в 2:2:2. В отличие от традиционных способов масштабирования в новом алгоритме использована первая итерация ДВП, после которой для дальнейшей обработки использована уменьшенная копия изображения низкочастотного диапазона. Таким образом, на этом этапе объем видеоданных сокращается в 4 раза.

2) Уменьшенная копия изображения кодируется с помощью прямого преобразования ДВП и поступает на вход адаптивного квантователя. В процессе квантования устанавливается параметр качества Q не менее 70%, что позволяет минимизировать ошибки. Отметим, что шкала квантования не изменилась относительно исходного изображения и при этом для масштабированной копии использованы коэффициенты квантования, соответствующие второй итерации ДВП.

3) В декодере после операций деквантования и обратного ДВП восстанавливается уменьшенная в два раза копия изображения. Для получения исходного разрешения в отличие от известных методов интерполяции использовано обратное преобразование ДВП с добавленными нулевыми коэффициентами в высокочастотные субдиапазоны.

Отметим, что масштабирование в два раза на входе кодера не означает повышение коэффициента сжатия точно в 4 раза относительно коэффициента сжатия адаптивного JPEG2000 [2], что обусловлено сохранением шкалы квантования коэффициентов ДВП для исходного разрешения изображения.

Для демонстрации результатов оценки качества адаптивного нового алгоритма сжатия выбраны тестовые изображения из популярного набора DIV2K [7]. Набор DIV2K содержит 800 оригинальных фотоизображений высокого качества с различной детальностью

и с разрешением не менее 2040×1356 пикселей. На рис. 1 показаны примеры вырезанных фрагментов с размером 1024×1024 пикселей девяти тестовых изображений.

В табл. 1 приведены коэффициенты сжатия ($K_{сж}$) и результаты объективных оценок качества тестовых изображений в формате 2K на выходе декодера, где для оценки качества использовались традиционная метрика PSNR и дополнительно авторская метрика MFSD (метрика искажений мелких структур) [3].

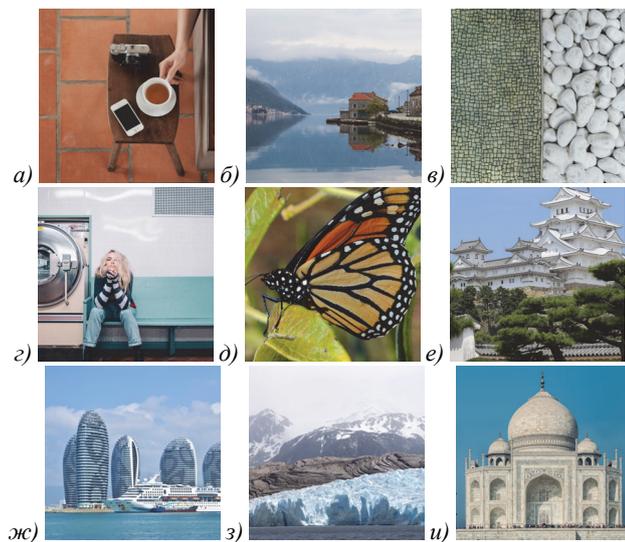


Рис. 1. Фрагменты тестовых изображений

Табл. 1. Оценка эффективности сжатия и качества тестовых изображений

Im	$K_{сж}$	PSNR	MFSD
a	50	33,4	1,70
б	63	32,7	1,94
в	25	30,8	1,44
г	39	31,3	1,69
д	53	34,8	1,53
e	36	30,3	1,70
ж	44	29,6	2,04
з	45	32,8	1,37
и	41	27,6	2,76
M	44	31,5	1,79

В отличие от других метрик, MFSD использует локальное сравнение изменений цветового контраста между соседними пикселями с учетом зрительных порогов. Теоретически и экспериментально доказано [3], что при значении $MFSD \leq 0,5$ искажения незаметны для глаза с оценкой «отлично» по пятибалльной шкале качества. При значении $MFSD \leq 0,8$ искажения мало заметны для глаза с оценкой «хорошо». Если $MFSD > 0,8$, то качество изображения соответствует оценкам «удовлетворительно» или «плохо». Такая градация позволяет объективно и эффективно оценивать искажения изображений после сжатия или других способов обработки. По метрике PSNR хорошее и отличное качество изображения обеспечивается при значении $PSNR > 36$ дБ.

В результате исследований получено, что новый алгоритм позволяет повысить коэффициент сжатия в 2...3 раза по сравнению с предыдущим вариантом [2] адаптивного JPEG2000, однако по объективным критериям качество изображений невысокое и соответствует субъективной оценке «удовлетворительно».

Такое же повышение коэффициента сжатия можно выполнить с помощью уменьшения параметра качества Q в стандартном алгоритме JPEG2000, реализованном в приложениях Photoshop, ACDSee, OpenCV и др. Поэтому возникает вопрос – зачем использовать масштабирование и последующую интерполяцию в новом алгоритме?

Для визуальной оценки искажений на рис. 2 показан фрагмент изображения рис. 1*и* до и после сжатия. Также на рис. 2*в* показан фрагмент тестового изображения, сжатого в 40 раз с помощью JPEG2000 фоторедактора Adobe Photoshop CS6 Extended. Оценка искажений оригинального изображения (рис. 1*и*) после сжатия Adobe Photoshop в формат *.jpf соответствует следующим значениям $PSNR = 31,8$ дБ и $MFSD = 1,34$. Сравнение с новым алгоритмом (табл. 1) показывает, что эти параметры лучше, чем для нового алгоритма ($PSNR = 27,6$ дБ и $MFSD = 2,76$), но также не обеспечивают высокого качества.



Рис. 2. Фрагмент изображения рис. 1*и*: а) оригинал; б) новый алгоритм; в) сжатие в Photoshop SC6

В результате исследований получено, что высокий коэффициент сжатия за счет грубого квантования (обнуления) коэффициентов ДВП в стандартном алгоритме JPEG2000 приводит к увеличению нелинейных искажений и к существенным потерям, которые практически не восстанавливаются с помощью глубоких нейронных сетей. В новом алгоритме шкала квантования коэффициентов ДВП не изменяется, т.е. нелинейные искажения минимальны. При этом вносятся искажения, обусловленные процессами масштабирования и интерполяции (рис. 2*б*). Такие искажения можно скорректировать с помощью современных систем искусственного интеллекта на основе глубоких нейронных сетей.

Для повышения качества изображения предложено использовать нейросетевой корректор четкости. Возможны два пути построения корректора:

- 1) восстановление уменьшенной копии изображения на выходе декодера до исходного разрешения с помощью SR-моделей, применяемых для повышения разрешения в два раза;
- 2) коррекция качества изображения после интерполяции уменьшенной копии до исходного разрешения с помощью новых моделей.

2. Корректор четкости

Рассмотрим первый вариант корректора. В настоящее время существуют нейросетевые модели, которые относятся к категории SISR – Single Image Super-Resolution, реализующие механизм восстановления качества масштабируемого одиночного изображения. К примерам таких моделей относятся ESRGAN, BSRGAN, SwinIR, AI Image Upscale, Waifu2x, Neural Love, Image Enlarger, Let's Enhance, BigJPG, Topaz Gigapixel AI, Deep Image и др.

В моделях SISR используются разные архитектуры, включая такие модели, как генеративно-сопоставительная сеть (GAN), трансформеры и др. Для сравнения выбраны три популярные модели: BSRGAN [8], ESRGAN [9], SwinIR [10].

BSRGAN – модель GAN, отличительной особенностью которой является комбинация случайных характеристик размытия, понижения дискретизации и снижения уровня шума. Размытие аппроксимируется двумя свертками с ядрами Гаусса, применяется билинейная и бикубическая интерполяция, шумовая составляющая обрабатывается с помощью конвейерной обработки изображения.

ESRGAN – еще одна модель GAN, отличающаяся применяемой моделью деградации для более эффективного применения этой сети на реалистичных изображениях, также используется модификация с применением дискриминатора U-Net для стабилизации динамики обучения.

SwinIR – нейронная сеть, построенная на архитектуре трансформера: в такой архитектуре сведены к минимуму последовательные вычисления. Работа SwinIR состоит из трех частей: извлечение поверхностных признаков на изображении, выделение «глубоких» признаков и реконструкция изображения на основе выделенных признаков.

Для исследования эффективности SR-моделей BSRGAN, ESRGAN и SwinIR использовались готовые модули с сайтов разработчиков. На входы обученных моделей подавались уменьшенные в два раза копии сжатых изображений после декодирования новым адаптивным алгоритмом.

В качестве примера использованы тестовые изображения, показанные на рис. 1. На рис. 3 показан фрагмент тестового изображения (рис. 1а) после увеличения разрешения в два раза с помощью SR-моделей. В сводной табл. 2 приведены результаты оценки качества для тестовых изображений рис. 1.

В результате тестирования получено, что лучшие показатели качества восстановленных изображений обеспечивает модель ESRGAN. Однако в среднем параметры четкости ниже установленных критериев по метрикам $MFSD \leq 0,8$ и $PSNR > 36$ дБ. В частности, после восстановления разрешения тестового изображения (рис. 1а) с помощью ESRGAN, параметры

качества $PSNR = 24,6$ дБ и $MFSD = 2,96$ только ухудшились по сравнению с данными в табл. 1.



Рис. 3. Фрагмент изображения рис. 1а после увеличения масштаба в два раза. а) BSRGAN, б) ESRGAN, в) SwinIR

По субъективным оценкам качество восстановленных изображений «удовлетворительное», т.е. не лучше, чем на входе модели.

В результате исследований получено, что SR-модели не улучшают качество изображения после адаптивного сжатия. Это объясняется тем, что они разработаны и обучены для повышения разрешения изображения, например, в 2 или 4 раза. Обучающая выборка, например, DIV2K (LR) представляет собой уменьшенные копии оригинальных изображений, и нейросеть в процессе обучения старается уменьшить ошибки интерполяции, возникающие в процессе увеличения масштаба. При этом артефакты сжатия здесь не учитываются.

Для повышения качества изображений до оценок «хорошо» и «отлично» разработана собственная модель, основанная на объединении нескольких методик, использованных для решения задач реконструкции изображений при разработке других архитектур, таких как MemNet [11], RDN [12] и EDSR [13].

Новая архитектура представляет собой сверточную глубокую сеть, построенную на основе архитектуры ResNet [14]. Разработанная новая архитектура получила название NeuroCorrector.

В сети присутствуют только локальные пути обхода в блоках нейросети. Блоки нейросети соединены между собой последовательно. Каждый из блоков имеет отдельный путь для соединения с блоком обработки иерархических признаков. Выход этого блока соединен с блоком восстановления изображений. Каждый из сверточных блоков состоит из 64 фильтров с размером ядра свертки 3 на 3 элемента. Блок нейросети представляет собой упрощенный блок из нейросети ResNet, состоящий из двух сверточных слоёв и пути обхода (Residual Path). От нейросетевой архитектуры EDSR был заимствован ([15], с. 532) упрощенный ResNet-блок. Данный блок изображен на рис. 4.

TensorFlow-версия новой модели состоит из 6 блоков Neuroglass Block. Каждый сверточный слой содержит 64 фильтра.

В версии для NeuroCorrector сверточный блок с ядром свертки 1 на 1 элемент заменен на сверточный блок с размером ядра свертки 3 на 3 элемента. Также в версии NeuroCorrector не содержит пути обхода на

уровне блока (Local Residual Path). Версия блока, реализованная в NeuroCorrector, показана на рис. 5.

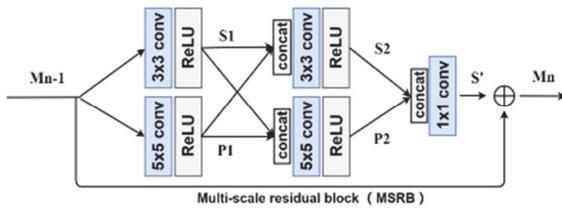


Рис. 4. Residual Block

От архитектур MemNet и RDN заимствован метод соединения блоков между собой, а также архитектура блока обработки иерархических признаков. Количество параметров сети для шести блоков равно 5 299 329.

Результаты исследования модели NeuroCorrector привели к созданию нового метода, позволяющего выполнять коррекцию изображения до высокого качества и при этом уменьшить количество вычислительных операций.

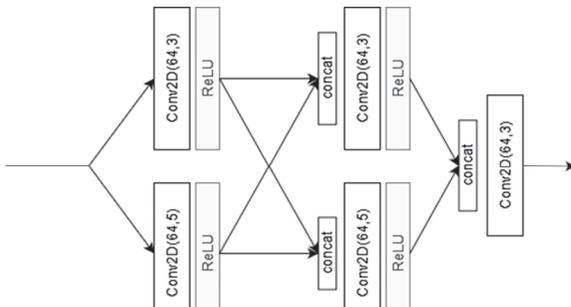


Рис. 5. NeuroCorrector Block

Суть метода заключается в том, что для обеспечения высокого качества изображений достаточно корректировать только яркостную составляющую Y . При этом интерполированные сигналы цветности U и V можно оставлять без изменений. Это утверждение обусловлено тем, что артефакты сжатия адаптивным алгоритмом [2], возникающие в каналах U и V , оказывают незначительное влияние на визуальное качество за счет особенностей зрения при выбранных параметрах квантователя.

Метод подготовки обучающей выборки реализован в виде следующего алгоритма:

- 1) преобразование тестового и искаженного изображения из цветового пространства RGB в пространство Lab [16] с последующей обработкой только яркостной составляющей L ;
- 2) разбиение изображения на блоки с заданным размером B_{size} , например 16×16 или 32×32 пикселей;
- 3) классификация блоков на фоновые и детальные и выделение (маркировка) детальных блоков;
- 4) вычисление среднего значения для каждого детального блока и получение разностных значений пикселей относительно среднего;
- 5) нормировка значений пикселей блока до диапазона $[-1, 0 \dots 1, 0]$;

- 6) случайное перемешивание всех блоков тестового и искаженного изображения;
- 7) формирование обучающей выборки для одного тестового и искаженного изображения с заданным количеством блоков N_B ;
- 8) подготовка обучающей выборки из выбранного набора тестовых изображений.

К особенностям подготовки обучающей выборки и обучения модели относится следующее. Количество блоков на входе сети (data train) определяется простым уравнением $N_{Input} = N_{Im} \times N_B$, где N_{Im} – количество изображений. Такой же размер данных необходим на выходе сети (data test).

Например, при использовании 800 изображений из набора DIV2K и $N_B = 256$ количество блоков в обучающей выборке будет равно $N_{Input} = 204800$. Отметим, что увеличение количества блоков ограничено вычислительными возможностями используемой видеокарты.

В качестве функции потерь $Loss$, в отличие от традиционной метрики MSE , использовано среднее значение модуля отклонения индекса яркости L в каждом блоке. Такое решение обусловлено тем, что цветовое пространство Lab используется в метрике MFSD [3] для объективной оценки цветовых различий деталей изображений. Поэтому выбранная функция $Loss$ позволяет более объективно оценивать потери качества изображения по индексу яркости L в процессе обучения модели.

Подготовка обучающей выборки отличается от известных методик следующими пунктами: обработка только яркостной составляющей L ; использование собственной функции потерь $Loss$; обучение блоками с небольшими размерами (32×32 пикселей); обучение только детальных блоков.

При обучении модели использовался метод Model.fit [17] с заданными параметрами перемешивания (shuffle = True) и процентом валидационных блоков (validation_split = 0,2). Для получения наилучших значений параметров модели использован метод ModelCheckpoint [17].

В процессе исследований необходимо было подобрать такие обучающие параметры, которые бы обеспечили лучшие показатели коррекции качества декодированных изображений. С этой целью изменялись: количество изображений N_{Im} в обучающей выборке, количество блоков N_B в одном изображении, размер блока B_{size} , тип оптимизатора и коэффициент обучения L_r , размер пакета (batch_size), количество эпох.

Экспериментальные исследования выполнены с помощью современных вычислительных систем на ПК с многоядерными процессорами и оперативной памятью не ниже 24 Гбайт. Для обучения и реализации нейронной сети использовались современные видеокарты (Nvidia RTX 2060, 3060) с установленным программным обеспечением CUDA [18] и TensorFlow [19].

В результате исследований были выбраны следующие параметры: количество блоков в одном изображении $N_B = 1024$; размер блока $B_{size} = 32 \times 32$ пикселей; размер пакета $batch_size = 16$; оптимизатор Nadam [19] с коэффициентом обучения $L_r = 0,0001$; количество эпох – 100.

Получено, что увеличение количества N_B анализируемых блоков в одном изображении до 1024 приводит к повышению качества, однако при большой обучающей выборке требуется значительное увеличение вычислительной мощности. Поэтому обучение модели выполнялось не по всему набору из 800 изображений DIV2K, а наборами из меньшего количества изображений с последовательной оптимизацией весовых коэффициентов от набора к набору.

Отметим, что эти параметры подобраны экспериментально только для набора изображений из DIV2K. Поэтому они могут быть использованы в качестве рекомендованных начальных параметров для других тестовых наборов изображений при формировании обучающей выборки.

В сводной табл. 2 приведены результаты тестирования новой модели, из которых следует, что разработанный нейросетевой корректор обеспечивает высокое качество декодированных изображений по критерию MFSD с визуальными оценками «отлично» и «хорошо». На рис. 6 показан фрагмент тестового изображения рис. 1*u*, иллюстрирующий качество изображения после коррекции. Анализ полученных результатов показывает, что для тестового набора изображений

улучшение качества по метрике PSNR составляет около 5 дБ и по метрике MFSD на 1,0.

Для сравнения в табл. 2 показаны результаты оценки качества изображений после сжатия стандартным алгоритмом JPEG2000 в Adobe Photoshop CS6 Extended.

Сравнение полученных результатов по метрикам PSNR и MFSD показывает, что использование стандартного алгоритма JPEG2000 не обеспечивает высокого качества декодированных изображений при высоких коэффициентах сжатия. Использование обученных SR-моделей также не дает положительных результатов по сравнению с NeuroCorrector.



Рис. 6. Фрагмент изображения рис. 1*u*: а) оригинал; б) на входе модели; в) на выходе модели

3. Конвертер фотоизображений

В результате исследований разработан новый конвертер фотоизображений. Конвертер реализован с помощью двух модулей. Первый модуль реализован на языке C++ в приложении RAD Studio, в котором реализован интерфейс и операции кодирования и декодирования. Второй модуль – на языке Python, в котором реализована функция коррекции четкости изображения с помощью обученной нейросети NeuroCorrector.

Табл. 2. Результаты тестирования моделей

Тест Im	Декодер	NeuroCorrector		BSRGAN		ESRGAN		SwinIR		Photoshop			
Im	К _{сж}	PSNR	MFSD	PSNR	MFSD	PSNR	MFSD	PSNR	MFSD	PSNR	MFSD		
a	50	33,4	1,70	36,5	0,75	31,2	2,09	31,1	2,13	30,6	2,04	35,5	0,81
б	63	32,7	1,94	39,2	0,76	32,2	2,09	33,1	1,86	30,9	2,23	34,9	1,28
в	25	30,8	1,44	35,8	0,62	25,0	2,66	28,5	1,56	25,5	2,39	33,2	0,89
г	39	31,3	1,69	35,6	0,78	28,4	2,25	28,7	2,17	28,4	2,16	34,1	0,88
д	53	34,8	1,53	36,0	0,72	27,5	2,33	29,9	1,73	27,5	2,12	35,5	1,07
е	36	30,3	1,70	36,3	0,69	24,9	2,72	25,6	2,31	23,8	2,93	32,7	1,15
ж	44	29,6	2,04	36,0	0,57	25,0	2,77	26,8	2,29	23,9	3,12	33,9	0,94
з	45	32,8	1,37	37,1	0,66	29,9	1,72	30,5	1,53	28,6	1,84	33,9	1,00
и	41	27,6	2,76	35,2	0,77	24,5	3,02	24,6	2,96	23,4	3,17	31,8	1,34
м	44	31,5	1,79	36,4	0,70	27,6	2,41	28,8	2,06	27,0	2,44	33,9	1,04

В конвертере и интерфейсе реализованы следующие функции:

- 1) открытие файлов высококачественных изображений с любым разрешением в форматах PNG, BMP, TIFF или JPEG;
- 2) установка параметра качества Q , который по умолчанию равен 70 %;
- 3) кодирование, вычисление коэффициента сжатия и сохранение сжатых файлов изображения в выбранную папку;
- 4) открытие и декодирование сжатых файлов;
- 5) визуализация оригинального и декодированного изображений и оценка искажений по метрикам PSNR и MFSD;

б) подключение модуля NeuroCorrector и коррекция четкости с последующей визуализацией и сохранением файла изображения для дальнейшей обработки.

Конвертер может быть использован для архивации и хранения больших объемов видеоданных с невысоким качеством, но с высоким коэффициентом сжатия. Если требуется высокое качество, то пользователь может использовать функцию коррекции четкости и далее сохранить изображения для дальнейшей обработки или визуализации.

Отметим, что функции сжатия и декодирование изображения выполняются достаточно быстро и не превышают 0,05 с. Основное время тратится на

коррекцию четкости и зависит от производительности видеокарты, размера изображения и количества детальных блоков. В результате экспериментов на видеокарте RTX 3060 получено, что время коррекции изображения 2К не превышает 4 с и для UHD не превышает 2 с. Относительно низкое быстродействие корректора четкости обусловлено большим количеством параметров разработанной глубокой нейронной сети, а также не самой высокой производительностью используемой видеокарты.

К достоинствам конвертера относится высокий коэффициент сжатия, который в среднем для изображений с различной детальностью составляет 40...45 раз и высокое качество восстановленных изображений с помощью корректора четкости.

Заключение

В результате исследований предложен новый адаптивный алгоритм JPEG2000 с масштабированием размеров матриц коэффициентов ДВП сигналов Y, U и V на первой итерации ДВП. Определены взаимосвязи результатов оценки качества и минимизации искажений в процессе сжатия новым алгоритмом.

Предложена архитектура нейросети, позволяющая решать задачу коррекции искажений, возникающих в процессе сжатия и восстановления изображений. Выбран оптимальный вариант обучения нейронной сети, позволяющий использовать обученную модель для коррекции искажений восстановленных изображений до высоких объективных и субъективных оценок качества. В результате экспериментов доказана более высокая эффективность нейросетевого корректора по сравнению с известными аналогами.

Следует отметить, что модель нейронной сети обучена для коррекции искажений, возникающих именно в авторском алгоритме сжатия. Следовательно, предлагаемый корректор не будет эффективным при использовании других алгоритмов сжатия. Корректор четкости хорошо справляется с нелинейными искажениями в режиме «мягкого» квантования и с линейными искажениями, обусловленными интерполяцией. Если использовать нейросеть для коррекции четкости JPEG или JPEG2000 изображений, сжатых в 40 – 50 раз, то «грубое» квантование приводит к существенным нелинейным искажениями, которые снижают эффективность коррекции. Универсальных корректоров пока что не существует!

К перспективным направлениям применения нейросетевого корректора относится обработка кадрового видео. В частности, при аэрофотосъемке поверхности Земли [20] возникает проблема в хранении и передаче больших объемов видеоданных. Предлагаемый алгоритм адаптивного JPEG2000 позволяет эффективно сжимать кадры в реальном времени с последующей передачей по каналу связи или хранением на носителе информации. Нейросетевой корректор предназначен для повышения качества и детализации

объектов изображения земной поверхности с целью анализа и более эффективной обработки видеоданных в задачах распознавания и сегментации.

Благодарности

Исследование выполнено за счет гранта Российского научного фонда № 24-11-20024 и Министерства образования и науки Хабаровского края (Соглашение № 124С/2024).

References

- [1] Svoboda P, Hradis M, Barina D, Zemcik P. Compression artifacts removal using convolutional neural networks. arXiv Preprint. 2016. Source: <https://arxiv.org/abs/1605.00366>. DOI: 10.48550/arXiv.1605.00366.
- [2] Sai SV. Method of adaptive quantization for the coefficients of a discrete wavelet transformation. Optoelectron Instrum Data Process 2023; 59(2): 167-176. DOI: 10.3103/S8756699023020115.
- [3] Sai SV. Metric of fine structures distortions of compressed images. Computer Optics 2018; 42(5): 829-837. DOI: 10.18287/2412-6179-2018-42-5-829-837.
- [4] Wang Z, Chen J, Hoi SCH. Deep learning for image super-resolution: A survey. arXiv Preprint. 2020. Source: <https://arxiv.org/abs/1902.06068>. DOI: 10.48550/arXiv.1902.06068.
- [5] Top 15 best image enlarger review. 2024. Source: <https://topten.ai/image-enlargers-review/>.
- [6] Lai W-S, Huang J-B, Ahuja N, Yang M-H. Fast and accurate image super-resolution with deep Laplacian pyramid networks. IEEE Trans Pattern Anal Mach Intell 2019; 41(11): 2599-2613. DOI: 10.1109/TPAMI.2018.2865304.
- [7] DIV2K dataset: DIVERse 2K resolution high quality images as used for the challenges @ NTIRE (CVPR 2017 and CVPR 2018) and @ PIRM (ECCV 2018). 2025. Source: <https://data.vision.ee.ethz.ch/cvl/DIV2K/>.
- [8] Zhang K, Liang J, Van Gool L, Timofte R. Designing a practical degradation model for deep blind image super-resolution. arXiv Preprint. 2021. Source: <https://arxiv.org/abs/2103.14006>. DOI: 10.48550/arXiv.2103.14006.
- [9] Wang X, Xie L, Dong C, Shan Y. Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. arXiv Preprint. 2021. Source: <https://arxiv.org/abs/2107.10833>. DOI: 10.48550/arXiv.2107.10833.
- [10] Liang J, Cao J, Sun G, Zhang K, Van Gool L, Timofte R. SwinIR: Image restoration using swin transformer. arXiv Preprint. 2021. Source: <https://arxiv.org/abs/2108.10257>. DOI: 10.48550/arXiv.2108.10257.
- [11] Tai Y, Yang J, Liu X, Xu C. MemNet: A persistent memory network for image restoration. 2017 IEEE Int Conf on Computer Vision (ICCV) 2017: 4549-4557. DOI: 10.1109/ICCV.2017.486.
- [12] Zhang Y, Tian Y, Kong Y, Zhong B, Fu Y. Residual dense network for image super-resolution. arXiv Preprint. 2021. Source: <https://arxiv.org/abs/1802.08797>. DOI: 10.48550/arXiv.1802.08797.
- [13] Lim B, Son S, Kim H, Nah S, Lee K. Enhanced deep residual networks for single image super-resolution. arXiv Preprint. 2017. Source: <https://arxiv.org/abs/1707.02921>. DOI: 10.48550/arXiv.1707.02921.

- [14] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. arXiv Preprint. 2015. Source: <<https://arxiv.org/abs/1512.03385>>. DOI: 10.48550/arXiv.1512.03385.
- [15] Li J, Fang F, Mei K, Zhang G. Multi-scale residual network for image super-resolution. In Book: Ferrari V, Hebert M, Sminchisescu C, Weiss Y, eds. Computer vision – ECCV 2018. 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part VIII. Cham, Switzerland: Springer Nature Switzerland AG; 2018: 527-542. DOI: 10.1007/978-3-030-01237-3_32.
- [16] Fairchild MD. Color appearance models. John Wiley and Sons Ltd; 2005. ISBN: 0-470-01216-1.
- [17] Keras 3 API documentation. 2025. Source: <<http://keras.io/>>.
- [18] CUDA Toolkit. 2025. Source: <<http://developer.nvidia.com>>.
- [19] TensorFlow. 2025. Source: <<http://tensorflow.org>>.
- [20] Bychkov IV, Ruzhnikov GM, Fedorov RK, Popova AK, Avramenko YV. Classification of Sentinel-2 satellite images of the Baikal Natural Territory. Computer Optics 2022; 46(1): 90-96. DOI: 10.18287/2412-6179-CO-1022.

Сведения об авторах

Сай Сергей Владимирович, 1960 года рождения, в 1983 году окончил Томский институт автоматизированных систем управления и радиоэлектроники (ТИАСУР) по специальности «Радиоэлектронные устройства». Доктор технических наук, профессор Тихоокеанского государственного университета (ТОГУ). Область научных интересов: объективные методы анализа изображений, распознавание образов. E-mail: 000493@pnu.edu.ru

Никонов Василий Сергеевич, 1998 года рождения. Аспирант Тихоокеанского государственного университета. В 2019 году получил степень бакалавра по направлению «Информатика и вычислительная техника». В 2021 получил степень магистра по направлению «Экология и природопользование». С 2021 проходит подготовку в аспирантуре Тихоокеанского государственного университета. Сфера научных интересов: цифровая обработка, анализ и распознавание изображений. E-mail: 2021104211@pnu.edu.ru

Фомина Екатерина Сергеевна, 1993 года рождения, в 2015 году окончила Тихоокеанский государственный университет (ТОГУ) по специальности «Многоканальные телекоммуникационные системы». Старший преподаватель Тихоокеанского государственного университета (ТОГУ). Область научных интересов: цифровая обработка сигналов, кодирование и передача информации. E-mail: 009373@pnu.edu.ru

ГРНТИ: 28.23.15

Поступила в редакцию 11 июля 2024 г. Окончательный вариант – 11 декабря 2024 г.

Adaptive compression algorithm based on JPEG2000 with neural network corrector of decoded image definition

S.V. Sai ¹, V.S. Nikonov ¹, E.S. Fomina ¹

¹Pacific National University, Tikhookeanskaya Str. 136, Khabarovsk, 680035, Russia

Abstract

The article describes features of digital image processing in the process of adaptive compression based on the discrete wavelet transform (DWT) used in the JPEG2000 compression standard. Unlike the well-known compression algorithm, the new algorithm uses a reduction (scaling) of the sizes of the matrices of the DWT coefficients of the Y, U, and V signals at the first DWT iteration, after which a reduced copy of the low-frequency range image is used for further processing. Thus, at this stage the volume of video data is reduced by four times. To obtain the original resolution, in contrast to the well-known interpolation methods, the inverse DWT transform with added zero coefficients in the high-frequency subranges of the first iteration is used. The proposed algorithm allows compressing image files on average by 40...50 times with satisfactory quality. To restore high image quality, an original neural network sharpness corrector based on a convolutional model with training of sample image blocks by the brightness index in the Lab space is developed. Based on the ResNet architecture, a proprietary deep neural network model is developed, based on combining several techniques used to solve image reconstruction problems in other architectures. An optimal neural network training option is selected, allowing the trained model to be used to correct the clarity of decoded images to high objective quality indicators with subjective assessments of "good" and "excellent".

Keywords: definition corrector, image analysis, distortion metric, discrete wavelet transform, compression efficiency, neural network.

Citation: Sai SV, Nikonov VS, Fomina ES. Adaptive compression algorithm based on JPEG2000 with neural network corrector of decoded image definition. *Computer Optics* 2025; 49(6): 986-993. DOI: 10.18287/2412-6179-CO-1586.

Acknowledgements: This work was funded by the Russian Science Foundation under grant No 24-11-20024 and the Ministry of Education and Science of Khabarovsk Krai under agreement No 124C/2024.

Author's information

Sergey Vladimirovich Sai, born in 1960, in 1983 graduated from the Tomsk Institute of Automated Control Systems and Radioelectronics (TIASUR), specializing in Radioelectronics Devices. Doctor of Technical Sciences, professor at the Pacific National University (PNU). Area of scientific interests: objective methods of image analysis, pattern recognition. E-mail: 000493@pnu.edu.ru

Vasily Sergeevich Nikonov, born in 1998. Postgraduate student of Pacific State University. In 2019, he received a bachelor's degree in "Informatics and Computer Science". In 2021 received a master's degree in Ecology and Nature Management. Since 2021 he has been studying at the graduate school of the Pacific National University. Research interests: digital image processing, analysis and recognition. E-mail: 2021104211@pnu.edu.ru

Ekaterina Sergeevna Fomina, born in 1993, in 2015 graduated from the Pacific National University (PNU), specializing in Multichannel telecommunication systems. Lecturer at the Pacific National University (PNU). Research interests: digital signal processing, coding and transmission of information. E-mail: 009373@pnu.edu.ru

Received July 11, 2024. The final version – December 11, 2024.
