

Министерство образования Республики Беларусь  
Учреждение образования  
Белорусский государственный университет  
информатики и радиоэлектроники

УДК 004.627

Дворко  
Андрей Вадимович

Квантователь в системе трансформационного кодирования изображений на  
основе искусственных нейронных сетей

**АВТОРЕФЕРАТ**

на соискание степени магистра технических наук  
по специальности 1-40 80 01 «Элементы и устройства вычислительной  
техники и систем управления»

Научный руководитель  
Петровский Николай Александрович  
Доцент, кандидат технических наук

Минск 2019

## ВВЕДЕНИЕ

Дискретное вейвлет-преобразование и векторное квантование являются двумя мощными и эффективными инструментами, применяемыми для сжатия изображений. В последние 10-15 лет было написано много работ и статей о них, были приложены огромные усилия объединить их для получения еще более эффективных методов. На сегодняшний день предложено немало алгоритмов и методов сжатия изображений в области ДВП, действует основанный на ДВП международный стандарт JPEG2000.

Увеличение степени сжатия изображений методами на основе ДВП приводит к росту значений ошибок восстановления вейвлет-коэффициентов и визуальному ухудшению восстановленных изображений. При использовании традиционных алгоритмов мелкие детали (границы) частично теряются при высоких степенях сжатия – наблюдаются размытости целых областей в изображении. Сохранение этих деталей является важным при сжатии фотографических изображений.

В связи с этим актуальной является задача поиска эффективных алгоритмов обработки значений вейвлет-коэффициентов, позволяющих сократить потери информации при квантовании и восстановлении значений вейвлет-коэффициентов. Кроме того, есть необходимость исследования и совершенствования существующих методов и алгоритмов сжатия на основе ВК в области ДВП для решения описанных выше проблем, а также улучшения их характеристик и расширения области применения.

В общем виде решение задачи сжатия цветных фотографических изображений можно свести к решению задачи эффективного сжатия полутоновых изображений – изображений с глубиной цвета 8 бит на пиксел. Это следует из того, что изображение с цветовой схемой RGB первоначально представляется в виде комбинации яркостной (Y) и двух цветных составляющих (U и V).

В данной работе будет проведено исследование применения многослойных искусственных нейронных сетей и преобразований на их основе к задаче сжатия цифрового изображения.

В информационных технологиях искусственная нейронная сеть представляет собой систему аппаратного и/или программного обеспечения. Применение нейронных сетей обычно сосредоточено на решении сложных задач обработки сигналов или распознавания образов.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

### **Актуальность темы исследования**

В настоящее время цифровая обработка и последующее распознавание изображений является одним из интенсивно развивающихся направлений научных исследований. Многие отрасли техники, имеющие отношение к получению, обработке, хранению и передаче информации, в значительной степени ориентируются в настоящее время на развитие систем, в которых информация имеет характер изображений.

В системах цифровой обработки изображений стремятся уменьшить число уровней и порогов квантования, так как от их количества зависит объем информации, необходимый для кодирования изображения. Однако при относительно небольшом числе уровней на квантованном изображении возможно появление ложных контуров. Они возникают вследствие скачкообразного изменения яркости проквантованного изображения и особенно заметны на пологих участках ее изменения. Ложные контуры значительно ухудшают визуальное качество изображения, так как зрение человека особенно чувствительно именно к контурам. Рассмотрение вопроса о моделировании системы квантования цифрового изображения с использованием вейвлет-преобразований на основе искусственных нейронных сетей является актуальным.

### **Цель и задачи исследования**

Целью данного исследования является исследование процесса квантования изображения при сжатии и создании модели нейронной сети, для последующего квантования на основе вейвлет преобразования.

В соответствии с поставленной целью, в работе сформулированы и решены следующие задачи:

1. Провести анализ современных алгоритмов кодирования изображения.
2. Провести анализ применения вейвлет-преобразований в современных алгоритмах кодирования изображения.
3. Выбрать архитектуру нейронной сети для решения поставленной задачи и разработать алгоритм ее обучения.
4. Экспериментально оценить разработанную схему кодирования изображений.
5. Реализовать программное средство.

**Объектом** исследования является цифровое изображение.

**Предметом** исследования выступают квантователь цифровых изображений и алгоритмы кодирования изображений.

**Область исследования** и содержание диссертационной работы соответствует образовательному стандарту высшего образования второй ступе-

ни (магистратуры) специальности 1-40 80 01 «Элементы и устройства вычислительной техники и систем управления».

**Научная новизна** диссертационной работы заключается в разработке и верификации алгоритма компрессора изображений, использующего все преимущества современных схем кодирования изображений, а именно нелинейного преобразования анализа, равномерного квантователя и нелинейного преобразования синтеза на основе аппарата нейронных сетей для получения максимально быстрого и качественного кодирования изображения.

#### **Положения, выносимые на защиту**

1. Структура нейросетевого компрессора изображений и сравнение структуры с современными алгоритмами кодирования изображений.
2. Результаты анализа применения вейвлет-преобразований в современных алгоритмах кодирования.
3. Результаты экспериментальной оценки разработанной схемы кодирования.

#### **Апробация результатов диссертации**

Основные положения и результаты диссертационной работы нигде не докладывались и не обсуждались на конференциях.

#### **Опубликованность результатов исследования**

По результатам исследований, представленных в диссертации, не было опубликовано печатных работ.

#### **Структура и объем диссертации**

Структура диссертационной работы обусловлена целью, задачами и логикой исследования. Работа состоит из введения, четырех глав и заключения, библиографического списка и приложений. Общий объем диссертации – 69 страниц. Работа содержит 3 таблицы, 19 рисунков. Библиографический список включает 35 наименований, графический материал включает 13 слайдов презентации.

## СОДЕРЖАНИЕ РАБОТЫ

Во **введении** рассмотрено современное состояние проблемы кодирования изображений, определены основные направления исследований, а также дается обоснование актуальности темы диссертационной работы.

В **общей характеристике работы** сформулированы ее цель и задачи, показана связь с научными программами и проектами, даны сведения об объекте исследования и обоснован его выбор, представлены положения, выносимые на защиту, приведены сведения о личном вкладе соискателя, апробации результатов диссертации и их опубликованность, а также, структура и объем диссертации.

В **первой главе** рассматриваются требуемые параметры современной схемы кодирования изображений, приводится описание существующих алгоритмов кодирования изображений и кодеров на их основе, рассматриваются алгоритмы и подходы к квантованию и кодированию параметров.

Основная цель кодирования изображений состоит в том, чтобы максимизировать воспринимаемое качество при определенной скорости передачи битов или минимизировать скорость передачи битов для определенного качества восприятия.

К требуемым характеристикам современного алгоритма кодирования изображения относят:

1. высокая степень компрессии;
2. высокое качество сжатого изображения (данное требование противоречит выполнению предыдущего требования, поэтому всегда приходится искать компромисс между степенью сжатия и качеством восстановленного изображения);
3. высокая скорость компрессии (данное требование актуально для приложений, занимающихся кодированием изображений в реальном масштабе времени: цифровых фотоаппаратов, видеокамер);
4. высокая скорость декомпрессии (данное требование актуально почти для всех приложений).

Применение нейронных сетей и преобразований на их основе для обработки изображений в настоящее время, как правило, сосредоточено на задачах распознавания или улучшения сигнала (удаление шума, изменение отдельных характеристик сигналов и т. п.). Применительно к задаче кодирования изображений существуют примеры применения нейронных сетей с линейным преобразованием данных.

Алгоритм JPEG основан на ДКП, применяемом к матрице непересекаю-

щихся блоков изображения, размером 8x8 пикселей. ДКП раскладывает эти блоки по амплитудам некоторых частот. В результате получается матрица, в которой многие коэффициенты, как правило, близки к нулю, которые можно представить в грубой числовой форме, т.е. в квантованном виде без существенной потери в качестве восстановления. Дискретное косинусное преобразование представляет собой преобразование информации без потерь и не осуществляет никакого сжатия. Дискретное косинусное преобразование подготавливает информацию для этапа сжатия с потерями или округления. После чего проходит этап квантования. Стандарт JPEG реализует эту процедуру через матрицу округления. Для каждого элемента матрицы ДКП существует соответствующий элемент матрицы округления. Результирующая матрица получается делением каждого элемента матрицы дискретного косинусного преобразования на соответствующий элемент матрицы округления и последующим округлением результата до ближайшего целого числа. Как правило, значения элементов матрицы округления растут по направлению слева направо и сверху вниз.

В JPEG2000 используется однобитное изображение-маска, задающее повышение качества в данной области изображения. Поскольку за качество областей у нас отвечают коэффициенты ДВП преобразования во 2, 3 и 4 квадрантах, то маска преобразуется т. о., чтобы указывать на все коэффициенты, соответствующие областям повышения качества. Эти области обрабатываются далее другими алгоритмами (с меньшими потерями), что и позволяет достичь более высоких показателей сжатия. В стандарте применяется однородное квантование (скалярное квантование) с центральной «мертвой зоной». Квантователь можно представить в виде функции, которая ставит в соответствие анализируемому коэффициенту некоторое множество на прямой действительных чисел. Для вейвлет-коэффициента квантователь выдает индекс (целое число со знаком).

Во **второй главе** приведена структура параметрического кодера изображений, и структура нейросети.

Непрерывно значимые данные (такие как векторы интенсивности пикселей изображения) должны быть квантованы до конечного набора дискретных значений, что приводит к ошибке. В этом контексте необходимо компенсировать две конкурирующие затраты: энтропию дискретизированного представления (степень сжатия) и ошибку, возникающую из-за квантования (искажения). Различные применения сжатия, такие как хранение или передача данных по каналам с ограниченной пропускной способностью, требуют различных компромиссов между степенью сжатия и искажением.

Большинство методов сжатия основаны на ортогональных линейных преобразованиях, выбранных для уменьшения корреляций в данных и, таким обра-

зом, для упрощения энтропийного кодирования. Но статистика откликов линейного фильтра демонстрирует сильные зависимости более высокого порядка. Они могут быть значительно снижены за счет использования совместных локальных нелинейных операций управления усилением, вдохновленных моделями зрительных нейронов, каскадные версии таких моделей использовались для захвата нескольких этапов визуальной трансформации. Некоторые более ранние результаты показывают, что включение локальной нормализации в методы кодирования с линейным блочным преобразованием может улучшить производительность кодирования и может улучшить производительность распознавания объектов в каскадных сверточных нейронных сетях. Однако параметры нормализации в этих случаях не были оптимизированы для задачи. Здесь используется преобразование GDN с оптимизированными параметрами, которое является высокоэффективным в Гауссификации локальной объединенной статистики естественных изображений гораздо больше, чем наборы линейных преобразований, за которыми следуют поточечные нелинейности.

Модель содержит несколько стековых модулей, каждый из которых является оптимизированным автоэнкодером. Исходное изображение сначала сжимается в базовом кодере, далее усиливающие слои будут принимать разницу между последним восстановленным изображением и исходным изображением в качестве входных данных, а затем сжимать остаток. Обучение структуры происходит послойно, что означает, что идет исправление всех предыдущих слоев при обучении текущего. В предложенной структуре усиливающие слои отвечают за итеративное кодирование остатков между восстановленным изображением из предыдущих слоев и исходным изображением. Первый усиливающий слой принимает ошибку между исходным изображением  $x$  и восстановленным изображением базового слоя в качестве входных данных. Как таковое, качество восстановления может быть улучшено с помощью ошибки, закодированной в улучшающем слое. Для последующих усиливающих слоев, взяв в качестве примера второй улучшающий слой, входные данные являются остатком между исходным  $x$  и реконструкцией из последнего слоя.

**В третьей главе** описано моделирование программного средства. Также сформулированы требования к методам и слоям нейросети.

**В четвертой главе** представлены результаты экспериментальных исследований.

На основании анализа зависимости ошибки от количества шагов обучения, было установлено, что предварительное обучение доказало свою эффективность, и обеспечивает хорошее приближение к решению задачи оптимизации и снижение ошибки.

Анализ ошибки реконструкции для тестовых образцов показал, что пред-

ложенный подход к представлению данных позволяет снизить ошибку реконструкции, но не всегда.

Сравнительный анализ ошибки реконструкции для нейросетевого квантователя показывает, что нейросетевой квантователь превосходит аналоги в частных случаях тестовых образцов при меньшем количестве вычислительных затрат и это показывает, что предложенная схема кодирования близка к аналогам по качеству реконструкции.

В приложениях представлен графический материал и исходное описание разработанной системы на языке Python.



## ЗАКЛЮЧЕНИЕ

В настоящей работе было проведено исследование применения многослойных искусственных нейронных сетей и преобразований на их основе применительно к задаче задачи квантования и кодирования изображений.

Также был представлен полный метод сжатия изображений, основанный на кодировании с нелинейным преобразованием. Данный метод сжатия предлагает улучшения в производительности степень сжатия-искажение по сравнению с JPEG и JPEG 2000 для большинства изображений и битрейтов. Что еще более примечательно, хотя метод был оптимизирован с использованием среднеквадратичной ошибки в качестве метрики искажения, сжатые изображения выглядят намного более естественно, чем сжатые с помощью JPEG или JPEG 2000, которые страдают от серьезных артефактов, обычно наблюдаемых у методов кодирования с линейным преобразованием. В соответствии с этим, качество восприятия (согласно оценке MS-SSIM) значительно улучшается во всех тестовых изображениях и битрейтах. Предполагается, что это визуальное улучшение возникает потому, что набор нелинейных преобразований в модели был оптимизирован для сбора характеристик и атрибутов изображений, которые представлены в статистике данных, параллельно процессам эволюции и развития, которые считаются формировать визуальные представления в человеческом мозге. Тем не менее, дополнительные визуальные улучшения могут быть возможны, если метод оптимизирован с использованием перцептуальной метрики вместо MSE.

Поскольку преобразование в дискретное представление можно рассматривать как форму классификации, стоит подумать о том, что может быть перенесено в более конкретные контролируемые проблемы обучения, такие как распознавание объектов. Например, аддитивный шум, используемый в целевой функции в качестве релаксации квантования, может также служить для повышения устойчивости контролируемых классификационных сетей.