

# ПРИМЕНЕНИЕ ГЛУБОКОГО АВТОЭНКODERA ДЛЯ СХЕМЫ СЖАТИЯ ИЗОБРАЖЕНИЯ НА ОСНОВЕ ПАРАУНИТАРНОГО БАНКА ФИЛЬТРОВ

Леченко А. В., Петровский Н. А.

Кафедра информатики,

Кафедра электронных вычислительных средств,

Белорусский государственный университет информатики и радиоэлектроники

Минск, Республика Беларусь

E-mail: {lechenko, nick.petrovsky}@bsuir.by

*В работе представлен комбинированный схема сжатия изображений на основе параунитарного банка фильтров в алгебре кватернионов ( $Q$ -ПУБФ) и глубокого автоэнкодера. Рассматривается сжатие изображений оттенков серого с разрешением  $1280 \times 768$ . Предложенный метод состоит из пространственной фильтрации изображения с помощью  $Q$ -ПУБФ с последующим сжатием низкочастотного канала автоэнкодером и квантованием остальных каналов. Результаты показывают, что восстановленные при помощи автоэнкодера изображения лучше по объективным характеристикам ( $PSNR$  и  $SIMM$ ) в сравнении с равномерным квантованием коэффициентного банка фильтров для случая низкой битовой скорости.*

## ВВЕДЕНИЕ

Алгоритмы сжатия изображений играют важную роль в передаче и хранении мультимедиа данных в условиях ограниченной ширины канала или лимитированном объеме хранилища данных. Современные стандарты сжатия изображений, как правило используют банки фильтров для разделения на субполосы, что позволяет более эффективно декоррелировать входные данные [1]. В настоящее время идет поиск методов сжатия изображений с помощью нейронных сетей [2], в том числе глубоких автоэнкодеров [3-4]. Применение нейронных сетей для сжатия изображений обсуждалось и ранее [5], однако комбинирование нейросетевых методов с классическими пространственно-частотными преобразованиями представляют интерес из-за высокой производительности.

В данной работе предлагается схема сжатия изображений, в которой изображение сначала разбивается на частотно-пространственные субполосы с помощью параунитарного банка фильтров ( $Q$ -ПУБФ) на основе алгебры кватернионов [6], после чего низкочастотный канал сжимается с помощью глубокого автоэнкодера [3-4], а остальные каналы квантуются с помощью равномерного квантователя в соответствии с их энергией и выделенным бюджетом бит [7].

## I. СХЕМА СЖАТИЯ

Пусть на вход кодера подается изображение размером  $M \times N$ . С помощью 1D 4-х канального  $Q$ -ПУБФ реализован 2D 16-канальный банк фильтров для декорреляции изображения. Банк фильтров концентрирует перцептуально значимую информацию в низкочастотном (НЧ)-канале для большинства изображений. Далее НЧ-канал делится на 4 равных части, каждая из которых кодируется соответствующим пред-

бученным автоэнкодером. Выходом автоэнкодера является 32-х слойный тензор, который стохастически бинаризуется для получения дискретных значений. Кодовой слой автоэнкодера кодирует содержимое НЧ-канала со скоростью 2 бита на значение исходного канала. Остальные высокочастотные каналы дискретизируются в соответствии с алгоритмом [7], учитывая, что 2 бита на значение уже распределены для НЧ части.

После того как изображение было квантовано подобным образом, НЧ-канал можно декодировать с помощью декодера глубокого автоэнкодера, остальные каналы можно домножить на соответствующий шаг квантования и подать получившиеся изображения на вход банка фильтров синтеза  $Q$ -ПУБФ, на выходе которого будет реконструированное изображение.

## II. СТРУКТУРА АВТОЭНКODERA

Рассмотрим подробнее архитектуру глубокого автоэнкодера. Выбранный автоэнкодер основан на работах [3-4] и состоит из блока энкодера с модулем стохастической бинаризации и блока декодера. Подробная схема автоэнкодера приведена на рисунке 1.

Первые два слоя автоэнкодера увеличивают количество слоев до 128 и ширину и высоту тензора в 4 раза. Далее следует линейный сверточный блок из двух слоев  $\text{LeakyReLU}$  между ними. К выходному тензору этого блока прибавляется значение на входе. Такой сверточный блок применяется три раза к остаточным значениям. Завершается энкодер еще одним сверточным слоем, который уменьшает количество каналов до 32. Полученный выходной тензор передается на вход функции  $\tanh(\cdot)$ , которая отображает значения тензора на интервал от  $[-1, 1]$ , пригодный для стохастической бинаризации. Стохастическая бинаризация вычисляется по формуле:  $b = \text{sign}(x + \psi)$ , где  $\psi$  – слу-

чайная величина с равномерным распределением в диапазоне  $[-1, 1)$ . Бинаризованный тензор и есть кодовый слой. Таким образом, часть изображения  $m \times n$  преобразуется в тензор размеров  $32 \times (n/4) \times (m/4)$  бит, или 2 бита на значение исходного части.

Декодер в основном повторяет структуру энкодера в обратном порядке: два сверточных слоя, которые увеличивают тензор до 128 каналов, три линейных блока свертки, и далее слой увеличивающие ширину и высоту до исходных.

### III. ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ

Для тестирования предложенной схемы, из Youtube-8m [8] были случайно выбраны 1700 изображений оттенков серого в разрешении  $1280 \times 768$ . Изображения были разделены на субполосы каскадом анализа  $Q$ -ПУБФ. Затем часть изображений делилась на обучающую и тестовую выборки. НЧ-каналы обучающей выборки использовались для обучения глубокого автоэнкодера. После того как автоэнкодер был обучен, генерировались восстановленные НЧ-каналы для изображений из тестовой выборки.

Для сравнения метода, в качестве опорного использовался метод [7], поэтому остальные каналы кодировались в соответствии с выделенным бюджетом бит с учетом того, что НЧ-канал потреблял только два бита на значение, т. е., нераспределенные биты на НЧ канал могли распределяться между остальными каналами. Далее полученные квантованные каналы и НЧ-канал, полученный при помощи автоэнкодера, использовались для восстановления изображения каскадом синтеза  $Q$ -ПУБФ, для полученного соотношения считались пиковое отношение сигнал-шум (PSNR) и индекс структурного сходства (SSIM). Сравнение средних значений этих метрик на всех изображениях с аналогичными для метода [7] приведены в таблице 1.

Как видно из таблицы, с ростом количества бит на пиксель (BPP), выделяемых на квантование методом [7], предложенное решение начинает проигрывать. В случае 0.375 BPP, метод [7] выделяет 6 бит на значение для НЧ-канала, что в три раза больше, чем доступно автоэнкодеру.

Перераспределение 4 бит на другие следующие по энергии каналы не позволяет получить заданную точность. Однако предложенная схема дает лучший результат для высоких степеней сжатия. Метрики качества предложенного решения почти в полтора раза выше, когда оба метода используют 0.125 BPP для кодирования изображения и все еще больше, даже если увеличить бюджет бит для метода [7] в полтора раза.

Таблица 1 – Сравнение эффективности сжатия предложенной схемой и методом [7]

BPP	SSIM	SSIM [7]	PSNR	PSNR [7]
0.125	0.834	0.444	32.1	22.7
0.1875		0.778		30.5
0.375	0.839	0.938	32.3	37.1
0.8	0.866	0.962	33.2	40.3
1.6	0.877	0.980	33.6	44.3

Из анализа полученных результатов следует вывод, что применение автоэнкодера показывает улучшение объективных характеристик для восстановленных изображений при низких битрейтах в представленном эксперименте.

1. Saramaki T., Bregovic R. Multirate systems and filterbanks //Multirate systems: design and applications. – IGI Global, 2002. – С. 27-85.
2. Ma S. et al. Image and video compression with neural networks: A review //IEEE Transactions on Circuits and Systems for Video Technology. – 2019. – Т. 30. – №. 6. – С. 1683-1698.
3. Theis L. et al. Lossy image compression with compressive autoencoders //arXiv preprint arXiv:1703.00395. – 2017.
4. Toderici G. et al. Variable rate image compression with recurrent neural networks //arXiv preprint arXiv:1511.06085. – 2015.
5. Dony R. D., Haykin S. Neural network approaches to image compression //Proceedings of the IEEE. – 1995. – Т. 83. – №. 2. – С. 288-303.
6. Парфенюк М., Петровский А. А. Параунитарные банки фильтров на основе алгебры кватернионов: теория и применение //Цифровая обработка сигналов. – 2008. – №. 1. – С. 22-36.
7. Петровский Н. А. Оптимальное распределение бит в параунитарном субполосном кодере изображений на основе алгебры кватернионов //Доклады БГУИР. – 2014. – №. 1. – С. 72-77.
8. Abu-El-Haija S. et al. Youtube-8m: A large-scale video classification benchmark //arXiv preprint arXiv:1609.08675 – 2016.

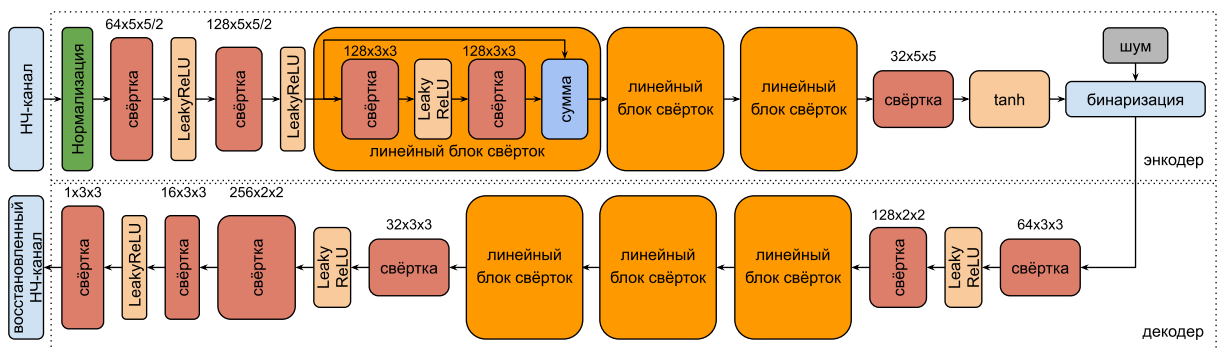


Рис. 1 – Схема глубокого автоэнкодера. LeakyReLU взят с коэффициентом утечки 0.01.