

# Энтропийное моделирование при сжатии видео на основе машинного обучения

М.А. Якубенко  
Самарский национальный исследовательский университет им. академика С.П. Королева  
Самара, Россия  
yamaksimkubenko@gmail.com

М.В. Гашинов  
Самарский национальный исследовательский университет им. академика С.П. Королева  
Самара, Россия  
mih-fast@yandex.ru

**Аннотация**— В работе исследуется энтропийная модель пространственно-временных характеристик при сжатии видео с использованием алгоритмов машинного обучения. Модель позволяет эффективно оценивать, как пространственные, так и временные характеристики сжимаемых видеоданных. Кроме того, универсальность энтропийной модели также позволяет устанавливать шаг квантования по пространственному каналу. Этот механизм квантования, адаптированный к контенту, не только помогает добиться плавной настройки скорости сжатия, но и улучшает конечную производительность за счет динамического распределения интервалов квантования. Результаты вычислительных экспериментов на реальных видеопоследовательностях подтверждают эффективность исследуемого метода сжатия видео.

**Ключевые слова**— машинное обучение, нейронные сети, сжатие видео, энтропийная модель, квантование

## 1. ВВЕДЕНИЕ

Стремительный рост темпов цифровизации и объёма цифровых данных в настоящее время влечёт повышение затрат на хранение и передачу данных, в частности видеоданных. Актуализация проблемы подталкивает к исследованиям и разработке эффективных средств снижения издержек.

Приложение машинного обучения к задаче компрессии, как подающее надежды возможное решение, вызывает интерес многих исследователей [1]. В данной работе исследуется энтропийная модель пространственно-временных характеристик и метод адаптивного квантования при сжатии видео с использованием алгоритмов машинного обучения.

## 2. АРХИТЕКТУРА МЕТОДА ЭНТРОПИЙНОГО МОДЕЛИРОВАНИЯ ПРИ СЖАТИИ ВИДЕО НА ОСНОВЕ МАШИННОГО ОБУЧЕНИЯ

Архитектура рассматриваемого метода компрессии видео, основанного на машинном обучении, представлена на «Рис. 1» Общий принцип работы этого метода компрессии можно описать следующим образом.

Входной кадр, совместно с реконструированным предыдущим кадром из буфера реконструированных кадров и признаков проходит через модуль оценки вектора движения. После кодирования и декодирования вектора движения реконструированный вектор движения совместно с реконструированным признаком предыдущего кадра из буфера декодированных кадров и признаков проходит через модуль анализа временного контекста. В модуле анализа формируются контексты разных масштабов, участвующие в работе контекстуального кодера-декодера и энтропийной модели.

Контекстуальный кодер преобразует входной кадр в скрытое представление, участвующее совместно с контекстами и реконструированным скрытым представлением предыдущего кадра в процессе формирования пространственно-канального шага квантования и оценке параметров собственного распределения. Арифметический кодер преобразует квантованное скрытое представление в битовый код.

Реконструкция кадра происходит по признаку наибольшего масштаба в генераторе кадров. В процессе реконструкции также заполняются буфер реконструированных кадров и признаков, а также буфер реконструированных скрытых представлений.

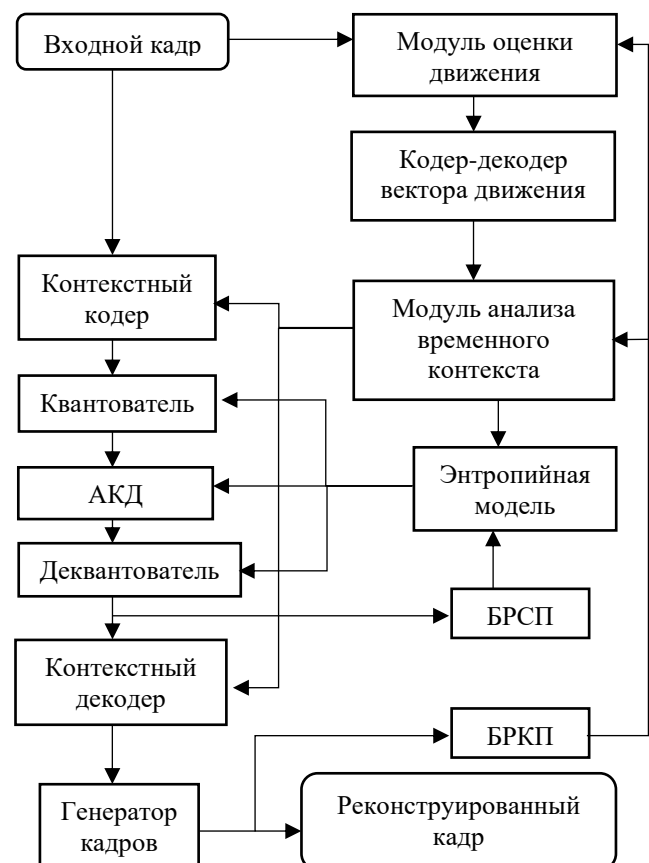


Рис. 1. Архитектура метода компрессии. АКД - арифметический кодер- декодер. БРСР - буфер реконструированных скрытых представлений. БРКР - буфер реконструированных кадров и признаков

### А. Оценка движения

Для оценки движения, результатом которой является вектор движения используется сетевая архитектура SPyNet [2]. Метод оценки оптического потока с использованием сетей пространственных пирамид

сочетает преимущества машинного обучения и классического подхода к оценке движения.

Пространственная пирамида позволяет эффективно оценивать, как большие, так и малые движения, благодаря располагающимся на разных уровнях пирамиды нейронным сетям.

#### Б. Кодер-Декодер вектора движения

Модули кодера и декодера спроектированы на основе структуры гиперприора с внесением асимметричности в декодере [3] для снижения вычислительных затрат на декодирование, не приводящих к значительным потерям. На входе модели имеем вектор движения, на выходе восстановленный вектор движения.

#### В. Анализ временного контекста

Используемый метод анализа временного контекста является преемником методов сжатия видео в области признаков. Анализ временного контекста необходим для оценки временной корреляции.

#### Г. Контекстуальный кодер-декодер

Подобно кодеру-декодеру вектора движения, контекстуальный кодер-декодер спроектирован на основе структуры гиперприора с некоторой асимметричностью. С помощью временных контекстов разного масштаба кодер преобразует текущий кадр в скрытое представление, а декодер реконструированное скрытое представление преобразует в реконструированный признак текущего кадра.

#### Д. Энтропийная модель

Энтропийная модель предназначена для оценивания параметра распределения вероятностей квантованного скрытого представления текущего кадра. Необходимость как можно более точной оценки исходит из условия минимизации перекрестной энтропии между истиной и оценочной распределениями вероятностей квантованного скрытого представления, показывающей среднее количество битов, необходимых для арифметического кодирования квантованного скрытого представления.

Путь улучшения оценки проходит через оценку корреляции между элементом скрытого представления и предыдущим реконструированным скрытым представлением, временным контекстным признаком наименьшего масштаба и гиперприором. Так же, для учёта пространственной корреляции и корреляции между каналами, используется двойной пространственный приор.

#### Е. Квантователь

Ещё одной особенностью данного подхода является возможность регулировки скорости без переобучения модели. Адаптивное квантование подразумевает использование трёх шагов квантования: глобального, канального и пространственно-канального, генерирующегося энтропийной моделью.

### 3. ЭКСПЕРИМЕНТ

Проведены вычислительные эксперименты на видеопоследовательностях из набора данных [4]. Оценивалась зависимость пикового отношения сигнал-шум (PSNR) от объема сжатых данных в битах на отсчёт (bits per pixel, BPP)

Усреднённые результаты показаны на «Рис. 2».

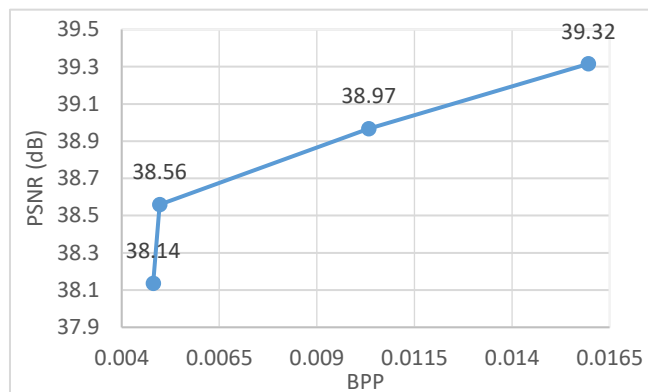


Рис. 2. Зависимость качества декомпрессированного видео (PSNR) от объёма сжатых данных в бит на отсчёт (BPP)

#### ЗАКЛЮЧЕНИЕ

Исследовался основанный на машинном обучении метод сжатия видео, рассматривающий универсальную энтропийную модель пространственно-временных характеристик и адаптивный квантователь. Ценность рассмотренного метода, в контексте развития приложений машинного обучения к задаче компрессии, заключается во введении универсальности, необходимой для приобретения методом прикладного значения. Проведенные вычислительные эксперименты подтвердили эффективность исследуемого метода при сжатии видео.

#### БЛАГОДАРНОСТИ

Исследование выполнено за счет гранта Российского научного фонда (проект № 22-21-00662).

#### ЛИТЕРАТУРА

- [1] Jiahao, Li. Hybrid Spatial-Temporal Entropy Modelling for Neural Video Compression / Jiahao Li, Bin Li, Yan Lu // Proceedings of the 30th ACM International Conference on Multimedia . – 2022. – P. 1503–1511. DOI 10.1145/3503161.3547845.
- [2] Ranjan, A. Optical Flow Estimation Using a Spatial Pyramid Network / A. Ranjan, M. J. Black // Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2017. – P. 2720-2729. DOI 10.1109/CVPR.2017.291
- [3] Temporal Context Mining for Learned Video Compression [Electronic resource]. — Access mode: <https://arxiv.org/abs/2111.13850> (07.11.2022).
- [4] Wang, H. MCL-JCV: A JND-based H.264/AVC video quality assessment dataset / H. Wang // Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP). – 2016. – P. 1509-1513. DOI 10.1109/ICIP.2016.7532610.