

# Разработка кодера на основе аппарата искусственных нейронных сетей для сжатия видеопотока эндоскопических изображений

А. П. Лосев

*Санкт-Петербургский государственный университет  
телекоммуникаций им. проф. М.А. Бонч-Бруевича*

losev.ap@sut.ru

А. Н. Бучатский

*Санкт-Петербургский государственный университет  
телекоммуникаций им. проф. М.А. Бонч-Бруевича*

abuchatsky@sut.ru

**Аннотация.** Современные видеокодеки, используемые на множестве устройств, постепенно начинают использовать аппарат искусственных нейронных сетей для сжатия видео. На данный момент, алгоритмы искусственных нейронных сетей активно применяются в задачах обнаружения объектов, отслеживания объектов, сегментирования изображений и множестве других прикладных задач, однако внедрение искусственных нейронных сетей для сжатия видео проводится лишь в малой доле новейших разработок в области видеокодеков. Архитектуры нейронных сетей позволяют преобразовывать информацию в сжатое контекстное представление — латентное пространство, что используется для контекстного сжатия видео. Главным аспектом внедрения искусственных нейронных сетей в практические сценарии является высокая требовательность к ресурсам из-за высокой вычислительной сложности, что является критичным при работе с медицинским аппаратным обеспечением. В рамках исследования представлен процесс разработки экспериментального программного модуля для сжатия видеопоследовательностей, с подробным профилированием затрат вычислительной мощности при работе алгоритма сжатия, рассчитанного для работы в аппаратных модулях для проведения эндоскопии.

**Ключевые слова:** видеокодек, сжатие видеoinформации, искусственные нейронные сети, вычислительная сложность, эндоскопия

## I. ВВЕДЕНИЕ

Алгоритмы сжатия видео являются необходимым элементом записи, воспроизведения, предобработки и передачи визуальной информации в сетях связи различного назначения. Практическая важность экономии скорости цифрового потока обоснована ограниченностью пропускных способностей каналов связи, а также ограниченностью объёмов носителей для сохранения визуальной информации. Именно по этим причинам классические кодеки находят широкое применение в различных сферах — от видеонаблюдения до беспилотных аппаратов.

Сжатие видеоданных в сфере медицины является крайне востребованной технологией: телемедицина, малоинвазивная хирургия, эндоскопические осмотры — все перечисленные области медицины требуют как потоковой передачи сжатого видео, так и сжатия видеоматериалов для архивного хранения, например, историй болезней пациентов. Однако в данной сфере существуют специфические для предметной области ограничения, которые необходимо учитывать при

разработке систем сжатия видеoinформации: области интереса видеопотока при медицинском обследовании, требования к цветопередаче изображения для корректной оценки состояния внутренних тканей организма, а также требования к стабильности передачи изображения. Методы работы с видео, использующиеся в медицинской сфере, должны быть построены на основе данных ограничений и требований.

Одной из областей, в которой критически важна как передача видео в реальном времени, так и сжатие видео для последующего анализа, является эндоскопия, в том числе, и частная область эндоскопии — гастроскопия. Современные стенды для эндоскопических обследований используют программные реализации видеокодеков широкого назначения (H.264, H.265, VP9) для задач сжатия видеопотока, получаемого с камеры эндоскопа. Подход использования кодеков общего назначения имеет как достоинства, так и недостатки. Достоинства данного способа — с учётом устойчивости циклов разработки данных кодеков и единообразия их стандартов, их внедрение позволяет реализовывать сжатие видеопотока с высоким уровнем предсказуемости результата, а также использовать аппаратное ускорение сжатия видео (при совместимости целевой платформы). Недостатки способа — искусственные дефекты работы кодека (пикселизация, потеря информации о мелких деталях в кадре, искажения в цветовом представлении), а также унифицированность применения кодеков, отражающаяся на частичной неоптимальности использования вычислительных ресурсов при работе с медицинскими видеоданными. С учётом перечисленных аспектов, системы сжатия видео могут быть улучшены с соответствующими улучшениями, ориентированными для предметной области в медицине. В качестве данных улучшений могут рассматриваться искусственные нейронные сети, обеспечивающие корректное распознавание областей интереса, а также корректное использование алгоритмов представления информации в латентном пространстве нейронной сети для контентно-ориентированного сжатия.

Данное исследование предоставляет информацию о технических требованиях к системам сжатия видеoinформации в медицинских системах для проведения эндоскопии, использующихся аппаратных ресурсах, на платформе которых планируется внедрение алгоритмов искусственных нейронных сетей, а также о существующих ограничениях для внедрения аппарата искусственных нейронных сетей в эндоскопические

стенды. В ходе работы была получена модель нейронной сети с наиболее предпочтительной архитектурой для использования в эндоскопических стендах с учётом аспектов использования и аппаратных возможностей стендов.

## II. ИССЛЕДОВАНИЯ В ОБЛАСТИ ПРЕОБРАЗОВАНИЯ ВИЗУАЛЬНОЙ ИНФОРМАЦИИ ПРИ ЭНДСКОПИИ

Алгоритмы передачи видео при эндоскопии видоизменялись в соответствии с требованиями предметной области на протяжении продолжительного времени существования эндоскопии как отрасли [1]. Преобразования видеосигнала в тракте эндоскопического стенда обоснованы необходимостью регулярно видоизменять настройки камеры эндоскопа в зависимости от обследуемой области, а также для корректного анализа картинки на наличие или отсутствие дефектов. В числе контрольных элементов управления эндоскопом существуют регуляторы экспозиции, а также наборы предустановленных корректировок, направленные на обнаружение дефектов в сосудах органов и на внутреннем покрытии.

Идея использования алгоритмов искусственных нейронных сетей также реализовывалась в различных теоретических исследованиях, преимущественно, для анализа изображений и автоматического обнаружения патологий во внутренних обследуемых органах [2]. Также, существующие методы обнаружения патологий встраиваются в программные разработки для эндоскопических стендов, в том числе, и с совместимостью со стандартом обмена медицинскими изображениями DICOM [3]. Это позволяет сделать вывод, что аппаратные ресурсы эндоскопических установок имеют возможность локального запуска специализированных моделей нейронных сетей, предобученных с целью распознавания патологий внутренних органов.

Задача сжатия визуальной информации при эндоскопии рассматривалась и ранее, но, преимущественно, для статичных изображений [4]. Однако современные исследования рассматривают нейронные сети в том числе и как способ анализа контекста передаваемой информации, и приоритизации фрагментов видеозаписи на основе контекстуальной ценности отдельных кадров или последовательностей кадров. Примером данного подхода может служить исследование [5], в котором авторы используют нейронную сеть MSRCTNet для очистки видеопотока от кадров, не имеющих ценности для повторного осмотра. Существуют также и методы общей оптимизации алгоритмов работы эндоскопических стендов, например, для ПЛИС [6] или для исправления артефактов в каналах для эндомикроскопии [7]. Направления исследований демонстрируют перспективу внедрения алгоритмов нейронных сетей в существующие эндоскопические системы, с целью оптимизации различных параметров работы с визуальной информацией.

## III. СУЩЕСТВУЮЩИЕ СПЕЦИФИКАЦИИ АППАРАТНЫХ УСТАНОВОК ДЛЯ ЭНДСКОПИИ

Установки для эндоскопии являются комплексом оборудования для приёма и преобразования видеосигнала на основе различных аппаратных платформ. Многие из существующих на рынке

аппаратных продуктов для проведения эндоскопии созданы на базе центральных процессоров от компании Intel, количество оперативной памяти в установках варьируется от 8 Гб (портативные модели) до 64 Гб (стационарные установки высшей категории). Часть эндоскопических станций оснащена встроенным накопителем данных, что позволяет вести запись обследования непосредственно на носитель, включённый в установку.

С учётом того, что оборудование для эндоскопии не должно быть источником акустического шума с высоким уровнем (выше 40 дБ абсолютной шкалы), центральный процессор не имеет возможности использовать активное охлаждение. Данная причина обосновывает отсутствие специализированных графических процессоров в установках для эндоскопии, так как графические процессоры требуют системы активного охлаждения, которая в помещении процедурного кабинета создала бы слишком высокий уровень шума, влияющий на комфорт пациента и медицинских работников в условиях медицинского учреждения.

Ограничения, наложенные на процессорное тепловыделение, отражаются и на максимальной мощности, которой может обладать процессор — что в свою очередь ограничивает вычислительные ресурсы для аппаратной платформы в целом. Поэтому в качестве целевой платформы для запуска были выбраны центральные процессоры Intel Core i5/i7 поколений с 12-е по 14-е, с диапазоном тепловыделения от 15 до 28 Вт. Данная линейка процессоров позволит обрабатывать Full HD-поток с частотой кадров 30-60 к/с, при этом сохраняя возможность работы на системах с пассивным или минимальным активным охлаждением.

## IV. МЕТРИКИ ОЦЕНКИ КАЧЕСТВА ЭНДСКОПИЧЕСКОГО ВИДЕОПОТОКА

При разработке нейронной сети стоит также установить, какая именно функция задана как функция расчёта потерь. С учётом того, что в сжатии эндоскопического видео используются конечный видеопоток (сжатый) и исходный (эталонный), то следует применить объективные метрики оценки качества как функцию потерь. В качестве функций потерь следует использовать расчёт двух метрик в качестве функций потерь: PSNR и MS-SSIM, так как их комбинация позволит проанализировать результат с точки зрения наибольшего количества сценариев применения — PSNR позволит получить статистическое соответствие исходному кадру, в то время как MS-SSIM позволит получить соответствие качества передачи мелких деталей между двумя потоками.

Стоит также указать, что рассмотрение объективных метрик оценки качества не позволит в полной мере оценить качество передачи эндоскопического изображения. При анализе результатов следует также учитывать и специфические критерии для различных зон внутренних органов, относящиеся к медицинским визуальным данным, которые необходимо учитывать при оценке качества сжатия. В число таких категорий стоит включать индивидуальные для области требования: дефекты стенок органов, состояние кровеносных сосудов и иную ценную визуальную информацию. Необходимые для оценки отображения зон внутренних органов приведены в таблице.

ТАБЛИЦА I.

Зона	Диагностическое значение	Требования к сохранности
Границы воспалительных очагов	Оценка распространённости процесса	Чёткость контура, отсутствие размыва
Язвенные дефекты	Определение стадии, риска кровотечения	Различимость дна, краёв, периферии
Сосудистый рисунок	Выявление неоплазии, воспаления	Сохранность мелких сосудов, паттернов
Цвет и текстура слизистой	Функциональная оценка состояния	Корректная цветопередача, контраст

## V. НАБОРЫ ДАННЫХ ДЛЯ ОБУЧЕНИЯ МОДЕЛИ СЖАТИЯ ВИДЕО ПРИ ЭНДОСКОПИИ

В ходе разработки алгоритма обучения нейронной сети для сжатия эндоскопического видео главный отрицательный критерий — острый дефицит видеоданных в соотношении со статичными изображениями, что ограничивает моделирование взаимосвязей для нейронной сети и формирование временных зависимостей. Данный аспект является критичным для разработки контентно-ориентированного сжатия, моделирующего контекстную важность во времени. Ниже приведены существующие наборы данных, наиболее часто используемые в разработках, связанных с эндоскопией.

ТАБЛИЦА II.

Набор данных	Область	Объём	Разметка	Ограничения
EDD2020	Гастроскопия/колоноскопия	386 изображений	Сегментация 7 классов патологий	Только статичные кадры
PolypGen	Колоноскопия	Мультиданные	Полипы толстой кишки	Специфика колоноскопии
HyperKvasir	Гастроинтестинальная эндоскопия	110 тысяч изображений, 374 видео	Классификация, сегментация, верификация качества	Видео — короткие клипы, не полные процедуры
GastroNet-5M	Гастроскопия	5 миллионов изображений	Н/Д	Анонсирован в 2025, детали уточняются

Текущее положение, связанное с недостатком продолжительных видеоданных, не позволяет провести полноценный цикл разработки искусственной нейронной сети для подтверждения теоретических гипотез на практике. Однако обогащение данными возможно за счёт использования генеративных моделей искусственных нейронных сетей при дополнительной генерации синтетических данных, также, допустимы запросы в медицинские учреждения и получение доступа к записям операций эндоскопии в индивидуальном порядке и с обезличиванием данных пациентов.

## VI. РАЗРАБОТКА АРХИТЕКТУРЫ МОДЕЛИ НЕЙРОННОЙ СЕТИ ДЛЯ СЖАТИЯ ЭНДОСКОПИЧЕСКОГО ВИДЕО

При разработке кодера с алгоритмом нейросетевого сжатия было принято использовать архитектуру вариационного автоэнкодера с компенсацией движения. Энкодер является свёрточной сетью с последовательным понижением пространственного разрешения с использованием страйдрованных свёрток с пакетной нормализацией и нелинейными активациями. Стандартная конфигурация включает 4-5 блоков понижения, что позволяет преобразовать входное изображение 1920 на 1080 с тремя каналами в компактное представление латентного пространства 120 на 68 на 192, что позволяет получить коэффициент сжатия 16x по сравнению с пространственным измерением. Общая схема работы тракта видеосигнала с данным энкодером представлена на рис. 1.

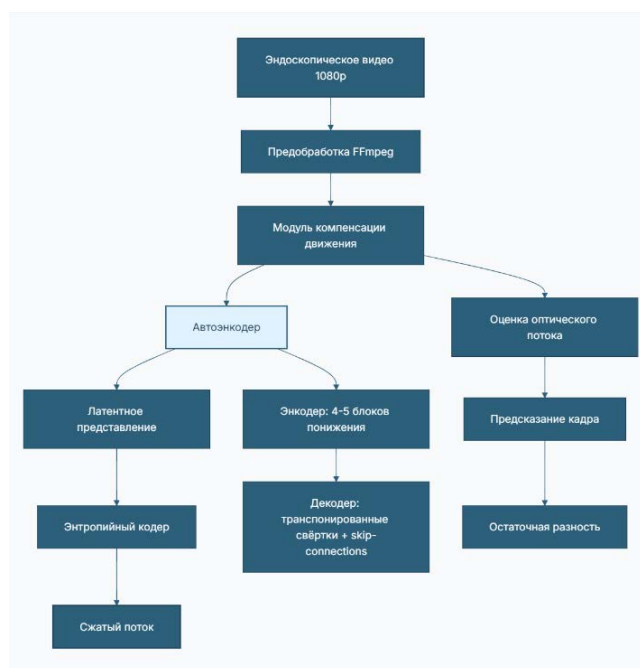


Рис. 1. Предлагаемая архитектура тракта видеосигнала в эндоскопическом стенде с модулем нейронного сжатия

В ходе проектирования декодера было принято решение использовать обратное свёрточное преобразование, с использованием соединений через пропуск, что позволяет компенсировать потерю информации на этапе понижения разрешения. Соединения через пропуск критически важны, так как их использование позволяет сохранить высокочастотные детали, которые необходимы для диагностической визуализации. Конечный слой с сигмоидной активацией восстанавливает оригинальное изображение. Полная схема слоёв энкодера и декодера представлена на рис. 2

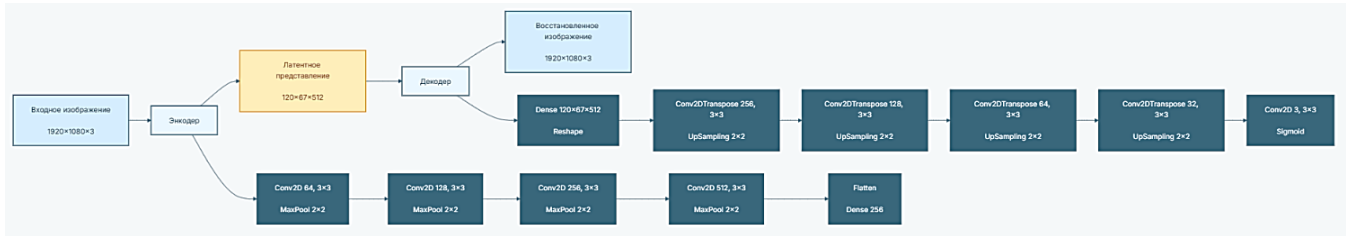


Рис. 2. Принципиальная схема архитектуры энкодера и декодера нейронной сети для сжатия эндоскопического изображения в разрешении Full HD

## VII. ЗАКЛЮЧЕНИЕ

В ходе исследования были выявлены основные аппаратные характеристики для эндоскопических станций, на основании данных характеристик были разработаны требования к использованию вычислительных мощностей в установках для эндоскопии. Также, была разработана архитектура нейронной сети, позволяющая сжимать видеoinформацию в латентное пространство, при этом, с возможностью работы на реальном эндоскопическом оборудовании. Недостаточное количество высококачественных данных, необходимых для корректного обучения модели нейронной сети, является препятствием к практической реализации данного кодека, однако, для тестовых сценариев, возможна генерация синтетических данных на основе статических последовательностей изображений.

## СПИСОК ЛИТЕРАТУРЫ

- [1] Авторское свидетельство № 1731156 А1 СССР, МПК А61В 1/04. Устройство для автоматического выбора экспозиции в эндоскопии: № 4780002: заявл. 09.01.1990 : опубл. 07.05.1992 / В.В. Балакирев; заявитель Всесоюзный научно-исследовательский институт медицинского приборостроения.
- [2] Применение искусственного интеллекта в видеокapsульной эндоскопии / И.Л. Клярская, Ю.А. Мошко, В.В. Кривой, И.А. Иськова, Е.В. Максимова, Е.И. Стилиди // Крымский терапевтический журнал. 2023. № 2. С. 11-22.
- [3] Свидетельство о государственной регистрации программы для ЭВМ № 2025669505 Российская Федерация. "Программное обеспечение "Рабочая станция для эндоскопии" EndoStation": заявл. 28.07.2025: опубл. 28.07.2025 / А.О. Маслов, И.А. Ларионов, А.Э. Талышинский; заявитель Общество с ограниченной ответственностью «БИС».
- [4] Метод сжатия изображений для беспроводной капсульной эндоскопии / Н.А. Козлов, В.А. Шурыгин, И.Ю. Жуков, Е.Д. Федоров, Е.В. Иванова, Д.М. Михайлов // Спецтехника и связь. 2011. № 6. С. 34-37.
- [5] Li Q. и др. MSRCTNet: a novel multi-scale capsule triplet network for efficient redundant frame removal in wireless capsule endoscopy videos // Sci Rep. 2026. Т. 16. № 1. С. 6902.
- [6] Zhang C. и др. Hardware-Aware Neural Architecture Search for Real-Time Video Processing in FPGA-Accelerated Endoscopic Imaging // Applied Sciences. 2025. Т. 15. № 20. С. 11200.
- [7] Hughes M. Real-time processing of fiber bundle endomicroscopy images in Python using PyFibreBundle. 2023.