

На правах рукописи



**Анциферова Анастасия Всеволодовна**

**Методы проектирования и тестирования алгоритмов оценки  
качества обработки и кодирования видеоданных**

Специальность 2.3.5 —  
«Математическое и программное обеспечение вычислительных систем,  
комплексов и компьютерных сетей»

**Автореферат**  
диссертации на соискание учёной степени  
кандидата физико-математических наук

Москва — 2024

Работа выполнена в лаборатории Компьютерной графики и мультимедиа Факультета вычислительной математики и кибернетики МГУ имени М.В.Ломоносова.

Научный руководитель: кандидат физико-математических наук  
**Ватолин Дмитрий Сергеевич**

Официальные оппоненты: **Визильтер Юрий Валентинович**,  
доктор физико-математических наук, профессор  
РАН,  
Федеральное автономное учреждение «Государственный научно-исследовательский институт авиационных систем»,  
руководитель научного комплекса «Искусственный интеллект и техническое зрение»

**Зайцев Алексей Алексеевич**,  
кандидат физико-математических наук,  
Автономная некоммерческая образовательная организация высшего профессионального образования «Сколковский институт науки и технологий»,  
старший преподаватель

Ведущая организация: Федеральное государственное учреждение «Федеральный исследовательский центр Институт прикладной математики им. М.В. Келдыша Российской академии наук»

Защита состоится 13 июня 2024 г. в 14 часов на заседании диссертационного совета 24.1.120.01 при Федеральном государственном бюджетном учреждении науки Институте системного программирования им. В. П. Иванникова Российской Академии Наук по адресу: 109004, г. Москва, ул. А. Солженицына, д. 25.

С диссертацией можно ознакомиться в библиотеке и на сайте Федерального государственного бюджетного учреждения науки Института системного программирования им. В. П. Иванникова Российской академии наук.

Автореферат разослан «\_\_\_» \_\_\_\_\_ 2024 года.

Ученый секретарь  
диссертационного совета  
24.1.120.01,  
кандидат физ.-мат. наук

Зеленов Сергей Вадимович

## Общая характеристика работы

**Актуальность темы.** На сегодняшний день видеоданные занимают более 70% всемирного интернет-трафика. Измерение качества видео имеет высокую значимость для разработки новых алгоритмов кодирования видео. Измерение качества видео проводится с помощью метода субъективного тестирования, в процессе которого происходит демонстрация последовательностей экспертам и сбор их мнения. Помимо субъективной оценки, осуществляемой человеком, существуют объективные методы оценки качества изображений и видео, производящие оценку алгоритмически. Данные алгоритмы основаны на решении задачи регрессии, входными данными в которой являются изображения или кадры видео, а метками — эталонные оценки качества, полученные от человека. Разработка методики сравнения и анализа устойчивости регрессионных моделей актуальна с прикладной точки зрения для решения задач, связанных с проектированием алгоритмов оценки качества видео.

При разработке и сравнении алгоритмов кодирования видео в основном используются традиционные алгоритмы объективной оценки качества, такие как PSNR и SSIM. Например, чтобы оценить эффективность нового стандарта кодирования видео, комитеты по стандартизации ITU-T, MPEG и JVET проводят тестирования новых стандартов на нескольких видео с помощью метода PSNR. При этом в многочисленных работах было показано, что PSNR имеет низкую корреляцию с эталонными оценками, полученными в процессе субъективных сравнений. Традиционные алгоритмы оценки качества, такие как PSNR и SSIM, были созданы задолго до появления новых стандартов кодирования видео, многие из которых используют нейросетевые компоненты. Поэтому традиционные алгоритмы показывают низкую точность при оценке искажений, производимых видеокодеками новых стандартов. Более новые алгоритмы оценки качества видео редко используются на практике, так как их точность обычно не воспроизводится на новых наборах данных. Сравнение алгоритмов объективной оценки качества видео производится в основном путем измерения корреляции с эталонными оценками, однако результаты могут отличаться в зависимости от выбора тестового набора видео и методики получения эталонной разметки качества.

В последние годы алгоритмы оценки качества видео используют подходы на основе машинного обучения и нейронных сетей. Нейросетевые методы чувствительны к возмущениям во входных данных, в том числе к состязательным атакам. Область устойчивости алгоритмов на основе нейронных сетей к состязательным атакам изучена в приложении к моделям классификации, например, в области компьютерного зрения. Однако методы состязательных атак могут также применяться к регрессионным моделям, в том числе алгоритмам оценки качества видео. С математической точки зрения, метод состязательной атаки является решением задачи оптимизации с ограничениями. Состязательная атака на алгоритмы оценки качества изображений или видео производится путем его обработки, которая увеличивает оценку качества объективным методом, и

при этом снижает визуальное качество. Ограничение может задаваться в виде нормы добавляемого в исходное изображение возмущения или максимально допустимого изменения какого-либо объективного метода оценки качества видео. Разработчики алгоритмов обработки и кодирования видео могут внедрять состязательные атаки в свои алгоритмы для получения более высоких позиций в рейтингах открытых сравнений. Также атака может быть совершена непреднамеренно. Например, если использовать неустойчивый к возмущениям алгоритм оценки качества видео в качестве критерия оптимизации для разработки какого-либо алгоритма обработки или сжатия видео, визуальное качество полученных видео может отличаться от объективных оценок, выдаваемых алгоритмом. Для появления надежных алгоритмов оценки качества необходимо в первую очередь разработать методику оценки их устойчивости, которая включает в себя методы состязательных атак на алгоритмы оценки качества видео.

Разработка методик сравнения и анализа устойчивости алгоритмов оценки качества видео актуальна для создания точных и надежных алгоритмов оценки качества видео. Однако, даже при наличии таких методик, невозможно разработать единственный общий для всех видео алгоритм объективной оценки качества, заменяющий субъективную оценку. В связи с большим разнообразием задач в области обработки и кодирования видео, объективные методы оценки качества являются узконаправленными, что позволяет им показывать высокую точность при оценке различных типов искажений в видеоданных. Для того чтобы разработать новый алгоритм объективной оценки качества, показывающий высокую корреляцию с эталонными оценками для определенных типов искажений в видео, требуется решить задачу регрессии с использованием набора видео, содержащего анализируемые искажения. Из-за высокой стоимости проведения субъективной оценки видео, необходимой для разметки их качества, количество видео в наборах данных, используемых для разработки алгоритмов оценки качества невелико по сравнению с наборами изображений. Задача получения эталонных оценок в случае измерения качества видео часто еще больше усложняется необходимостью использования специального оборудования, например, мониторов с поддержкой широкого диапазона битности. Из-за малых размеров наборов видео, актуальными являются методы оптимизации для решения задачи регрессии, устойчивые к переобучению и выбросам в обучающих данных. Одной из задач, для которых на практике полезно применение таких методов является оценка качества стереоскопического видео. Получение эталонных оценок для таких видео является дорогим, так как для субъективной оценки необходимо оборудование с поддержкой воспроизведения стереовидео (проектор или 3D-телевизор). В то же время оценка качества стереоскопического видео является актуальной, так как при просмотре некачественного стереоскопического видео у зрителей может возникать головная боль и дискомфорт. Низкое качество стереоскопического видео может характеризоваться разными причинами, среди которых искажения, появляющиеся в процессе съемки из-за

недостаточно точной калибровки камер. Существующие методы автоматического контроля качества стереофильмов способны обнаружить такие искажения, однако на сегодняшний день не существует алгоритмов, оценивающих качество стереоскопических фильмов в терминах оценки уровня дискомфорта. Так как бюджеты и график подготовки фильмов редко позволяют исправить все искажения стереопары, возникшие при съемке, актуальной становится задача разработки метода оценки качества стереоскопических видео с такими искажениями.

Таким образом, для развития области объективной оценки качества видеоданных необходимо проводить комплексное тестирование разрабатываемых алгоритмов, включающее оценку их качества и устойчивости к состязательным атакам, а также использовать методы проектирования алгоритмов в условиях малого количества обучающих данных.

**Целью** данной работы является разработка методов тестирования и проектирования алгоритмов оценки качества видеоданных, включающих в себя методику сравнения алгоритмов оценки качества кодированных видео, методы состязательный атак на алгоритмы оценки качества видео для тестирования их устойчивости, и метода проектирования алгоритма оценки качества видео в условиях нехватки данных.

Для достижения поставленной цели необходимо было решить следующие **задачи**:

1. Разработать методику сравнения алгоритмов оценки качества видео. Провести сравнение существующих алгоритмов для оценки качества кодирования видео.
2. Разработать методы состязательных атак на алгоритмы оценки качества видео и провести тестирование уязвимости алгоритмов оценки качества обработки изображений и видео.
3. Разработать метод оценки качества видео в условиях нехватки данных. Реализовать программную систему на основе предложенного метода для решения задачи оценки качества стереоскопических видео.

#### **Основные положения, выносимые на защиту:**

1. Методика тестирования алгоритмов оценки качества кодирования видео на основе нового набора видеоданных и вероятностная гарантия применимости модели Брэдли-Терри для сравнения качества видео.
2. Новый метод состязательной атаки на неререференсные алгоритмы оценки качества изображений и видео в режиме «белого ящика» с помощью создания универсального возмущения и аналитическая оценка его сходимости.
3. Новые методы состязательных атак на полнореференсные алгоритмы оценки качества изображений и видео с помощью оптимизации параметров обработки видео генетическим алгоритмом и с помощью обучения дифференцируемой аппроксимации.

4. Метод и реализация программной системы оценки качества стереоскопических видео.

**Научная новизна:** Предложена новая методика сравнения алгоритмов оценки качества видео, включающая подготовку и разметку видео. Теорема о вероятностной гарантии применимости модели Брэдли-Терри для оценки качества видео позволяет оценить вероятность получения однозначного ранжирования качества видео в ходе субъективных экспериментов. Разработаны новые методы состязательных атак на алгоритмы оценки качества видео: метод состязательной атаки с помощью обучения универсального возмущения, метод состязательной атаки путем оптимизации обработки видео генетическим алгоритмом, метод состязательной атаки путем создания аппроксимации недифференцируемого алгоритма оценки качества видео. Теорема о методе состязательной атаки на основе обучения универсального возмущения позволяет оценить точность приближения градиента предложенным методом.

**Практическая значимость** Предложенные в данной работе методы были реализованы в виде программных инструментов.

1. Создан новый набор видео и эталонных оценок качества, превосходящий существующие по количеству различных типов кодирования (47 видеокодексов, 2486 видео, более 766000 субъективных оценок). Набор данных опубликован на сайте <https://videoprocessing.ai/datasets/cvqad.html>.
2. Проведено самое большое в мире сравнение алгоритмов оценки качества видео по количеству проанализированных алгоритмов (41 алгоритм). Результаты сравнения опубликованы на сайте <https://videoprocessing.ai/benchmarks/video-quality-metrics.html>. Получены положительные отзывы на проведенное исследование от компаний Huawei, Tencent, Яндекс, Google (YouTube Media Algorithms), Dr Alan Bovik (самый цитируемый ученый, занимающийся разработкой алгоритмов оценки качества видео) и других.
3. Предложенный метод состязательной атаки на алгоритм VMAF был внедрен разработчиками из Google в видеокодек libaom. Впоследствии разработчики из Netflix выпустили улучшенную версию метода VMAF, которая более устойчива к данным преобразованиям.
4. Проведено сравнение устойчивости 15 методов оценки качества видео к 7 состязательным атакам. Результаты сравнения опубликованы на сайте <https://videoprocessing.ai/benchmarks/metrics-robustness.html>.
5. Создан новый набор стереоскопических видео и эталонных оценок их качества, превосшедший аналоги по количеству полученных субъективных оценок (302 зрителя и более 22000 субъективных оценок). Набор данных опубликован на сайте <https://videoprocessing.ai/datasets/mvdsvd.html>.

Результаты работы могут быть использованы в научных исследованиях и индустриальных разработках при тестировании и проектировании методов оценки качества видео.

**Методология и методы исследования.** В работе применялись методы линейной алгебры, теории алгоритмов, а также методы машинного обучения.

**Достоверность** Все предложенные методы и разработанная методика были реализованы и прошли экспериментальную проверку. Предложенные наборы данных были опубликованы.

**Апробация работы.** Основные результаты работы докладывались на:

1. International Conference on 3D Immersion (г. Брюссель, Бельгия, 11-12 декабря 2017 года).
2. X Международная научно-практическая конференция «Запись и воспроизведение объемных изображений в кинематографе и других областях», (ВГИК, г. Москва, Россия, 16-18 апреля 2018).
3. 29-я Международная конференция по компьютерной графике и машинному зрению ГрафиКон-2019 (г. Брянск, Россия, 23-26 сентября 2019).
4. GraphiCon2021 (31st International Conference on Computer Graphics and Vision) (Нижний Новгород, Россия, 24-30 сентября 2021).
5. 2021 Asia Digital Image Processing Conference (г. Киото, Япония, 17-19 декабря 2021).
6. Открытая конференция ИСП РАН (Москва, Россия, 2-3 декабря 2021).
7. Семинар кафедры интеллектуальных информационных технологий ВМК МГУ (26 мая 2022).
8. Семинар центра визуализации и спутниковых информационных технологий НИИСИ РАН (3 октября 2022).
9. Семинар института информационных технологий, математики и механики ННГУ им. Н.И.Лобачевского (10 ноября 2022).
10. Семинар им. М.Р. Шура-Бура ИПМ им. М.В. Келдыша (3 ноября 2022).
11. **CORE A** British Machine Vision Conference 2022 (Лондон, United Kingdom, 21-24 ноября 2022).
12. **CORE A+** Neural Information Processing Systems 2022 (Новый Орлеан, США, 28 ноября – 9 декабря 2022).
13. Семинар кафедры информатики и программного обеспечения БГТУ (17 марта 2023).
14. Семинар подразделения интеллектуального анализа данных и технического зрения ГосНИИАС (13 апреля 2023).
15. **CORE A+** 2023 International Conference on Learning Representations (ICLR) (Кигали, Руанда, 1-5 мая 2023).
16. Открытая конференция ИСП РАН (Москва, Россия, 4-5 декабря 2023).
17. Технологии ИИ в науке и образовании, Московский государственный университет им. М. В. Ломоносова (9 декабря 2023).
18. **CORE A+** Thirty-Eighth AAAI Conference on Artificial Intelligence (AAAI-24) (Ванкувер, Канада, 20-27 февраля 2024).

19. Семинар по системному программированию под руководством академика А.И. Аветисяна (27 марта 2024).

**Личный вклад.** В работах [1], [2] автору принадлежит разработка методики оценки качества видео. В работах [3], [4] автору принадлежит постановка задачи, выполнение теоретических исследований и курирование создания программной реализации предложенных методов состязательных атак на алгоритмы оценки качества видео. Автором лично была реализована программная система для оценки качества стереоскопических видео [5].

**Публикации.** Основные результаты по теме диссертации изложены в четырех работах, две из которых изданы в журналах, рекомендованных ВАК, и две в журналах, входящих в международные базы цитирования Scopus и Web of Science. Получено свидетельство о государственной регистрации программы для ЭВМ. Также результаты работы представлены в 12 статьях, опубликованных в трудах конференций, включая конференции рейтинга A/A+: NeurIPS 2022, BMVC 2022, ICLR 2023, AAAI 2024.

Диссертационная работа была выполнена при поддержке грантов:

- «Умник» Фонда содействия инновациям в рамках проекта «Разработка автоматической системы оценки степени усталости зрителей, вызываемой просмотром стереовидео»
- Поддержка молодых ученых и ученых без степени от Некоммерческого Фонда развития науки и образования «Интеллект»
- РФФИ 15-01-08632, РФФИ 19-01-00785.

## Содержание работы

Во **введении** обосновывается актуальность исследований, проводимых в рамках данной диссертационной работы, приводится обзор научной литературы по изучаемой проблеме, формулируется цель, ставятся задачи работы, излагается научная новизна и практическая значимость представляемой работы.

**Первая глава** содержит обзор существующих подходов к оценке качества видеоданных, включая методы субъективного тестирования и объективные алгоритмы. Приводится описание проблем с существующими методами, такими как трудность тестирования их точности на новых данных и сложность оценки устойчивости нейросетевых методов к атакам на оценку качества. На сегодняшний день результаты работы многих новых алгоритмов оценки качества изображений и видео часто не воспроизводятся на новых данных, отличных от использованных авторами. Одной из причин низкой точности работы на других данных является неполнота наборов данных, используемых для разработки алгоритмов оценки качества видео. Большинство существующих открытых наборов видео с субъективными оценками содержат только кодирование алгоритмами стандартов H.264/AVC и H.265/HEVC, которые отличаются от современных стандартов кодирования. В новых стандартах кодирования, таких

как H.266/VVC, LCEVC, AV1, на разных этапах кодирования и декодирования используются подходы на основе машинного обучения и нейронных сетей. Использование нейросетевых методов приводит к появлению новых типов искажений. Например, в стандарте LCEVC на этапе декодирования применяется повышение разрешения закодированного видео, современные реализации которого в основном нейросетевые. Видео, закодированное видекодеком стандарта LCEVC, визуально отличается от аналогичного закодированного с использованием стандарта H.265/HEVC: повышение разрешения, используемое LCEVC, восстанавливает больше деталей. Однако традиционные алгоритмы оценки качества были разработаны с использованием видео, закодированных старыми стандартами кодирования, и они часто не различают такие кадры или даже дают более низкую оценку качества видео с восстановленными деталями, так как данные восстановленные детали отличаются от деталей в исходном кадре.

Современные алгоритмы оценки качества видео, основанные на машинном обучении и нейронных сетях, значительно превосходят традиционные методы по корреляции с визуальным качеством. Однако, как и другие нейросетевые методы, алгоритмы оценки качества изображений и видео уязвимы к изменению входных данных, в том числе к состязательным атакам. Для анализа устойчивости современных методов вводится понятие состязательной атаки на алгоритм оценки качества: это любой вид обработки видео, при котором объективная оценка качества улучшается, а визуальное качество при этом снижается. На рисунке 1 продемонстрирован пример подобной атаки на алгоритм оценки качества PaQ-2-PiQ.



Рис. 1 — Пример состязательной атаки на алгоритм оценки качества PaQ-2-PiQ.

Примеры сценариев, в которых возможны атаки на алгоритмы оценки качества, включают в себя снижение визуального качества обработанных изображений и видео, возможность манипулирования результатами открытых сравнений и манипулирование результатами поиска изображений в интернете. Обработка видео, ориентированная на повышение оценок методов измерения качества видео, уже реализована во многих видекодеках. Например, видекодеки libaom и LCEVC имеют режимы, оптимизирующие выходной видеопоток для повышения оценки качества метода VMAF. Разработчики методов обработки изображений и видео могут использовать уязвимости алгоритмов оценки качества для достижения лучших результатов в открытых соревнованиях. В

некоторых соревнованиях, публикующих результаты визуальных сравнений совместно с объективными оценками качества алгоритмами, можно заметить разницу в лидерах этих двух номинаций. Другой пример это манипуляция результатами поисковой выдачи. Поисковые системы для ранжирования результатов выдачи изображений используют не только ключевые слова и описания, но и измерение их качества. Например, разработчики Microsoft Bing использовали качество изображения как один из показателей для улучшения точности поисковой выдачи.

Помимо задач в области тестирования алгоритмов оценки качества видео, обсуждается проблема, связанная с высокой стоимостью разметки данных для создания новых алгоритмов оценки качества видео. Разработка алгоритма оценки качества видео в условиях нехватки данных рассматривается на примере оценки качества стереоскопических видео. Задача устранения стереоскопических артефактов, вызывающих дискомфорт, остается актуальной из-за отсутствия соответствующих методов оценки качества. В отличие от обычного видео, где искажения кадра лишь снижают визуальное качество, артефакты стереоскопического видео могут вызвать различные неприятные симптомы, такие как визуальный дискомфорт. В данной главе исследуется влияние на визуальный дискомфорт искажений, появляющихся в стереофильмах, произведенных методом стереоскопической съемки, и рассматриваются следующие типы искажений: геометрические (поворот одного из ракурсов относительно другого, различные масштаб ракурсов), искажения цвета в стереопаре, временной сдвиг между ракурсами стереопары. Подобные искажения возникают в процессе съемки фильма из-за недостаточно точной калибровки камер, снимающих видео для левого и правого ракурса стереофильма. Такие искажения провоцируют высокую нагрузку на зрительную систему человека, которая приводит к появлению дискомфорта, головной боли и других симптомов. Существуют методы, способные обнаружить стереоскопические искажения, однако они обладают большой долей ошибок первого рода, а ручное исправление всего некачественного материала может значительно повысить стоимость и продлить сроки производства стереофильма.

**Вторая глава** посвящена разработке новой методики сравнения алгоритмов оценки качества видео, закодированных современными стандартами видеокodeков. Постановка задачи сформулирована в Разделе 2.1. Раздел 2.2 посвящен обзору существующих сравнений алгоритмов оценки качества видео, а также существующих наборов данных, используемых для сравнений. Большинство существующих сравнений опубликованы в статьях, авторы которых предложили новый алгоритм оценки качества видео и сравнили его с небольшим количеством существующих методов. Подобные сравнения часто основаны на открытых наборах данных, используемых и для разработки сравниваемых алгоритмов оценки качества **1**.

Предложенная в разделе 2.3 методика сравнения алгоритмов оценки качества закодированных видео состоит из трех этапов: создание набора тестовых

Набор	Кол-во видео	Средн. длит. (с)	Кол-во иск. видео	Искажение	Метод эталон. оценок	Зрители	Ответы
MCL-JCV (2016)	30	5	1560	Кодирование	In-lab	150	78 тыс.
VideoSet (2017)	220	5	45760	Кодирование	In-lab	800	-
UGC-VIDEO (2020)	50	> 10	550	Кодирование	In-lab	30	16,5 тыс.
CVD-2014 (2014)	5	10-25	234	In-capture	In-lab	210	-
LIVE-Qualcomm (2016)	54	15	208	In-capture	In-lab	39	8.1 тыс.
GamingVideoSET (2018)	24	30	576	Кодирование	In-lab	25	-
KUGVD (2019)	6	30	144	Кодирование	In-lab	17	-
KoNViD-1k (2017)	1200	8	1200	In-the-wild	Crowdsorce	642	205 тыс.
LIVE-VQC (2018)	585	10	585	In-the-wild	Crowdsorce	4776	205 тыс.
YouTube-UGC (2019)	1500	20	1500	In-the-wild	Crowdsorce	>8000	600 тыс.
LSVQ (2020)	39075	5-12	39,075	In-the-wild	Crowdsorce	6284	5 млн.
Предложенный: открытая часть	36	10, 15	1022	Кодирование (32 кодека)	Crowdsorce	10,800	320 тыс.
Предложенный: скрытая часть	36	10, 15	1464	Кодирование (51 кодека)	Crowdsorce	10,800	446 тыс.
Предложенный	36	10, 15	2486	Кодирование (83 кодека)	Crowdsorce	10800	766K

Таблица 1 — Обзор наборов видео с эталонными оценками и предложенный набор данных.

видео, получение эталонных оценок качества и сравнение методов на полученном наборе данных. Методика основана на новом подходе к созданию набора данных: он основан на видео, закодированных современными стандартами кодирования (включая стандарты H.266/VVC, LCEVC, AV1 и др.), и оценках визуального качества, полученных с помощью парных сравнений. Раздел 2.3.1 описывает процесс создания нового набора видео. Исходные видео должны обладать различной пространственно-временной сложностью. Предлагается использовать кластеризацию по пространственно-временной сложности с помощью алгоритма K-Means. Набор исходных видео кодируется с помощью нескольких видеокодеков различных стандартов, который должен включать H.264/AVC, H.265/HEVC, AV1, VVC, но может так же включать видеокодеки других стандартов. Чтобы увеличить разнообразие артефактов кодирования, предлагается использовать две разные предустановки кодирования для всех видеокодеков и три целевых битрейта: 1000 кбит/с, 2000 кбит/с и 4000 кбит/с.

В разделе 2.3.2 описан метод получения эталонных оценок качества видео. Оценки визуального качества для созданного набора видео предлагается получать с помощью краудсорсинговых парных сравнений. По сравнению с традиционным методом оценки видео зрителями в лабораторных условиях, краудсорсинговый подход позволяет получить оценки от большего количества зрителей. Разметка качества видео в режиме парных сравнений обладает рядом преимуществ перед независимой оценкой качества каждого видео: она не требует нормировки оценок каждого из зрителей, а также при добавлении новых видео в сравнение не требует калибровки шкалы оценок.

На вход алгоритму получения эталонных оценок качества подается набор видео, который состоит из  $N$  исходных видео и  $M_k$  закодированных версий каждого из исходных видео различными видеокодеками с  $K_k$  битрейтами. Количество видеокодеков и битрейтов может отличаться для разных исходных видео. Набор видео разбивается на  $N$  групп по  $M_k \times K_k + 1, k = 1, 2, \dots, N$  видео в

каждой, соответствующие закодированным версиям каждого исходного видео. Внутри каждой группы видео разбиваются на пары для дальнейшего попарного сравнения, всего  $(M_k \times K_k + 1 \times M_k \times K_k)/2$ ,  $k = 1, 2 \dots N$  пар. Каждая пара получает голоса от зрителей, где голос содержит информацию о том, какое из двух видео было выбрано как имеющее лучшее визуальное качество, либо были ли видео одинакового качества по мнению зрителя.

Для преобразования результатов голосования в конечные оценки качества видео применяется алгоритм Брэдли-Терри. Необходимо получить вектор из оценок  $\beta = \beta_1, \dots, \beta_n$ , отражающих качество видео из входного набора. Результат голосования  $\beta_i - \beta_j$  можно рассматривать как бернуллиевскую случайную величину, для которой вероятность  $p_{ij}$  того, что видео  $c_i$  лучше, чем  $c_j$  моделируется как:

$$\log \frac{p_{ij}}{1 - p_{ij}} = \beta_i - \beta_j \quad (1)$$

Начиная с произвольного  $p$ , путем максимизации логарифмической функции правдоподобия, получим искомый вектор эталонных субъективных оценок  $\hat{\beta} = \hat{\beta}_1, \dots, \hat{\beta}_n$ :

$$\hat{\beta} = \operatorname{argmax}_{\beta, \sum_i \beta_i = 0} l(\beta) = \operatorname{argmax}_{\beta, \sum_i \beta_i = 0} \sum_{i,j} C_{i,j} \ln \frac{e^{\beta_i}}{e^{\beta_i} + e^{\beta_j}}. \quad (2)$$

Стандартная ошибка для  $\beta_i - \beta_j$  может быть выражена как:

$$se_{\beta_i - \beta_j} = \sqrt{A_{i,i}}, A = \left( -\nabla_{\beta}^2 \{l(\beta)\} \Big|_{\hat{\beta}} \right)^{-1}. \quad (3)$$

**Теорема 1.** Пусть для всех возможных упорядоченных пар  $(i, j)$ ,  $i \geq j$  видео выполняется  $\min_{i,j} \delta_{ij} > 0$ , где  $\delta_{ij} = \operatorname{sign}(\hat{\beta}_i - \hat{\beta}_j) * \left( \hat{\beta}_i - \hat{\beta}_j \right) - \hat{se}_{\hat{\beta}_i - \hat{\beta}_j} * z_{\frac{1-\alpha}{2}}$ . Тогда можно установить отношение порядка между всеми видео из набора в соответствии с эталонными оценками  $\beta_1, \dots, \beta_N$  с вероятностью хотя бы  $p = (1 - \alpha)^{C_N^2}$ .

Раздел 2.3.4 описывает экспериментальную оценку предложенной методики. Исходные видео были выбраны из коллекции, состоящей из более чем 18000 видео с высоким битрейтом. Данная коллекция была получена из открытых источников, в основном <https://vimeo.com/>, на котором представлены как профессиональные ролики, так и видео, загружаемые пользователями. Путем кластеризации по пространственно-временной сложности было выбрано 36 исходных видео, которые далее были закодированы с помощью 22 различных видеокодеков: 11 видеокодеков стандарта H.265/HEVC, 5 видеокодеков стандарта AV1, 2 видеокодека стандарта H.264/AVC и 4 видеокодека других стандартов.

Для проведения краудсорсинговых сравнений использовались платформы <https://Subjectify.us/> и <https://toloka.ai/>. Из набора закодированных видео были сформированы пары, и видео из каждой пары показывались

зрителям последовательно в полноэкранном режиме. Участникам сравнений необходимо было выбрать видео с наилучшим качеством изображения или отметить, что оба видео одинакового качества. У зрителей также была возможность воспроизвести видео заново. В ходе процедуры одного сравнения участники сравнивали 12 пар видео. Ответы участников, давших неправильные ответы на проверочные вопросы, исключались. Для каждой пары видео было получено не менее 10 оценок. В ходе визуальных сравнений было получено 766362 оценок от более чем 10800 человек.

В Разделе 2.3.5 представлены результаты сравнения алгоритмов оценки качества видео, проведенного по предложенной методике. Созданный набор данных был разделен на две части: открытую и закрытую (40% и 60% от всего набора данных, соответственно). Открытая часть была опубликована поле публикации результатов проведенного сравнения, а закрытая часть используется для дальнейшего внутреннего тестирования новых методов, присылаемых авторами для сравнения с существующими. Наличие закрытой части позволяет более объективно оценивать работу методов, основанных на машинном обучении, и предотвращает возможность их переобучения на всем наборе данных. Новый набор данных использовался для проведения сравнения современных методов оценки качества изображений и видео с открытой реализацией. Всего было проанализировано 14 полнореференсных методов и 12 неререференсных методов оценки качества видео. Сравнение показало, что некоторые новые неререференсные методы, такие как MDTVSFA, по точности близки к полнореференсным. Метод VMAF показал самую высокую корреляцию с эталонными оценками. Результаты проведенного сравнения методов оценки качества изображений и видео были опубликованы онлайн. С момента публикации и на начало 2024 года было получено 15 заявок добавления новых методов в опубликованное сравнение от различных команд разработчиков.

В разделе 2.4 кратко описано проведенное по предложенной методике и опубликованное онлайн сравнение алгоритмов оценки качества видео.

**Третья глава** посвящена разработке методов состязательных атак на алгоритмы оценки качества изображений и видео. В разделе 3.1 сформулирована постановка задачи. Входные и выходные данные алгоритма состязательной атаки зависят от типа атаки. Большинство современных неререференсных методов оценки качества дифференцируемы, поэтому для атак на них применимы методы оптимизации, основанные на градиентном спуске. Для состязательных атак на неререференсные алгоритмы оценки качества видео на вход алгоритму подается изображение (кадр видео)  $x$  и алгоритм оценки качества  $M(x)$ . На выходе алгоритм должен выдавать изображение  $x^*$  такое, что

$$M(x^*) > M(x), \|x - x^*\|_q < \epsilon, \quad (4)$$

где  $q, \epsilon$  — параметры алгоритма. На сегодняшний день самую высокую корреляцию с визуальным качеством показывает метод VMAF. Однако данный метод

не дифференцируем, поэтому для атаки на него применимы методы оптимизации, основанные на генетических алгоритмах. Для состязательной атаки на алгоритм оценки качества видео VMAF на вход алгоритму подается набор кадров  $X = x_{i=1,N}$ , а на выходе алгоритм должен выдать набор преобразованных кадров видео  $X^* = x_{i=1,N}^*$  такой, что

$$VMAF(X^*) > VMAF(X), \|SSIM(X^*) - SSIM(X)\|_q \leq \epsilon, \quad (5)$$

где  $\epsilon$  — параметр алгоритма. Для атаки на метод VMAF была также предложен метод состязательной атаки, основанный на создании аппроксимации. Постановка задачи для нее совпадает с постановкой для предыдущего метода атаки.

Раздел 3.2 посвящен обзору существующих методов состязательных атак на алгоритмы оценки качества изображений и видео. Ранее в области оценки устойчивости нейросетевых моделей в основном рассматривались только модели распознавания объектов. Существует лишь несколько работ, посвященных анализу устойчивости методов оценки качества видео. В работе Ciaramello и Reibman впервые предложили проводить проверку надежности методов оценки качества изображений как дополнение к стандартному тестированию их производительности. Однако рассмотренные методы атак скорее были способами демонстрации ошибок в работе методов, не распространяемых на новые изображения и видео. В других работах авторы предложили итеративные методы состязательных атак на алгоритмы оценки качества видео, однако из-за низкой скорости работы и необходимости проведения субъективных сравнений для генерации атак данные методы трудно применять для видео: они слишком медленные, а значит разработчики алгоритмов не будут внедрять такие атаки в свои методы для победы в сравнениях. Таким образом, для анализа устойчивости методов оценки качества видео в условиях реальных задач необходимо разработать методы состязательных атак, работающие с высокой скоростью и распространяемые на другие видео помимо тех, для которых данные атаки обучались.

Раздел 3.3 посвящен описанию предложенных методов состязательных атак на алгоритмы оценки качества видео. Предложенный в 3.3.1 метод состязательной атаки на неререференсные алгоритмы оценки качества видео основан на создании универсальных возмущений, которые при добавлении к исходному кадры видео повышают оценку его качества. Пусть  $x_i$  — изображение из обучающего набора  $x_i \in D$ ,  $D \in \mathbb{R}^{3 \times W \times H}$ ,  $N$  — размер обучающего набора, а  $M$  — метод оценки качества изображений или видео. Состязательная атака на метод оценки качества видео с помощью универсального возмущения  $U$  формулируется следующим образом:

$$U = \underset{f}{argmax} \frac{1}{N} \sum_{i=1}^N (M(x_i + U) - M(x_i)), \|U\|_\infty < \epsilon. \quad (6)$$

На каждой итерации путем вычисления градиента атакуемого нейросетевого метода оценки качества по входному изображению определяется возмущение, необходимое для увеличения оценки целевого алгоритма оценки качества.

Обновление возмущения происходит с помощью оператора проекции. Алгоритм работает до тех пор, пока средний показатель атакуемого алгоритма оценки качества на обучающем наборе изображений не будет незначительно отличаться между соседними итерациями или пока не будет достигнуто максимальное количество итераций. Алгоритм 1 описывает предложенный метод обучения универсального возмущения  $U$ .

**Input:** обучающий набор изображений  $\mathbf{D}$ , атакуемый неререференсный алгоритм оценки качества  $\mathbf{M}$ ,  $\mathbf{L}_p$ -норма возмущения величины  $\epsilon$

**Output:** Универсальное возмущение  $U$

Initialize  $U \leftarrow 0$ ;

**while** *not converged* **do**

**for** *each image*  $x$  **in**  $\mathbf{D}$  **do**

$x' \leftarrow x + U$ ;

$x' \leftarrow \text{clamp}(x', [0, 1])$ ;

        Compute the gradient of  $\mathbf{M}$  with respect to  $U$ ;

        Update  $U$  using the gradient;

$U \leftarrow \mathbf{P}_{p,\epsilon}(U)$ ;

**end**

**end**

**Algorithm 1:** Генерация универсального возмущения предложенным методом

**Теорема 2.** Пусть дан обучающий набор данных  $D = (I^1, I^2, \dots, I^N)$  и  $U$  — универсальное возмущение, полученное с помощью предложенного метода и набора данных  $D$ . Пусть значение градиента  $G^k$  всех изображений из обучающего набора  $D$  в точке  $(u, v)$  имеет один знак. Тогда в точке  $(u, v)$  точность приближения градиента  $G^k$  произвольного изображения может быть вычислена по обучающему набору данных  $D$  и задается выражением:

$$\left| |U_{(u,v)}| - |G^k_{(u,v)}| \right| = \left| \left( \left| C_\epsilon \sum_{i=1}^N G^i \right| - |G^k| \right)_{(u,v)} \right|. \quad (7)$$

**Следствие 1.** Предложенный метод сходится за одну итерацию, то есть за один проход по обучающему набору данных.

В разделе 3.3.2 предложены методы состязательной атаки на недифференцируемые полнореференсные алгоритмы оценки качества видео на примере алгоритма VMAF. Первый предложенный метод основан на оптимизации параметров алгоритмов обработки видео, повышающих VMAF, с помощью генетического алгоритма.

Предложенный метод атаки на VMAF основан на оптимизации параметров расширенного набора алгоритмов обработки видео. В него входят алгоритмы

тонового картирования (метод тонового картирования Драго, метод тонового картирования Мантюка, Retinex и метод тонового картирования Рейнхарда, адаптивная эквализация гистограммы с ограничением контраста (CLANE), эквализация гистограммы, преобразования на основе контраста (гамма-коррекция, преобразование контраста и линейное преобразование контраста) и нерезкое маскирование. Большинство из данных методов имеют несколько параметров, и каждый параметр экспоненциально увеличивает вычислительную сложность поиска оптимальной комбинации. Чтобы уменьшить количество тестируемых параметров и ускорить сходимость, предлагается использовать генетический алгоритм ( $\mu + \lambda$ ). Так как кодирование видео сглаживает добавленные возмущения, его применение к атакованным видео снижает эффективность атаки. Поэтому в процессе подбора параметров алгоритмов обработки видео предлагается использовать кодирование видеокодеком x264.

Для ускорения метода предложена модификация состязательной атаки на алгоритм оценки качества видео VMAF с помощью обучения дифференцируемой аппроксимации VMAF. Архитектура предложенной нейронной сети, осуществляющей обработку видео для атаки с помощью создания аппроксимации VMAF, показана на Рис. 2.

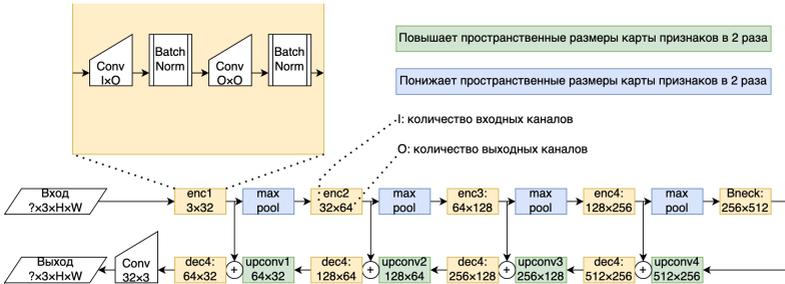


Рис. 2 — Архитектура сети на базе U-Net для состязательной атаки на алгоритм оценки качества видео VMAF.

Нейронная сеть, выполняющая атаку путем обработки видео, принимает на вход набор трехканальных изображений  $B_{in} \in R_{[0..1]}^{batch \times 3 \times W \times H}$ . На выходе — набор трехканальных изображений  $B_{out} \in R_{[0..1]}^{batch \times 3 \times W \times H}$ . Обозначим данную нейронную сеть  $G, G(B_{in}) = B_{out}$ . Аппроксимирующая оценки VMAF сеть  $P$  принимает на вход набор пар трехканальных изображений  $B_{in} \in R_{[0..1]}^{batch \times 6 \times W \times H}$  и возвращает  $Q_{pred} \in R_{[0..1]}^{batch}$  — значение оценки качества VMAF для каждой пары входных изображений. Обозначим аппроксимирующую нейронную сеть  $P$ .  $P_{frozen}$  обозначает, что в ходе применения параметры нейронной сети заморожены и не накапливают градиенты функции потерь.

В ходе обучения оптимизировалась функция потерь, состоящая из взвешенной суммы нескольких компонент  $L = L_{vmaf} + \lambda_{proxy} L_{proxy} + \lambda_{stable} L_{stable}$ .

Основная сеть обучалась обрабатывать изображение так, чтобы максимизировать оценку VMAF полученного на выходе изображения по сравнению со входным. В силу того, что сама функция вычисления VMAF не является дифференцируемой, в функции потерь использовалась вспомогательная нейронная сеть:  $L_{\text{vmaf}} = M - P_{\text{frozen}}(G(B_{\text{in}}), B_{\text{in}})$ , где  $M$  — произвольная константа, принятая равной 1. Вспомогательная нейронная сеть  $P$  решает задачу аппроксимации VMAF для искаженного и исходного изображения:  $L_{\text{proxu}} = \left( P(G_{\text{frozen}}(B_{\text{in}}), B_{\text{in}}) - \text{VMAF}(G_{\text{frozen}}(B_{\text{in}}), B_{\text{in}}) \right)^2$ . Также в ходе обучения использовалась мера близости в евклидовом пространстве изображений в качестве стабилизирующей добавки к функции потерь:  $L_{\text{stable}} = \|G(B_{\text{in}}) - B_{\text{in}}\|_2$ . Вместо L2-нормы можно использовать другие меры близости, например, алгоритм оценки качества SSIM:  $L_{\text{SSIM}} = \|\text{SSIM}(G(B_{\text{in}})) - \text{SSIM}(B_{\text{in}})\|_2$ .

В разделе 3.3.3 описан метод тестирования устойчивости алгоритмов оценки качества видео к состязательным атакам. В нем предлагается использовать различные алгоритмы оценки качества для измерения потерь визуального качества видео после применения методов состязательных атак, а так же разную силу атаки путем добавления универсального возмущения с разной амплитудой. Также в предложенном методе учитывается разная степень сжатия видео. С помощью данного метода проведено тестирование устойчивости семи неререференсных алгоритмов оценки качества изображений и видео.

Эффективность предложенных методов продемонстрирована с помощью объективной и визуальной оценки в разделе 3.4. С помощью первого метода на наборе изображений COCO были обучены универсальные возмущения для семи неререференсных методов оценки качества видео (PaQ-2-PiQ, Linearity, VSFA, MDTVSFA, Koncept512, Nima и SPAQ). Полученные универсальные возмущения для всех целевых методов были применены к 200 изображениям размером  $256 \times 256$  из набора данных COCO. Также предложенный метод атаки был применен с восемью уровнями амплитуды: к исходному изображению было добавлено обученное возмущение, умноженное на масштабирующий коэффициент (0,02, 0,04, 0,06, 0,08, 0,1, 0,2, 0,4 и 0,6). В проведенных экспериментах атака методом универсального возмущения показала высокую скорость работы, превосходящую ближайший аналог — атаку MADC. В сравнении с этой атакой, предложенный метод обеспечивает 70% повышения значения метода всего за 3% времени обработки.

С помощью предложенного метода сравнения устойчивости методов оценки качества изображений было вычислено повышение целевого атакуемого метода оценки качества и снижение прокси-методов оценки качества изображений. Для этого было использовано 20 видео разрешения FullHD из набора данных Xiph.org, четыре уровня амплитуды атаки (0,02, 0,04, 0,06 и 0,08) и кодирование с использованием видекодека x264 стандарта H.264 с четырьмя битрейтами (200 кбит/с, 1000 кбит/с, 5000 кбит/с и 12000 кбит/с). Согласно полученным результатам, четыре протестированных метода (PaQ-2-PiQ, Linearity,

VSFA и MDTVSFA) оказались уязвимыми к универсальным атакам возмущения, в то время как три (Koncept512, Nima и SPAQ) оказались устойчивыми к ним.

Результаты анализа эффективности атаки на VMAF при помощи генетических алгоритмов описаны в 3.4.2 и показали, что метод CLANE обеспечил наилучшее увеличение VMAF: среднее повышение оценок метода составило 86,8%, медианное значение — 50,2%; самое большое повышение — 218,8%. Для VMAF NEG комбинация гамма-коррекции и unsharp mask показал повышение значений метода в среднем на 6%. При этом видео, обработанные предложенным методом, при выполнении атаки на алгоритм VMAF NEG не содержат видимых искажений, в отличие от атаки на алгоритм VMAF. В сравнении предложенных методов повышения VMAF и VMAF NEG с методами, описанными в других работах, предложенный метод обеспечивает наибольшее повышение оценок VMAF.

Для предложенного метода на основе создания аппроксимации была рассмотрена процедура обучения на наборах данных Pascal VOC и Vimeo 90K. В качестве тестового набора видео использовался набор видео Derf's collection, который содержит видео длиной 60 кадров. Для этого видео с и без обученной обработки были закодированы видеокодеком х264 с битрейтами 60 кбит/с, 250 кбит/с, 1000 кбит/с, 2000 кбит/с, 4000 кбит/с, 8000 кбит/с и вычислены значения экономии битрейта на основе метода BSQ-Rate. Наилучшая обученная модель получена с использованием SSIM-регуляризации и фильтрацией 50% наиболее трудной части тренировочной выборки, она показывает BSR-Rate равный 0.541 и повышает VMAF в сценарии без дальнейшего кодирования на 19.07 и 17.03 единицы в среднем на двух наборах данных.

В разделе 3.5 кратко описано проведенное и опубликованное онлайн сравнение устойчивости алгоритмов оценки качества видео, использующее предложенные методы атак.

**Четвертая глава** посвящена созданию метода оценки качества в условиях нехватки данных на примере оценки качества стереофильмов. В разделе 4.1 сформулирована постановка задачи, которая заключается в разработке метода оценки качества стереоскопического видео а с различными стереоскопическими искажениями. Метод на вход должен принимать значения  $\{X_i^k\}_{i=1,N;k=1,S}$  стереоскопических искажений и признаки видео, а на выходе выдавать оценку качества для сцен стереофильма  $\{Y_k\}_{k=1,S}$ , где  $N$  — количество используемых показателей (типов искажений и признаков видео),  $S$  — количество сцен фильма.

В разделе 4.2 рассмотрены существующие исследования по оценке визуального качества стереоскопических видео. В основном существующие методы в данной области были посвящены оценке визуального качества, а не визуального дискомфорта. Последние были основаны в основном на низкоуровневых характеристиках стереоскопических видео, таких как величина диспаратности, временная и пространственная сложность, скорость движения объектов в кадре и скорость движения камеры. Однако ранее исследователями практически не рассматривались стереоскопические искажения, возникающие при стереосъемке.

Раздел 4.3 посвящен разработке метода оценки качества стереоскопического видео.

В разделе 4.5 приведено описание предложенного метода оценки качества стереоскопического видео. Поставленная задача является задачей регрессии и для ее решения были применены методы машинного обучения с использованием полученного набора данных. Предложенный метод заключается в оптимизации функции потерь Хьюбера с L2-регуляризацией. Формально, выбор коэффициентов  $\omega_i$  для признаков качества стереофильма проводился оптимизацией следующего функционала:

$$Y_k = \alpha + \sum_{i=1}^N X_i^k \omega_i + \epsilon, \quad (8)$$

где  $\alpha, \omega$  — параметры регрессии, определяемые при решении задачи,  $\epsilon$  — аддитивная случайная величина, имеющая нормальное распределение с нулевым средним и дисперсией  $\sigma^2$ ,  $N$  — количество признаков. Функция потерь Хьюбера  $H_m(z)$  накладывает меньший штраф на точки с большим отклонением, и при решении задачи регрессии производится оптимизация по параметрам  $\omega$  (коэффициенты вклада признаков) и  $\sigma$  (масштабирующий параметр, позволяющий задавать пороговое значение  $\epsilon$  функции потерь Хьюбера):

$$\min_{\omega, \sigma} \sum_{k=1}^S \left( \sigma + H_m \left( \frac{X^k \omega - Y_k}{\sigma} \right) \sigma \right)^2, \text{ где} \quad (9)$$

$$H_m(z) = \begin{cases} z^2, & |z| < \epsilon; \\ 2\epsilon|z| - \epsilon^2, & |z| \geq \epsilon. \end{cases} \quad (10)$$

Обучение метода проводилось с использованием предложенного нового набора данных. В качестве признаков, описывающих сцены, продемонстрированные во время эксперимента зрителям, были использованы 15 показателей технического качества стереофильмов, включающие оценки яркости, движения, геометрических искажений, временного сдвига, цвета и резкости для левого и правого ракурсов.

Результаты экспериментальной оценки предложенного метода приведены в разделе 4.4. Для обучения метода был подготовлен новый набор данных, описанный в Разделе 4.4.1. Для получения информации о восприятии зрителями различных стереоскопических артефактов была проведена серия экспериментов, в которой участникам предлагалось оценивать качество специально подготовленного стереоскопического видео. Подготовленное видео состояло из 60 тридцатисекундных фрагментов, взятых из четырех художественных стереофильмов. Были выбраны сцены с различной интенсивностью движения, глубиной и яркостью. Видео были модифицированы путем добавления одного из четырех исследуемых типов искажений: поворот одного из ракурсов относительно другого, несоответствие масштаба ракурсов, цветовое несоответствие

и временное несоответствие между ракурсами, 20 видео содержали комбинацию двух различных искажений. Каждое искажение присутствовало с одним из пяти уровней интенсивности, которые были выбраны в соответствии с распределением значений данного искажения в художественных фильмах. Общая длительность составленного тестового стереофильма составила сорока минут.

Эксперимент проводился в аудитории П-13 факультета ВМК МГУ, оборудованной двумя профессиональными проекторами Digital Projection Titan 1080p-700 с линейной поляризацией, разрешение каждого проектора  $1920 \times 1080$ , экран с серебряным покрытием диагональю 9 метров. Из полученных с помощью проведенных экспериментов ответов были исключены выбросы — ответы участников, которые сильно отличались от остальных. С помощью комбинации четырех методов обнаружения аномалий в данных (метода изолирующего леса, метода эллипсоидной аппроксимации, метода обнаружения локальных аномалий и кластеризации с помощью DBScan) были исключены ответы 13 зрителей. В таблице 2 приведено сравнение проведенного эксперимента и полученного набора данных с существующими исследованиями визуального качества стереоскопического видео.

Таблица 2 — Эксперименты, проведенные в рамках исследований визуального качества стереоскопического видео.

Университет	Кол-во участников	Длит.	Год	Метод показа 3D	Тип искажений
Catholic University of Korea	20	18 мин	2011	Пассивная	Снятые S3D послед., разл. параллакс и движение
Telecom Innovation Labs	24	50 мин	2011	Активная	База S3D видео, разл. параллакс
Philips Research Labs	24	24 мин	2011	Авто-стереоскопич.	Стереофильм (конвертир.), разл. параллакс
Beijing Institute of Ophthal.	30	30 мин	2012	Активная, пассивная	Стереофильм (снятый)
LUNAM University	29	28 мин	2012	Активная	База S3D видео, 10 типов деградаций
Tampere University of Tech.	10	45 мин	2013	Пассивная	Снятые S3D послед., разл. кодирование
LUNAM and Yonsei U., Acreo Inst.	29, 28, 48	29 мин	2013	Активная, пассивная	База S3D видео, 10 типов деградаций
Yonsei University	56	10,5 мин	2014	Авто-стереоскопич.	База S3D видео, разл. SI/ТИ
Roma Tre University	854	~90 мин	2015	Пассивная	Опрос после просмотра стереофильмов в кинотеатрах
University College Dublin	4 для кажд. видео	5.5 мин	2015	Активная	База S3D видео, имитирование потерь при трансляции
University of British Columbia	88	35 мин	2016	Пассивная	База S3D видео, разл. SI/ТИ, кодирование
University North	18	8 мин	2016	Активная	База S3D видео, 22 типа деградаций (кодир., размер, геометрич. и др.)
<b>Предложенный</b>	302	40 мин	2017	Пассивная	Сцены из снятых стереофильмов, 4 типа искажений, 5 уровней

В Разделе 4.4.2 приведено сравнение предложенного метода с аналогами. Точность предложенного метода при тестировании составила 0,881. Большинство открытых наборов данных, состоящих из стереоскопических видео и оценок визуального качества для них, нацелены на оценку качества кодирования и не содержат стереоскопические артефакты. Корреляция значений, полученных предложенной моделью на наиболее близком по тематике наборе данных 3DVQA, содержащем видео с разной диспаратностью, составила 0,61, что превосходит результаты другого метода. Предложенный метод был применен для анализа 60 полнометражных стереофильмов. Анализ показал, что большинство фильмов, который могут вызвать визуальный дискомфорт при просмотре, имеют низкий бюджет производства.

В заключении приведены основные результаты работы, которые заключаются в следующем:

1. Разработана методика тестирования алгоритмов оценки качества видео, основанная на новом наборе видео. Предложенный набор состоит из 2486 видео, закодированных 47 видеокodeками 10 различных стандартов, и эталонных оценок от 18000 зрителей. С помощью предложенной методики проведено и опубликовано сравнение 41 алгоритма оценки качества видео. Получена вероятностная гарантия применимости модели Брэдли-Терри для оценки качества видео.
2. Разработаны новые методы состязательных атак на алгоритмы оценки качества видео: на основе универсальных возмущений, обработки видео и создания аппроксимации. Экспериментальная оценка их эффективности показала, что метод на основе создания универсальных возмущений превосходит аналоги по скорости работы в 10-200 раз и обеспечивает сопоставимое повышение объективных оценок качества, а методы атак на основе обработки видео и создания аппроксимации обеспечивают повышение объективных оценок качества на 20-220%.
3. Разработан метод оценки качества стереоскопического видео, основанный на оценке стереоскопических искажений. Экспериментальная оценка предложенного метода проведена на новом наборе данных, содержащем 22200 эталонных оценок, и показала, что предложенный метод превзошел нейросетевой аналог, показав корреляцию 0,88 с визуальными оценками. Метод прошел экспериментальную апробацию при оценке 60 полнометражных стереофильмов.
4. Предложенные методы были реализованы и прошли экспериментальную оценку. Предложенные новые наборы данных были опубликованы.

## Публикации автора по теме диссертации

1. *Zvezdakova (Antsiferova) A. V., Kulikov D. L., Zvezdakov S. V., Vatolin D. S.* BSQ-rate: a new approach for video-codec performance comparison and drawbacks of current solutions [Text] // Programming and computer software. — 2020. — Vol. 46, no. 3. — P. 183—194. — (Scopus, WoS).
2. *Звездакова (Анциферова) А. В., Куликов Д. Л., Звездаков С. В., Ватолин Д. С.* BSQ-rate: новый подход к сравнению производительности видеокодеков и недостатки существующих решений [Текст] // Труды Института системного программирования РАН. — 2020. — Т. 32, № 1. — С. 89—108. — (БАК).
3. *Shumitskaya E., Antsiferova A., Vatolin D.* Towards adversarial robustness verification of no-reference image- and video-quality metrics [Text] // Computer Vision and Image Understanding. — 2024. — Vol. 240. — P. 103913. — (Scopus, WoS).
4. *Solovev A., Antsiferova A., Vatolin D., Galaktionov V.* Development of neural network-based video preprocessing method to increase the VMAF score relative to source video using distillation [Текст] // Keldysh Institute preprints. — Moscow, 2023. — № 66. — С. 1—11. — (БАК).
5. *Свидетельство о гос. регистрации программы для ЭВМ.* Программный комплекс для определения степени дискомфорта зрителей при просмотре стереофильма по данным его технического качества [Текст] / А. В. Анциферова ; А. А. В. — № 2018614727 ; заявл. 17.04.2020 ; опубл. 17.04.2017, 1234567890 (Рос. Федерация).
6. *Antsiferova A., Lavrushkin S., Smirnov M., Gushchin A., Vatolin D., Kulikov D.* Video compression dataset and benchmark of learning-based video-quality metrics [Text] // Advances in Neural Information Processing Systems. Vol. 35. — 2022. — P. 13814—13825.
7. *Antsiferova A., Abud K., Gushchin A., Shumitskaya E., Lavrushkin S., Vatolin D.* Comparing the robustness of modern no-reference image- and video-quality metrics to adversarial attacks [Text] // Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 38. — 2024. — P. 700—708.
8. *Shumitskaya E., Antsiferova A., Vatolin D. S.* Fast Adversarial CNN-based Perturbation Attack on No-Reference Image- and Video-Quality Metrics [Text] // Tiny Papers @ ICLR 2023, Kigali, Rwanda, May 5, 2023. — 2023.
9. *Shumitskaya E., Antsiferova A., Vatolin D. S.* Universal Perturbation Attack on Differentiable No-Reference Image- and Video-Quality Metrics [Text] // 33rd British Machine Vision Conference 2022, BMVC 2022, London, UK, November 21-24, 2022. — 2022.
10. *Antsiferova A., Vatolin D.* The influence of 3D video artifacts on discomfort of 302 viewers [Text] // 2017 International Conference on 3D Immersion (IC3D). — IEEE. 2017. — P. 1—8. — (Scopus).

11. *Zvezdakova (Antsiferova) A., Zvezdakov S., Kulikov D., Vatolin D.* Hacking VMAF with video color and contrast distortion [Text] // Graphicon-Conference on Computer Graphics and Vision. Vol. 2485. — 2019. — P. 53—57. — (Scopus).
12. *Antsiferova A., Yakovenko A., Safonov N., Kulikov D., Gushin A., Vatolin D.* Applying Objective Quality Metrics to Video-Codec Comparisons: Choosing the Best Metric for Subjective Quality Estimation [Text] // Graphicon-Conference on Computer Graphics and Vision. Vol. 31. — 2021. — P. 199—210. — (Scopus).
13. *Zvezdakova (Antsiferova) A., Kulikov D., Kondranin D., Vatolin D.* Barriers towards no-reference metrics application to compressed video quality analysis: on the example of no-reference metric NIQE [Text] // Graphicon-Conference on Computer Graphics and Vision. Vol. 2485. — 2019. — P. 22—27. — (Scopus).
14. *Siniukov M., Antsiferova A., Kulikov D., Vatolin D.* Hacking VMAF and VMAF NEG: vulnerability to different preprocessing methods [Text] // 2021 4th Artificial Intelligence and Cloud Computing Conference. — 2021. — P. 89—96. — (Scopus).
15. *Анциферова А., Ватолин Д.* Прогнозируемая оценка степени дискомфорта зрителей, вызываемого геометрическими, цветовыми и временными искажениями стереоскопического видео, для 60 снятых стереофильмов [Текст] // Запись и воспроизведение объёмных изображений в кинематографе и других областях. IX Международная научно-практическая конференция. (Москва, 17-18 апреля 2018 г.) Материалы и доклады. — 2018. — С. 73—89.
16. *Анциферова А., Ватолин Д., Звездаков С.* Исследование влияния геометрических, цветовых и временных искажений стереоскопических видео на дискомфорт зрителей [Текст] // Запись и воспроизведение объёмных изображений в кинематографе и других областях. IX Международная научно-практическая конференция. (Москва, 17-18 апреля 2017 г.) Материалы и доклады. — 2017. — С. 88.
17. *Анциферова А., Ватолин Д.* Автоматический метод оценки степени усталости от просмотра 3D-видео [Текст] // Запись и воспроизведение объёмных изображений в кинематографе и других областях. IX Международная научно-практическая конференция. (Москва, 17-18 апреля 2016 г.) Материалы и доклады. — 2016. — С. 106—122.

*Анциферова Анастасия Всеволодовна*

Методы проектирования и тестирования алгоритмов оценки качества обработки и кодирования видеоданных

Автореф. дис. на соискание ученой степени канд. физ.-мат. наук

Подписано в печать \_\_\_\_\_.\_\_\_\_\_.\_\_\_\_\_. Заказ № \_\_\_\_\_

Формат 60×90/16. Усл. печ. л. 1. Тираж 100 экз.

Типография \_\_\_\_\_