

Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный университет имени М. В. Ломоносова»



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»

Диссертация на соискание учёной степени
кандидата физико-математических наук

Научный руководитель:
кандидат физико-математических наук
Ватолин Дмитрий Сергеевич

Москва — 2024

Оглавление

	Стр.
Введение	5
Глава 1. Обзор существующих методов тестирования и проектирования алгоритмов оценки качества видеоданных.	14
1.1 Особенности терминологии	14
1.2 Субъективная оценка качества видео	15
1.3 Объективная оценка качества видео	16
1.4 Проблемы существующих методов оценки качества видео	18
1.5 Выводы	23
Глава 2. Методика сравнения алгоритмов оценки качества видео	24
2.1 Описание и постановка задачи	24
2.2 Обзор существующих сравнений алгоритмов оценки качества видео	25
2.3 Описание предложенной методики	26
2.3.1 Создание набора тестовых видео	26
2.3.2 Получение эталонных оценок визуального качества для видео	29
2.3.3 Сравнение алгоритмов оценки качества видео	31
2.3.4 Экспериментальная оценка	32
2.3.5 Результаты сравнения	35
2.4 Открытое онлайн сравнение	37
2.5 Выводы	38
Глава 3. Методы состязательных атак на алгоритмы оценки качества видео	39
3.1 Описание и постановка задачи	39
3.2 Обзор существующих методов	40
3.3 Описание предложенных методов	44
3.3.1 Метод состязательной атаки на неререференсные алгоритмы оценки качества видео	44
3.3.2 Методы состязательных атак на полнореференсные алгоритмы оценки качества видео	46

	Стр.	
3.3.3	Метод тестирования устойчивости алгоритмов оценки качества видео к атакам	50
3.4	Экспериментальная оценка	52
3.4.1	Оценка методов состязательных атак на нереференсные алгоритмы оценки качества на основе построения универсальных возмущений	52
3.4.2	Оценка методов состязательных атак на полнореференсные алгоритмы оценки качества видео	57
3.5	Открытое онлайн сравнение	62
3.6	Выводы	64
Глава 4.	Метод оценки качества стереоскопического видео	65
4.1	Постановка задачи	65
4.2	Обзор существующих методов	66
4.3	Разработка метода оценки качества стереоскопических видео	68
4.3.1	Выбор признаков	68
4.3.2	Исключение аномалий в данных	70
4.3.3	Описание предложенного метода	71
4.3.4	Формулы построенных моделей	73
4.4	Экспериментальная оценка	73
4.4.1	Подготовка набора данных	73
4.5	Программная реализация	84
4.6	Выводы	85
Заключение	87
Список литературы	88
Список рисунков	101
Список таблиц	105
Приложение А. Приложение	107

A.1	Глава 1: зависимость прогнозируемого уровня дискомфорта зрителей при просмотре стереофильма от бюджета его производства и года выхода в прокат.	107
A.2	Глава 1: визуализации выбора признаков для обучения метода прогнозирования дискомфорта.	107
A.3	Глава 1: исследование влияния коэффициента регуляризации на точность предложенной модели.	107
A.4	Глава 2: результаты сравнения методов оценки качества видео с эталонными оценками, полученными с помощью предложенной методики.	108

Введение

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 видекодеками 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 чув-

ствительны 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, количе-

ство 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.

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 для тестирования 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 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 нереперенсные алгоритмы оценки качества изображений и видео в режиме «белого ящика» с помощью создания универсального возмущения и аналитическая оценка его схожести.
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:

1. Предложена новая методика сравнения алгоритмов оценки качества видео, включающая подготовку и разметку видео, и получена вероятностная гарантия применимости модели Брэдли-Терри для оценки качества видео.
2. Разработаны новые методы состязательных атак на алгоритмы оценки качества видео: метод состязательной атаки с помощью обучения универсального возмущения, метод состязательной атаки с помощью оптимизации обработки видео генетическим алгоритмом, метод состязательной атаки путем создания аппроксимации недифференцируемого алгоритма оценки качества видео.

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

видеокодеков, 2486 видео, более 766000 субъективных оценок). Набор данных опубликован на сайте <https://videoprocessing.ai/datasets/cvqad.html>.

2. Проведено самое большое в мире сравнение методов оценки качества видео по количеству проанализированных алгоритмов (41 метод). 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 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 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, превзошедший аналоги по количеству полученных субъективных оценок (302 зрителя и более 22000 субъективных оценок). Набор данных опубликован на сайте <https://videoprocessing.ai/datasets/mvdsvd.html>.

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

Достоверность Все предложенные алгоритмы и разработанная методика были реализованы и прошли экспериментальную проверку.

Апробация работы. Основные результаты работы докладывались на:

1. International Conference on 3D Immersion (г. Брюссель, Бельгия, 11-12 декабря 2017 года).
2. Международная научная конференция студентов, аспирантов и молодых учёных «Ломоносов-2018» (МГУ имени М.В. Ломоносова, г. Москва, Россия, 10-11 апреля 2018).

3. X Международная научно-практическая конференция «Запись и воспроизведение объемных изображений в кинематографе и других областях», (ВГИК, г. Москва, Россия, 16-18 апреля 2018).
4. 29-я Международная конференция по компьютерной графике и машинному зрению ГрафиКон-2019 (г. Брянск, Россия, 23-26 сентября 2019).
5. XXVIII Международная научная конференция студентов, аспирантов и молодых ученых «Ломоносов 2021» (МГУ им. М.В. Ломоносова, Москва, Россия, 12-23 апреля 2021).
6. GraphiCon2021 (31st International Conference on Computer Graphics and Vision) (Нижний Новгород, Россия, 24-30 сентября 2021).
7. 2021 Asia Digital Image Processing Conference (г. Киото, Япония (удаленно), 17-19 декабря 2021).
8. Открытая конференция ИСП РАН (Москва, Россия, 2-3 декабря 2021).
9. OpenTalks.ai 2022 (г. Москва, Россия, 17 февраля 2022).
10. Семинар кафедры интеллектуальных информационных технологий ВМК МГУ (26 мая 2022).
11. Семинар центра визуализации и спутниковых информационных технологий НИИСИ РАН (3 октября 2022).
12. Семинар института информационных технологий, математики и механики ННГУ им. Н.И.Лобачевского (10 ноября 2022).
13. Семинар им. М.Р. Шура-Бура ИПМ им. М.В. Келдыша (3 ноября 2022).
14. Семинар кафедры информатики и программного обеспечения БГТУ (17 марта 2023).
15. Семинар подразделения интеллектуального анализа данных и технического зрения ГосНИИАС (13 апреля 2023).
16. Открытая конференция ИСП РАН (Москва, Россия, 4-5 декабря 2023).
17. Семинар по системному программированию под руководством академика А.И. Аветисяна (27 марта 2024).

Личный вклад. Автору диссертации принадлежит выполнение основного объема теоретических и экспериментальных исследований для разработки нового метода оценки качества стереофильмов в работах ([1],[2], [3], [4]). В работах ([5], [6], [7], [8], [9]) автору принадлежит разработка методики оценки качества видео. Автору также принадлежит постановка задачи, выполнение теоретических исследований и курирование создания программной реализации предложенных методов состязательных атаки на алгоритмы оценки качества видео в работах

([10], [11], [12], [13], [14], [15], [16]). Автором лично была реализована программная система для оценки качества стереоскопических видео [17].

Публикации. Основные результаты по теме диссертации изложены в трех работах, две из которых изданы в журналах, рекомендованных ВАК, одна в журнале, входящем в международные базы цитирования Scopus и Web of Science. Также получено одно свидетельство о государственной регистрации программы для ЭВМ.

Объем и структура работы. Диссертация состоит из введения, 4 глав, заключения и 1 приложения. Полный объем диссертации составляет 115 страниц, включая 38 рисунков и 15 таблиц. Список литературы содержит 135 наименований.

Данная работа посвящена разработке методов тестирования и проектирования алгоритмов оценки качества обработки и кодирования видеоданных. Во Введении обсуждается актуальность работы, цели и задачи исследования, научная новизна, научная и практическая значимость диссертации, формулируются положения, выносимые на защиту, а также приводится список работ, в которых опубликованы результаты данного исследования. Кратко представлено содержание диссертации.

В главе 1 приведен обзор существующих подходов для оценки качества видеоданных. Они включают в себя методы субъективного тестирования и объективные методы, реализованные в виде алгоритмов. Приведено описание проблем существующих методов, которые заключаются в трудности тестирования их точности для новых данных, трудности оценки устойчивости нейросетевых методов к атакам на оценку качества. Внедрение таких атак в алгоритмы обработки видео приводит к появлению нескольких негативных последствий, в том числе позволяет их разработчикам получать более высокие позиции в рейтингах сравнений. Уязвимость методов оценки качества видео к атакам может привести к неверной интерпретации результатов сравнений алгоритмов обработки и кодирования видео. Также на примере оценки качества стереоскопических видео обсуждаются трудности разработки новых методов, предназначенных для узкоспециализированных задач, имеющих малое количество размеченных данных. Из-за отсутствия автоматических методов оценки качества стереоскопического видео, задача устранения стереоскопических артефактов, вызывающих головную боль и дискомфорт, остается актуальной. В отличие от обычного видео, где искажения кадра, вызванные сильным кодированием или состязательными атаками,

лишь снижают визуальное качество, артефакты стереоскопического видео могут вызвать множество неприятных симптомов, например визуальный дискомфорт.

Глава 2 посвящена разработке методики сравнения современных алгоритмов оценки качества изображений и видео в рамках оценки визуального качества кодированных видео. Задача сравнения существующих методов оценки качества видео была поставлена во многих работах, посвященных разработке новых алгоритмов оценки качества, описанных в разделе 2.2. В разделе 2.3 предложена новая методика сравнения алгоритмов оценки качества видео. Новый набор данных, описанный в 2.3, состоит из 2435 видео, закодированных 83 видеокодеками, содержит субъективные оценки, полученные от 10800 респондентов. Как показано в 2.2, предложенный набор видео превосходит аналоги по объему и репрезентативности искажений, возникающих при кодировании видео. Раздел 2.3 содержит результаты сравнения алгоритмов оценки качества видео, проведенного согласно предложенной методике. Проведенное исследование показало, что для оценки качества кодирования видео некоторые современные методы, которые не требуют эталонного видео для оценки, по точности становятся очень близки к полнореференсным, что позволяет использовать их для решения практических задач в областях кодирования и потоковой передачи видео.

Глава 3 посвящена разработке новых методов состязательных атак на алгоритмы оценки качества видео. Методы оценки качества видео, основанные на машинном обучении, могут быть подвержены возмущениям во входных данных, в том числе состязательным атакам. Существующие исследования в области состязательных атак проводились в основном для оценки устойчивости алгоритмов компьютерного зрения, таких как классификация и обнаружение объектов, и существует лишь несколько исследований устойчивости алгоритмов оценки качества видео к атакам, которые описаны в разделе 3.2. Предложенные подходы для состязательных атак на методы оценки качества видео, описанные в разделе 3.3, основаны на нескольких подходах: метод состязательной атаки в режиме белого ящика с обучением универсального возмущения, добавляемого к входному изображению, а также методы состязательных атак в режиме черного ящика с использованием генетических алгоритмов и обучения аппроксимации недифференцируемого метода оценки качества видео. Раздел 3.4 посвящен экспериментальной оценке разработанных методов состязательных атак применительно к современным алгоритмам оценки качества видео. Практическая значимость разработанных методов подтверждается повышенным интересом к их

результатам со стороны индустрии и внедрением одного из опубликованных методов в исходный код открытого видеокодека libaom, разрабатываемого Alliance for Open Media, в который входят компании Google, NETFLIX, Mozilla, и другие.

В Главе 4 рассматривается задача создания метода оценки качества стереоскопического видео. В разделе 4.2 рассмотрены существующие исследования визуального восприятия стереоскопических видео. В основном существующие методы в данной области посвящены оценке визуального качества, а не визуального дискомфорта. Также существующие исследования в основном посвящены анализу характеристик стереоскопических видео, таких как величина диспаратности, временная и пространственная сложность и другие, но не рассматривается влияние на визуальное качество и дискомфорт артефактов, возникающих при стереосъемке из-за недостаточной калибровки и синхронизации камер: это геометрические искажения (вертикальный сдвиг и поворот), цветовое искажение и временной сдвиг между ракурсами стереопары. Предложенный метод оценки качества описан в разделе 4.3. Он основан на обучении модели линейной регрессии с устойчивой к выбросам функцией потерь. Для обучения использовался предложенный набор данных, полученный в результате проведения серии субъективных экспериментов. Разработанный метод был опробован для оценки качества 60 стереоскопических фильмов, результаты экспериментальной оценки описаны в разделе 4.4.

В Заключение сформулированы основные результаты диссертационного исследования, рассмотрены возможные варианты их применения и обозначены перспективы дальнейших исследований.

Глава 1. Обзор существующих методов тестирования и проектирования алгоритмов оценки качества видеоданных.

Оценка качества является одной из классических задач в обработке сигналов. Видеоданные обладают высокой пространственно-временной избыточностью, поэтому для сокращения размера хранимых и передаваемых данных для уменьшения избыточности используются специальные программы — видеокодеки. Они осуществляют кодирование и декодирование видеоданных с потерями, что позволяет уменьшить размер файлов. Однако, квантизация частот при кодировании сигналов приводит к потере информации и качества сигнала. Существует два основных способа измерения качества сигналов: субъективная оценка, получаемая с помощью привлечения экспертов и их опроса, и объективная, получаемая с помощью алгоритмов.

1.1 Особенности терминологии

В данной работе использованы следующие понятия:

- Визуальное качество — критерий качества видео, определяемый в ходе субъективного тестирования с привлечением экспертов.
- Визуальный дискомфорт — критерий качества стереоскопического видео, определяемый в ходе субъективного тестирования, и оценивающий негативные симптомы у зрителя.
- Состязательная атака — возмущение, которое при добавлении к изображению или видео, изменяет его объективную оценку качества.
- Атака в режиме «белого ящика» — подход к разработке методов состязательных атак, при котором атакуемый алгоритм известен и доступен для использования при оптимизации атакующих возмущений.
- Атака в режиме «черного ящика» — подход к разработке методов состязательных атак, при котором атакуемый алгоритм недоступен.
- Артефакт — искажение изображения или видео, приводящее к снижению визуального качества.
- Сцена — набор кадров, связанных во времени и пространстве.

- Стереофильм — кинофильм, изобразительный ряд которого представляет собой последовательность стереокадров, полученных путем съемки.
- Стереопара — два изображения одного и того же объекта, зафиксированные с двух ракурсов или созданные искусственно.
- Ракурс — положение глаза или объектива, используемого для наблюдения или съемки, по отношению к рассматриваемому объекту. Изображения стереопары — это изображения левого и правого ракурсов.
- Диспаратность — различное взаимное положение точек, отображаемых на сетчатках левого и правого глаза. Термин также используется для обозначения величины бинокулярного параллакса, выраженной в угловых или дуговых градусах.

1.2 Субъективная оценка качества видео

Субъективная оценка возможна только с привлечением экспертов для разметки данных, например, для прослушивания аудиозаписей, просмотра изображений или просмотра видео. При проведении визуальных сравнений видео экспертам требуется оценить качество изображения. Для получения достоверной оценки требуется привлечение большого количества экспертов и агрегация их ответов. Проведение субъективной оценки в лаборатории позволяет проконтролировать соблюдение условий просмотра видео, при необходимости проверить остроту зрения экспертов или провести другие тесты. Однако, из-за дороговизны таких экспериментов, в последнее время набирает популярность подход к субъективной оценке, называемый краудсорсинговым. В нем эксперты принимают участие в разметке качества видео удаленно, например, скачивая видео и оставляя оценку через форму на сайте. Такой подход применим только для оценки качества видео, не требующих специального оборудования. Его достоинством является высокая скорость получения эталонной разметки, что позволяет получить больше оценок для каждого видео или другого сигнала. Однако, недостатком краудсорсингового подхода является низкая достоверность и зашумленность полученных оценок, возникающая из-за невозможности контролировать условия просмотра и внимательность эксперта. Несмотря на развитие краудсорсинговых подходов к оценке качества сигналов и снижение стоимости получения оценок, субъектив-

ная оценка используется в основном для валидации результатов работы методов обработки сигналов.

1.3 Объективная оценка качества видео

Алгоритмы объективной оценки качества можно разделить на три категории [18]: полнореференсные, сокращенореференсные и неререференсные. Полнореференсные методы наиболее распространены, поскольку их результаты легко интерпретируются. Как правило, такие методы используются для оценки степени искажений в видео и их заметности для зрителя. Основным недостатком данного подхода по сравнению с сокращенным референсным и неререференсным подходами является требование наличия оригинального видео для сопоставления со сравниваемым, что часто неосуществимо. Наиболее популярные алгоритмы для оценки качества кодирования видео являются полнореференсными: PSNR, SSIM [19], VMAF [20].

Методы оценки качества изображений и видео широко используются при разработке и сравнении алгоритмов обработки изображений и видео. До недавнего времени наиболее широко используемыми являлись полнореференсные методы, так как из-за наличия исходного видео в доступе, они обладают наибольшей точностью. Однако в последнее время неререференсные методы становятся более популярными благодаря более широкой области применимости. Например, неререференсные методы стали часто использоваться для оценки качества потоковой передачи на стороне зрителей, оценки качества видеоконференцсвязи, сбора статистики качества контента для коммуникации в реальном времени и потокового вещания, а также для оптимизации параметров обработки изображений и видео [21]. При этом, новые методы обладают высокой точностью работы и используют подходы на основе нейронных сетей. За счет обучения на больших наборах данных и адаптации к новым типам искажений видео, вызванных обработкой нейросетевыми методами, современные неререференсные алгоритмы оценки качества стали показывать более высокую корреляцию с визуальными оценками, чем традиционные подходы.

Неререференсные методы обычно используют нейросетевые подходы или статистики, оценивающие распределение естественных изображений. Напри-

мер, метод VIDEVAL (2021) [22] использует 60 признаков видео (измерение движения, искажений и визуального качества) из ранее разработанных моделей качества. Этот метод хорошо работает на существующих наборах данных пользовательских видео, но требует дополнительной настройки параметров. В большинстве последних работ по оценке качества особое внимание уделяется подходам, основанным на глубоком обучении. MEON (2017) [23] — это модель, построенная из двух нейронных сетей: сети идентификации искажений и сети прогнозирования качества. Она может определять тип искажения. VSFA (2019) [24] использует предварительно обученную сверточную нейронную сеть, основанную на ResNet-50 [25] и также экстрактор признаков, учитывающий содержание, за которым следует слой временной памяти (temporal-pooling layer). Точность данного метода была низкой при оценке с использованием различных открытых наборов данных, поэтому авторы предложили MDTVSAFA (2021) [26] как улучшенную версию VSFA. Он использует стратегию обучения на смешанных наборах данных и имеет высокую вычислительную сложность из-за использования рекуррентных слоев и полноразмерных входных изображений. PaQ-2-PiQ (2020) [27] использует архитектуру, анализирующую области изображения, обученную на большом наборе изображений с эталонными оценками из 40000 фотографий. Koncept512 (2020) [28] основан на сети InceptionResNetV2 и был обучен на предложенном авторами наборе данных KonIQ-10k. SPAQ (2020) [29] содержит три дополнительные модификации своей базовой модели: обработка данных EXIF (MT-E), учет атрибутов изображения (MT-A) и учет высокоуровневой семантики сцены (MT-S). Авторы LINEARITY (2020) [30] ввели собственную функцию потерь “norm-in-norm”, которая сходится в 10 раз быстрее, чем функции потерь MAE и MSE. NIMA (2018) [31] была обучена на крупномасштабном наборе данных Aesthetic Visual Analysis (AVA) и прогнозирует распределение оценок качества.

Полнореференсные методы оценки качества видео PSNR и SSIM [19] являются одними из самых популярных методов оценки качества изображений и видео. Существуют различные вариации суммирования цветовых компонент SSIM и MS-SSIM [32]; последняя является усовершенствованной версией SSIM, вычисленная по нескольким разрешениям с использованием субдискретизации. LPIPS (2019) [33] основан на AlexNet и VGG. DISTTS (2020) [34] был разработан для учета дискретизации текстур и чувствителен к структурным различиям. Он объединяет измерения структурного и текстурного сходства фрагментов изоб-

ражений и основан на предварительно обученной сети VGG. Tencent DVQA (2020) [35] основан на сети C3DVQA [36]. Эта архитектура использует 3D-конволюционные слои для учета пространственно-временных особенностей и 2D-конволюционные слои для извлечения пространственной информации. Главной особенностью FovVideoVDP (2021) [37] является учет периферийной остроты зрения. Этот метод моделирует реакцию зрительной системы человека на временные изменения в поле зрения. Он может оценить мерцание, дрожание и другие временные искажения, а также пространственно-временные артефакты, возникающие при различной степени периферийного зрения. ST-GREED (2021) [38] может количественно оценить эталонное и искаженное видео с различной частотой кадров без предварительной обработки. Он предлагает две основные функции: SGreed и TGreed. Последняя оценивает статистику временных полосовых откликов, учитывающих как пространственные, так и временные искажения. Первая получает пространственные оценки с помощью схемы локальной фильтрации. При вычислении итоговой оценки качества видео ST-Greed использует регрессор, основанный на методе опорных векторов. В настоящее время VMAF (2018) [39] является одним из самых популярных методов оценки качества видео. Он вычисляет три базовых признака: оценку потери деталей (DLM) [40], верность визуальной информации (VIF) [41] и временную информацию (TI), и комбинирует их с помощью регрессии методом опорных векторов.

1.4 Проблемы существующих методов оценки качества видео

В настоящее время разрабатывается большое количество новых методов оценки качества видео. Однако, точность многих методов не была воспроизведена исследователями на новых данных или при независимых сравнениях. В связи с развитием технологий кодирования видео, искажения, вносимые при кодировании, стали иначе оцениваться методами, традиционно использовавшимися для данной задачи. Например, повышение разрешения, используемое в LCEVC и в других новых стандартах кодирования на основе нейронных сетей приводит к искажениям, с которыми традиционные методы не могут справиться. На рисунке 11 показана разница между кадрами видео с кодировкой x265 и видео с кодировкой

lcevc_x265: второе содержит больше деталей при более низких показателях PSNR и SSIM.



Рисунок 11 — Фрагменты кадров из видео, закодированных видекодеком стандарта H.265 и видекодеком нового стандарта LCEVC-H.265. LCEVC использует повышение разрешения, которое позволяет восстановить больше деталей и приводит к появлению искажений нового типа.

Одной из проблем невоспроизводимости результатов новых методов оценки качества видео для решения задачи оценки кодирования связана с устаревшими наборами данных для измерения качества кодирования видео. Большинство существующих наборов данных со сжатым видео и эталонными оценками включают только кодирование H.264/AVC. Эталонные оценки для многих из них были получены с помощью лабораторных тестов. Из-за высокой сложности и стоимости таких сравнений в разметке участвовало небольшое количество зрителей, и на каждое видео приходилось всего несколько оценок. Из-за этого искажения, создаваемые видекодеками, использующими современные стандарты, часто не учитываются при разработке методов оценки качества.

Как и другие нейросетевые подходы, методы оценки качества, основанные на машинном обучении и нейронных сетях, чувствительны к изменению входных данных, в том числе к состязательным атакам. Так, например, традиционный метод оценки качества PSNR более устойчив к модификациям входных изображений, направленным на манипулирование его оценками (так как любое изменение пикселей входного изображения уменьшит оценку PSNR). Напротив, поведение нейросетевых методов гораздо сложнее. На рисунке 12 продемонстрирован пример подобной атаки на метод оценки качества PaQ-2-PiQ: после добавления атакующего возмущения к исходному изображению оценка качества методом повысилась, однако его визуальное качество ухудшилось, так как в изображении появился шум.

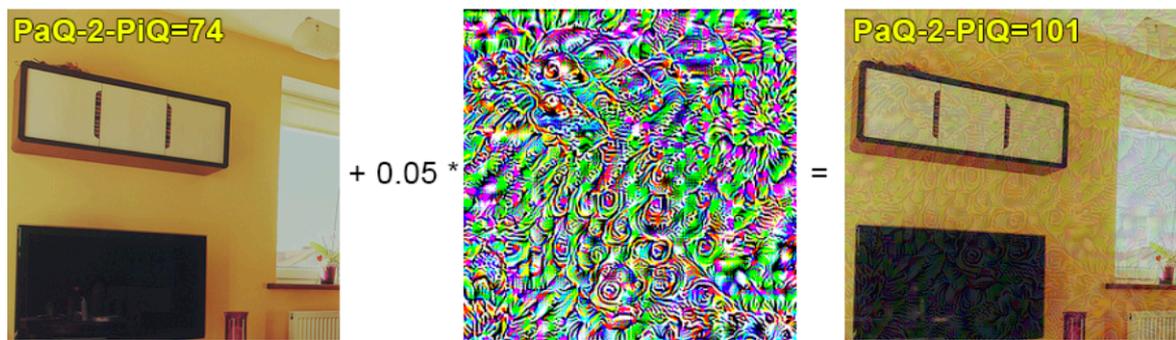


Рисунок 12 — Атака на метод оценки качества изображений PaQ-2-PiQ.

Существует несколько способов оценки надежности методов, основанных на нейронных сетях, к возмущениям входных данных и состязательным атакам. Например, устойчивость к возмущениям можно исследовать с помощью теоретических оценок, таких как оценка липшицевости. Однако подобные подходы имеют множество ограничений, в основном связанные с высокой вычислительной сложностью, поэтому сегодня только нейросети с небольшим количеством параметров или обучающих данных могут быть верифицированы теоретически. Другой подход к оценке устойчивости основан на эмпирическом подходе и применении состязательных атак. Данная область широко исследована для моделей компьютерного зрения, однако для методов оценки качества изображений и видео пока получено мало результатов.

В настоящее время в сравнениях алгоритмов оценки качества изображений и видео в основном оценивается их корреляция с эталонными оценками качества, но не учитывается их устойчивость к возмущениям во входных данных. В то же время возможность манипулировать оценками алгоритмов измерения качества изображений и видео уже встречается в различных жизненных сценариях. Ниже приведены некоторые примеры таких сценариев и потенциальных негативных последствий использования ненадежных методов.

Непреднамеренное снижение визуального качества обработанных изображений и видео. Обработка видео, ориентированная на повышение оценок методов измерения качества видео, уже реализована во многих видеокодеках. Например, видеокодеки libaom [42] и LCEVC [43] имеют режимы, оптимизирующие выходной видеопоток для повышения оценки качества метода VMAF. Такая настройка была разработана для улучшения визуального качества закодированного видео; однако, поскольку VMAF является методом, основанным на машинном обучении, такая обработка видео может снижать визуальное качество [14; 15]. Использование при оптимизации функции потерь неустойчивых к возмущени-

ям методов оценки качества изображений может привести к снижению качества восстановления видео [44]. Например, LPIPS [33] широко используется в качестве показателя для оценки визуального восприятия, но оптимизация его оценок приводит к увеличению яркости [45], что может быть нежелательно или даже привести к ошибочным результатам (например, при непреднамеренном повышении яркости медицинских изображений).

Манипулирование результатами открытых сравнений. 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. Например, лидеры по VMAF в визуальном сравнении видеокодеков МГУ 2021 отличаются от лидеров по визуальному качеству [46]. Метод LPIPS по-прежнему используется в качестве основного показателя в некоторых тестах, например для сравнения методов сверхвысокого разрешения [47]. При этом для обоих методов VMAF и LPIPS уже были опубликованы примеры их уязвимостей к преобразованию входного видео.

Манипулирование результатами поиска изображений в интернете. 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 [48]. Как было показано во время конкурса MediaEval 2020 «Pixel Privacy: Quality Camouflage for Social Images» [49], существует множество способов обмануть методы оценки качества изображений.

Разработка новых методов оценки качества видео требует получения большого количества размеченных данных в ходе субъективных лабораторных или краудсорсинговых экспериментов. Стоимость проведения субъективной оценки обычно высока, и может усложняться необходимостью использования специализированного оборудования для демонстрации сигнала. Например, оценка качества стереоскопического видео требует использования двух проекторов и 3D-очков или 3D-телевизора.

При этом, оценка качества стереоскопического видео является актуальной задачей. Съемка стереофильмов производится двумя камерами, закрепленными на стереориге или с помощью бимсплиттера — специального устройства с зерка-

лом, позволяющим записывать два ракурса видео одновременно. Процесс калибровки камер требует синхронизации большого количества параметров, например, фокусного расстояния, масштаба, цветовых характеристик. Из-за сложности их настройки часто камеры остаются не полностью синхронизированы, что приводит к возникновению, например, геометрических искажений между правым и левым ракурсами стереокадра. Получившиеся неестественные для зрительной системы сцены стереофильма могут вызывать дискомфорт или даже головную боль при просмотре [50]. У дискомфорта, возникающего во время просмотра стереоскопического видео, есть много причин. Они включают в себя низкое качество контента (например, стереоскопические артефакты, неестественные искажения, вызванные конвертацией из 2D-видео в стереовидео и др.), а также несоблюдение рекомендаций для комфортного просмотра стереоскопического контента (некачественное оборудование, нахождение вне допустимой комфортной зоны во время просмотра).

Сегодня оценка качества стереофильмов проводится путем субъективных экспериментов с привлечением экспертов. Понятие визуального качества стереоскопического видео тесно связано с визуальным дискомфортом. Стереоскопическое видео может иметь низкое визуальное качество, например, в нем может присутствовать размытие, неестественные цвета и другие артефакты, но при этом просмотр такого видео не будет приводить к появлению визуального дискомфорта. Однако, визуально нормальное видео может вызывать появление дискомфорта у зрителя при его просмотре. Визуальный дискомфорт является одним из субъективных критериев оценки качества стереоскопического видео. Для сокращения количества стереоскопических искажений в выходящих в прокат стереофильмах требуется проведение контроля качества стереоскопического контента на этапе его производства. В настоящее время существуют алгоритмы, способные обнаружить различные искажения в стереопаре [51], однако величина искажения, определяемая алгоритмом (например, величина относительного поворота в градусах), не всегда отражает то, насколько болезненным оно может быть для зрителя. Различные типы искажений, а также их интенсивность, по-разному влияют на уровень дискомфорта [50]. Поэтому в случае необходимости исправления большого количества сцен в короткие сроки важно в первую очередь найти и исправить те из них, которые потенциально могут вызвать наиболее сильный дискомфорт при просмотре.

1.5 Выводы

В первой главе проведен обзор основных подходов к оценке качества видеоданных. Описаны трудности, связанные с получением эталонных оценок с помощью субъективного метода. Также рассмотрены объективные методы, которые позволяют получить оценку качества с помощью алгоритмов. Рассмотрены проблемы существующих методов объективной оценки качества, которые заключаются в сложности их тестирования и разработки. Сложность тестирования связана с двумя факторами. Во-первых, существующие методы оценки качества обладают низкой точностью при оценке видео, обработанных нейросетевыми моделями обработки и кодирования видео. Во-вторых, методы оценки качества видео, уязвимы к состязательным атакам, что затрудняет их использование при разработке новых алгоритмов обработки и кодирования видео. Сложность проектирования новых алгоритмов заключается в нехватке размеченных данных и высокой стоимостью их получения, особенно для задач, требующих специального оборудования. Данная проблема рассмотрена на примере оценки качества стереоскопического видео. Для разработки точных и надежных методов оценки качества видео требуется решение обозначенных в данной главе проблем.

Глава 2. Методика сравнения алгоритмов оценки качества видео

Целью исследования, описанного в данной главе, является разработка методики и независимый анализ современных методов оценки качества изображений и видео с использованием нового набора данных, представляющих различные артефакты кодирования от различных видеокодеков, в том числе новых стандартов. Под методикой сравнения алгоритмов оценки качества кодированных видео подразумевается совокупность программных средств и алгоритмов для проведения сравнения методов оценки качества закодированных видео.

2.1 Описание и постановка задачи

Задача формулируется следующим образом: по входному набору тестовых видео $\{X_j\}_{j=1}^N$, где $X_j = \{x_i\}_{i=1}^P$ — последовательность кадров $x_i \in \mathbb{R}^{3 \times W \times H}$ и результатам оценки качества каждого видео K сравниваемыми методами $M_k(X_j)$, $k = 1..K, j = 1..N$ необходимо получить упорядоченную последовательность $\{r_i\}_{i=1}^K$ методов оценки качества M_k , такую что для любого $k_1, k_2 \in 1..K, k_1 < k_2$ метод M_{k_1} лучше соответствует визуальному качеству, чем M_{k_2} .

Требуется разработать методику для ранжирования методов оценки качества видео по соответствию их оценок оценкам визуального качества закодированных видео. Также требуется провести экспериментальную оценку предложенной методики, которая требует решения следующих задач:

- Подготовить набор данных для тестирования алгоритмов оценки качества видео, состоящий из исходных видеопоследовательностей и видеопоследовательностей, закодированных различными видеокодеками с разными уровнями качества
- Получить эталонные оценки визуального качества для подготовленных видео от зрителей.
- Провести сравнение современных методов оценки качества видео на новом наборе данных и определить методы, наилучшим образом отражающие эталонные оценки.

2.2 Обзор существующих сравнений алгоритмов оценки качества видео

Большинство существующих сравнений методов оценки качества изображений и видео опубликованы в работах, авторы которых сами в них предложили новый метод и сравнили его с существующими. Подобные сравнения часто основаны на открытых наборах видео, на которых существующие методы могли быть обучены. Также в таких сравнениях часто содержится небольшое количество методов. В [52] было опубликовано сравнение на нескольких различных наборах данных, но оценивались только 4 метода. Существует только два сравнения методов оценки качества видео, которые принимают новые методы. В [53] авторы сравнили методы оценки качества видео, используя три набора данных пользовательского видеоконтента (UGC) и провели ряд исследований. Они проанализировали точность работы методов используя видео с различными типами контента, разрешением и качеством, временной сложностью, а также сравнили их вычислительную сложность. Артефакты кодирования, однако, играли незначительную роль в этом исследовании. Основная идея работы [54] заключалась в сравнении полнореференсных и неререференсных методов посредством субъективной оценки пользовательского видеоконтента, закодированного с использованием различных стандартов и степени кодирования видео.

Существующие открытые наборы данных, содержащие видео и эталонные оценки для них, можно разделить на два типа: с синтетическими искажениями (в основном за счет искажений при кодировании и передаче, ухудшения качества захвата, артефактов обработки и гауссова размытия) и пользовательский контент (UGC). В наборах данных [33; 55–61] оригинальные видео были модифицированы синтетически. Наборы данных UGC [52; 62–66] набирают популярность, поскольку в настоящее время видео, созданные непрофессиональными создателями контента, содержат широкий спектр нестандартных искажений. Многие новые методы оценки качества видео были протестированы только на видео пользовательского контента. В большинстве последних опубликованных исследований методов оценки качества видео используются похожие наборы данных, которые показаны в таблице 1.

Набор	Кол-во видео	Средн. длит. (с)	Кол-во иск. видео	Искажение	Метод эталон. оценок	Зрители	Ответы
MCL-JCV (2016) [55]	30	5	1560	Кодирование	In-lab	150	78 тыс.
VideoSet (2017) [67]	220	5	45760	Кодирование	In-lab	800	-
UGC-VIDEO (2020) [68]	50	> 10	550	Кодирование	In-lab	30	16,5 тыс.
CVD-2014 (2014) [63]	5	10-25	234	In-capture	In-lab	210	-
LIVE-Qualcomm (2016) [62]	54	15	208	In-capture	In-lab	39	8.1 тыс.
GamingVideoSET (2018) [69]	24	30	576	Кодирование	In-lab	25	-
KUGVD (2019) [70]	6	30	144	Кодирование	In-lab	17	-
KoNViD-1k (2017) [64]	1200	8	1200	In-the-wild	Crowdsourcing	642	205 тыс.
LIVE-VQC (2018) [52]	585	10	585	In-the-wild	Crowdsourcing	4776	205 тыс.
YouTube-UGC (2019) [65]	1500	20	1500	In-the-wild	Crowdsourcing	>8000	600 тыс.
LSVQ (2020) [66]	39075	5-12	39,075	In-the-wild	Crowdsourcing	6284	5 млн.
Предложенный: открытая часть	36	10, 15	1022	Кодирование (32 кодека)	Crowdsourcing	10,800	320 тыс.
Предложенный: скрытая часть	36	10, 15	1464	Кодирование (51 кодека)	Crowdsourcing	10,800	446 тыс.
Предложенный	36	10, 15	2486	Кодирование (83 кодека)	Crowdsourcing	10800	766К

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

2.3 Описание предложенной методики

Предложенная методика сравнения алгоритмов оценки качества закодированных видео состоит из трех этапов:

1. Создание набора тестовых видео.
2. Получение эталонных оценок качества тестовых видео.
3. Сравнение методов на полученном наборе данных.

На рисунке 21 представлена схема предложенной методики.

2.3.1 Создание набора тестовых видео

Для анализа работы алгоритмов оценки качества закодированных видео необходима подготовка специального набора видео, содержащего различные искажения, возникающие при кодировании. Для оценки визуального качества кодирования видео важно, чтобы оригинальное видео имело высокий битрейт.

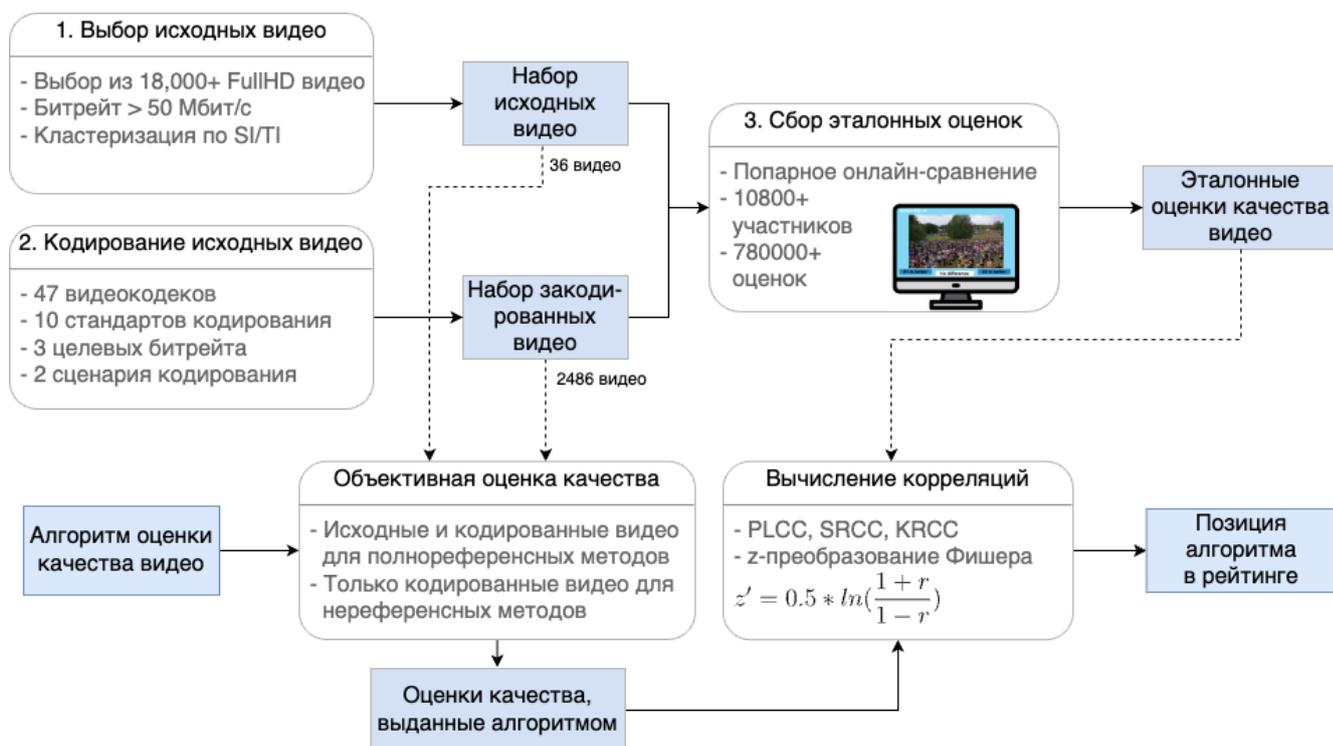


Рисунок 21 — Схема предложенной методики тестирования алгоритмов оценки качества видео.

Это необходимо для исключения появления артефактов, возникающих при перекодировании. Артефакты перекодирования могут присутствовать в реальных видео, передаваемых по сети, но чаще пользователи смотрят видеоконтент, подготовленный для трансляции и закодированный единожды, так как однократное кодирование позволяет достичь оптимального соотношения размера файла и минимизировать искажения. Для минимизации искажений видео предлагается использовать в качестве референсных видео только видео разрешения не менее FullHD с битрейтом более 20 Мбит/с и открытой лицензией.

Набор исходных видео должен обладать высокой репрезентативностью по пространственно-временной сложности. Список жарнов фиксируется пользователем и жанр каждого видео определяется субъективно. Он отражает содержание видео, например, “спорт”, “интервью”, “природа” и т.д. В данной работе предлагается оценивать пространственную и временную сложность с использованием алгоритмов, описанных в [71]. Пространственная сложность представляет собой нормализованный средний размер I-кадра в байтах, а временная сложность представляет собой средний размер P-кадра, деленный на средний размер I-кадра. Как показано на 22, распределение пространственно-временной сложности реальных пользовательских видео отличается от сложности видео, представленных в по-

пулярных открытых наборах media.xiph.org, который используется для сравнения стандартов кодирования видео.

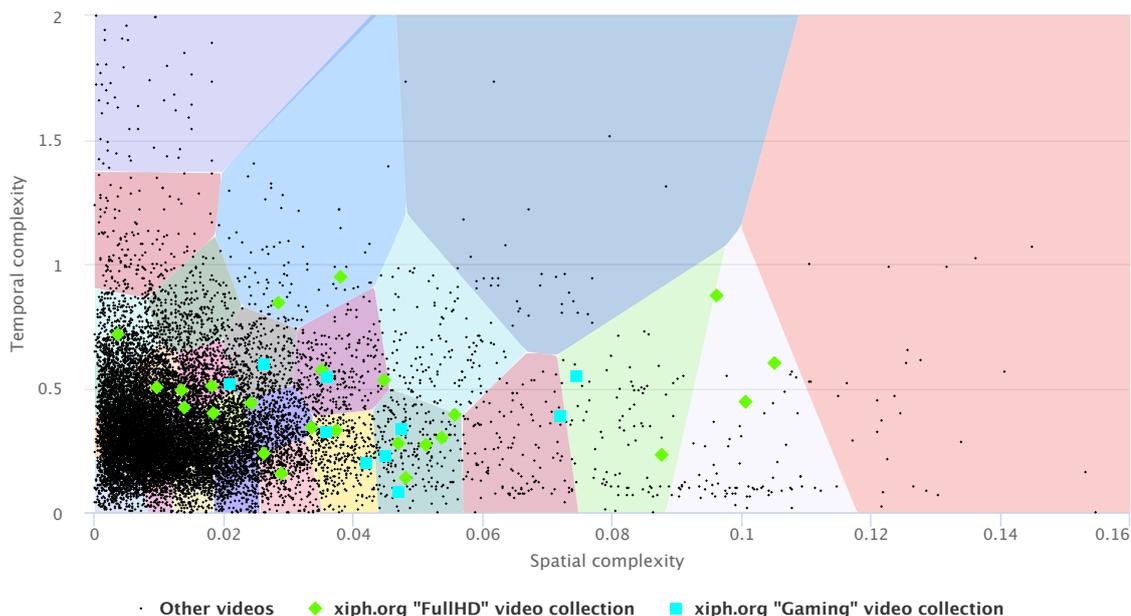


Рисунок 22 — Сравнение пространственной и временной сложности 18 418 пользовательских видео с ресурса vimeo.com с популярными открытыми коллекциями видео media.xiph.org.

Для репрезентативности набора видео предлагается использовать кластеризацию по пространственно-временной сложности с помощью алгоритма K-Means [72]. Количество кластеров соответствует необходимому количеству исходных видео. Для каждого кластера случайным образом отбираются до 10 видеороликов-кандидатов, которые находятся близко к центру кластера. Из кандидатов каждого кластера выбирается вручную одно видео с учетом его жанра так, чтобы в итоговом наборе присутствовали видео разных жанров.

На следующем этапе необходимо провести моделирование широкого набора типов искажений, возникающих при кодировании. Набор исходных видео кодируется с помощью нескольких видеокодеков различных стандартов. Список стандартов 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: одна обеспечивает скорость кодирования 30 кадров в секунду, а другая — 1 кадр в секунду, что приводит к более высокому качеству. Видео сжимаются с тремя целевыми битрейтами: 1000 кбит/с, 2000 кбит/с и 4000 кбит/с с использованием режимов VBR (для видеокодеков,

поддерживающих режим VBR) или с соответствующими значениями QP/CRF, обеспечивающими такие битрейты. Существуют рекомендации сервисов потокового видео относительно использования целевых битрейтов для разрешения FullHD. В основном рекомендуется кодировать видео с битрейтом не более 4500–8000 кбит/с [73—75].

2.3.2 Получение эталонных оценок визуального качества для видео

Получение эталонных оценок для созданного набора видео предлагается с помощью проведения попарных сравнений онлайн. Онлайн-сравнения называют также краудсорсинговым подходом. По сравнению с традиционным подходом к оценке видео зрителями в лабораторных условиях, краудсорсинговый подход позволяет получить большее количество данных и снизить влияние факторов среды на результат за счет широкого охвата аудитории. Разметка качества видео в режиме попарных 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 перед разметкой [76]. В данной работе предлагается для разметки использовать платформу Subjectify.us — это сервис для попарных сравнений, который использует модель Брэдли-Терри для преобразования результатов парного голосования в эталонную оценку качества каждого видео. Ниже приведено описание алгоритма получения эталонных оценок для видео.

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. Формально голосование и

его результаты можно представить в виде квадратной таблицы, строки и столбцы которой соответствуют сравниваемым видео, а в каждой ячейке содержится значение, обозначающее количество раз, когда зрители проголосовали за видео v_{ij} из строки против видео v_{ji} из столбца. Если видео были отмечены как имеющие одинаковое качество, то в обе ячейки v_{ij} и v_{ji} прибавляется единица.

Для преобразования результатов голосования в конечные оценки качества видео применяется алгоритм Брэдли-Терри. Обозначим (i, j) каждую пару видео c_i, c_j , где c_i лучше, чем c_j . Обозначим $C_{i,j}$ — количество сравнений, в которых видео i было лучше, чем видео j . 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:

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

$$\text{Тогда } p_{ij} = P(i > j) = \frac{e^{\beta_i}}{e^{\beta_i} + e^{\beta_j}}.$$

Выразим логарифмическую функцию правдоподобия:

$$L(p) = \ln \prod_{i,j} \left(\frac{e^{\beta_i}}{e^{\beta_i} + e^{\beta_j}} \right)^{C_{i,j}} = \sum_{i,j} \ln \left(\frac{e^{\beta_i}}{e^{\beta_i} + e^{\beta_j}} \right)^{C_{i,j}} = \sum_i \sum_j C_{i,j} \ln \left(\frac{e^{\beta_i}}{e^{\beta_i} + e^{\beta_j}} \right).$$

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

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

Так как метод максимизации правдоподобия дает асимптотически нормальную оценку, стандартная ошибка для $\beta_i - \beta_j$ может быть выражена как:

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

α -доверительный интервал для $\beta_i - \beta_j$ выражается как $(\hat{\beta}_i - \hat{\beta}_j) \pm \hat{se}_{\hat{\beta}_i - \hat{\beta}_j} * z_{\frac{1-\alpha}{2}}$. Рассмотрим случай, когда случайная величина $\beta_i - \beta_j$ принимает всего два значения: 0, если видео i лучше, чем видео j , и 1 иначе. Тогда верно следующее утверждение.

Теорема 1. Пусть для всех возможных упорядоченных пар $(i, j), i \geq j$ видео выполняется $\min_{i,j} \delta_{ij} > 0$, где $\delta_{ij} = \operatorname{sign}(\hat{\beta}_i - \hat{\beta}_j) * (\hat{\beta}_i - \hat{\beta}_j) - \hat{se}_{\hat{\beta}_i - \hat{\beta}_j} * z_{\frac{1-\alpha}{2}}$. Тогда

можно установить отношение порядка между всеми видео из набора в соответствии с эталонными оценками β_1, \dots, β_N с вероятностью хотя бы $p = (1 - \alpha)^{C_N^2}$.

Доказательство. Рассмотрим пару видео (i, j) с эталонными оценками β_i и β_j . Зафиксируем уровень значимости α . Для установления порядка при ранжировании этой пары с вероятностью хотя бы $1 - \alpha$ необходимо, чтобы доверительный интервал статистики $\hat{\theta}_{ij} = \hat{\beta}_i - \hat{\beta}_j$ целиком располагался в области одного знака. Для этой статистики доверительный интервал имеет вид $(l_{ij}, r_{ij}) = (\hat{\theta}_{ij} - \hat{s}e_{\hat{\beta}_i - \hat{\beta}_j} * z_{\frac{1-\alpha}{2}}, \hat{\theta}_{ij} + \hat{s}e_{\hat{\beta}_i - \hat{\beta}_j} * z_{\frac{1-\alpha}{2}})$. Если границы интервала принимают значения одного знака, то с вероятностью хотя бы $1 - \alpha$ значение θ_{ij} имеет тот же знак. Тогда оценка порядка для пары видео (i, j) будет соответствовать эталонной с вероятностью хотя бы $1 - \alpha$. Теперь рассмотрим все возможные C_N^2 пар (i, j) . Введем случайную величину ξ :

$$\xi = \begin{cases} 1, & \text{если для всех упорядоченных пар } (i, j) \theta_{ij} \in (l_{ij}, r_{ij}), \\ 0, & \text{иначе.} \end{cases} \quad (2.1)$$

Заметим, что

$$\mathbb{P}(\xi = 0) \leq \sum_{i \geq j} \mathbb{P}(\theta_{ij} \notin (l_{ij}, r_{ij})) \leq \sum_{i \geq j} \alpha = \alpha * C_N^2. \quad (2.2)$$

Таким образом, $\mathbb{P}(\xi = 1) \geq 1 - \alpha * C_N^2$. Иными словами, при выполнении требования $\delta_{ij} > 0$ для всех упорядоченных пар (i, j) все видео можно ранжировать в соответствии с эталонными оценками с вероятностью хотя бы $1 - \alpha * C_N^2$.

2.3.3 Сравнение алгоритмов оценки качества видео

Для каждого анализируемого алгоритма оценки качества видео можно вычислить корреляцию его значений с эталонными оценками $\beta = \beta_1, \dots, \beta_n$ на подготовленном наборе видео. Однако, вектор $\beta = \beta_1, \dots, \beta_n$ оценок качества для набора видео получен независимо для каждой из групп, соответствующих N исходным видео. Поэтому для вычисления корреляции оценок алгоритмов с эталонными оценками на всем наборе видео необходимо вычислить корреляции отдельно для каждой из N групп, соответствующим разным исходным видео, а затем вычислить среднюю корреляцию используя z-преобразования Фишера [77].

Обозначим s_k коэффициенты корреляции Спирмена между вектором значений анализируемого алгоритма объективной оценки качества $f = f_1^k, \dots, f_n^k$ и вектором эталонных оценок $\beta = \beta_1^k, \dots, \beta_n^k$. Тогда вычисление среднего коэффициента корреляции для N наборов видео будет выполняться для преобразованных коэффициентов z'_k , где:

$$z'_k = 0.5 \ln \left(\frac{1 + s_k}{1 - s_k} \right).$$

Далее вычисляется взвешенное среднее:

$$Z = \frac{1}{N} \sum_{k=1}^N z'_k \times (M_k \times K + 1),$$

и применяется обратное преобразование:

$$\hat{Z} = \frac{e^{2Z} - 1}{e^{2Z} + 1}$$

Доверительный интервал для полученной оценки:

$$\frac{(e^{2Z} - 1)}{(e^{2Z} + 1)} \leq \hat{Z} \leq (e^{2Z} - 1).$$

2.3.4 Экспериментальная оценка

Создание набора видео

Видео для нового набора данных были выбраны из коллекции, состоящей из более чем 18 000 клипов с высоким битрейтом. Данная коллекция была получена путем скачивания видео из открытых источников, в основном на сайте <https://vimeo.com/>. Видео были найдены с использованием различных ключевых слов и предлогов при поиске по сайту, которые могут обеспечить максимальный охват результатов поиска, таких как “a”, “the”, “of”, “in”, “be”, “to” и т.д. Были загружены только видео с лицензиями “CC BY” и “CC0” и с минимальным битрейтом 20 Мбит/с. Среднее значение битрейта видео из полученной коллекции составило 130 Мбит/с. Все видео были преобразованы в субдискретизацию YUV 4:2:0.

Видео из собранной коллекции были разделены на 36 кластеров. Итоговый набор видео включал видео различных жанров (спорт, игры, природа, интервью, пользовательский контент и т.д.). Таким образом, было получено 36 видео разрешения FullHD для дальнейшего кодирования.

Кодирование проводилось 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.

На рисунке 23 показано распределение битрейтов видео для созданного набора видео. Фактические значения битрейтов отличаются от целевых значений кодирования из-за использования режима кодирования VBR, однако по-прежнему удовлетворяют общим рекомендациям. Набор данных был разделен на две части: открытую и скрытую (40% и 60% от всего набора данных, соответственно). Открытая часть была опубликована на сайте, скрытая часть используется только для внутреннего тестирования. Наличие скрытой части позволяет более объективно оценивать работу методов, основанных на машинном обучении, и предотвращает возможность их переобучения на всем наборе данных. Характеристики полученных наборов данных приведены в таблице 1.

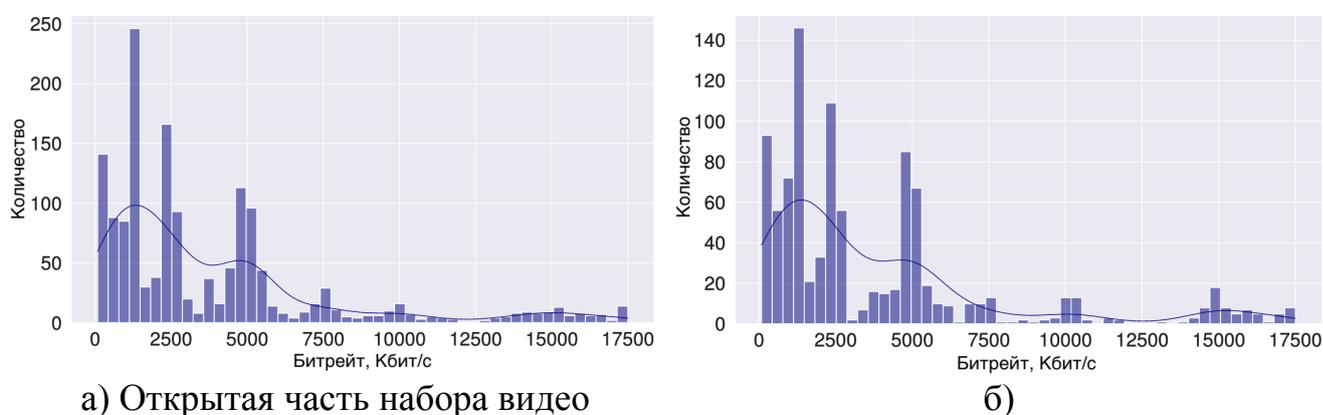


Рисунок 23 — Распределение битрейта видео в открытой и скрытой частях набора видео.

Для оценки разнообразия искажений в полученных видео был проведен анализ покрытия и плотности значений PSNR с использованием подхода, описанного в [78]. Как показано на рисунке 24, новый набор данных обеспечивает широкий диапазон качества и степени кодирования.

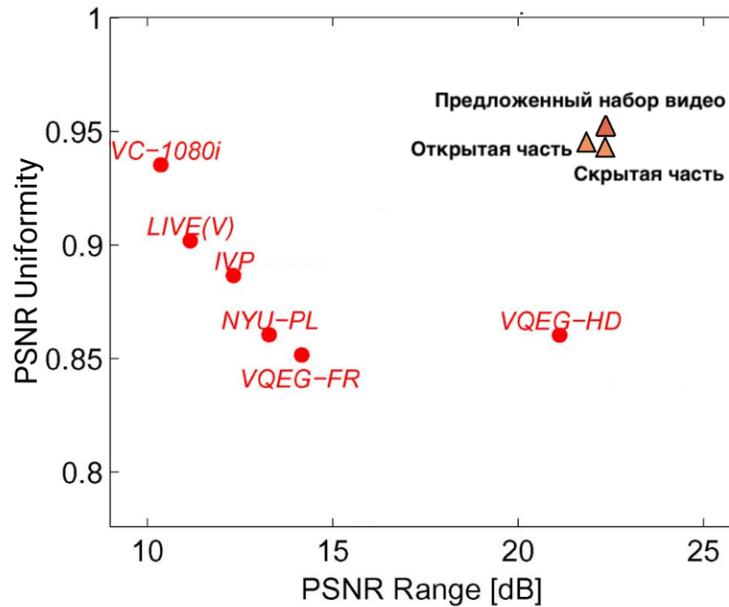


Рисунок 24 — Сравнение PSNR (диапазон и равномерность) предложенного набора видео с другими наборами видео.

Получение эталонных оценок для видео

Для проведения сравнений закодированных видео использовались платформы Subjectify.us и toloka.ai. Поскольку необходимое количество парных сравнений растёт экспоненциально с увеличением количества исходных видео, набор видеоданных был разделен на 5 подмножеств по исходным видео. Каждое подмножество содержало группу исходных видео и их закодированные версии. В каждом сравнении генерировались и сравнивались все возможные пары кодированных видео для одного исходного видео. Таким образом, в каждой паре были только видео из одного источника. Исходные видео также были включены в набор для сравнения. Видео из каждой пары предъявлялись зрителям последовательно в полноэкранном режиме. Участников просили выбрать видео с наилучшим качеством изображения или указать, что видео одинакового качества. У них также была возможность воспроизвести видео заново. В ходе процедуры сравнения участники должны были сравнить в общей сложности 12 пар видео. Две пары имели явного лидера по визуальному качеству и использовались в качестве проверочных вопросов. Ответы участников, давших неправильные ответы на проверочный вопрос, отбрасывались. Для повышения релевантности результатов было собрано не менее 10 ответов по ранжированию каждой пары видео. В

общей сложности было собрано 766 362 ответа от более чем 10 800 человек. После применения модели Брэдли-Терри к таблице парных рангов были получены эталонные оценки, согласованные в пределах группы одного исходного видео.

Поскольку эталонные оценки были получены попарным сравнением видео, полученных путем кодирования разными алгоритмами одного исходного видео, эти эталонные оценки сравнимы только внутри данных групп видео. Для каждой пары “эталонное видео — настройка кодирования” были вычислены коэффициенты корреляции Спирмена и Кендалла (SROCC и KROCC) между оценками алгоритмов и эталонными оценками. Затем для исключения нерелевантных значений, где коэффициенты корреляции вычислялись по маленьким выборкам, были выбраны только те значения, которые были рассчитаны для групп с количеством образцов больше определенного порога (15 и 6 для SROCC и KROCC соответственно).

Сравниваемые алгоритмы оценки качества видео

Для сравнения были отобраны алгоритмы оценки качества изображений и видео с открытым исходным кодом. Для тестирования у всех алгоритмов были установлены параметры по умолчанию. Алгоритмы оценки качества изображений были применены к видео покадрово, а затем было вычислено среднее значение полученных оценок качества для всего видео. Алгоритмы оценки качества видео выдают оценку для всей последовательности и не требуют дополнительного усреднения.

2.3.5 Результаты сравнения

В таблице 2 представлен анализ сравнения, проведенного на основе предложенной методики, и его сравнение с другими существующими исследованиями алгоритмов оценки качества видео.

Для более детального анализа точности методов оценки качества видео было выделено несколько категорий видео. Этими категориями являются: поль-

Сравнение	Кол-во видео	Кол-во методов	Кол-во зрителей	Искажение
Z. Sinno and A. Bovik (2018) [52]	585	4	4,776	In-the-wild видео, 80 смартфонов, 18 разрешений
Y. Li <i>et al.</i> (2020) [54]	550	15	28	H.264, H.265, QP: 22, 27, 32, 37, 42
UGC-VQA (2021) [53]	3,108 (LIVE-VQC, YouTube-UGC, KoNViD-1k)	13	>13,000	Кодирование, передача
Предложенный	2,486	26	10,800	Кодирование (H.264, H.265, AV1, VVC, и др.)

Таблица 2 — Обзор существующих сравнений алгоритмов оценки качества видео.

зовательский контент (UGC), тряска (Shaking), спорт (Sports), природа (Nature), анимация (Gaming/Animation), битрейт до 1000 кбит/с (Low Bitrate) и битрейт свыше 6000 кбит/с (High Bitrate).

Кодирование с высоким и низким битрейтом. Алгоритмы показали самые низкие корреляции для видео, закодированных с высоким битрейтом (6000 Кбит/с и выше). Причина может заключаться в низкой достоверности эталонных оценок для этой категории. Зрителям при проведении субъективных сравнений трудно различить артефакты кодирования в видео высокого качества. Для этой категории видео нереперенсные методы MDTVSA, VSFA и Linearity показали более высокие корреляции, чем все полнореференсные методы.

Кодирование при помощи стандартов HEVC, AV1 и VVC. Точность работы алгоритмов оценки качества для видео, закодированных видеокодеками стандарта H.265, выше, чем для других стандартов. Поскольку H.265 является более старым и популярным стандартом, алгоритмы оценки качества могли быть адаптированы для него. Для нового стандарта VVC лидеры отличаются по сравнению с другими стандартами кодирования: лучшим нереперенсным алгоритмом является LINEARITY, а лучшим полнореференсным — SSIM. Высокие показатели SSIM могут быть объяснены его универсальностью. Кроме того, выборка для этой категории относительно мала, поэтому для дальнейшего анализа лучших алгорит-

мов, подходящих для оценки качества кодирования VVC, может потребоваться больший набор данных.

Лидеры по категориям видеоконтента: “UGC”, “Shaking”, “Sport”, “Nature” и “Gaming/Animation”. VMAF сохранил свое лидерство во всех категориях. Его оригинальная версия, вычисляемая только для яркостной компоненты (Y-VMAF) показала лучшие результаты для контента “UGC” и “Shaking”; его модифицированная версия с использованием компонентов цветности (YUV-VMAF) является лучшей для других категорий. Нереперенсный алгоритм MDTVSFA лидирует в категориях “UGC”, “Shaking” и “Nature”; LINEARITY лидирует в категории “Sports”; а VSFA лучшая в категории “Gaming/Animation”. PaQ-2-PiQ также точно оценивает качество игрового и анимационного контента. Большинство алгоритмов оценки качества видео имеют неравномерное распределение значений для кодированных видео. Например, визуальная разница качества изображения в диапазоне SSIM 0,8–0,9 будет более заметна, чем в более низком диапазоне. Для анализа диапазонов значений других алгоритмов на рисунке 25 показано распределение нормализованных оценок алгоритмов оценки качества видео для созданного набора данных. Как показано на графике, среднее значение SSIM составляет около 0,85, что соответствует реальным данным.

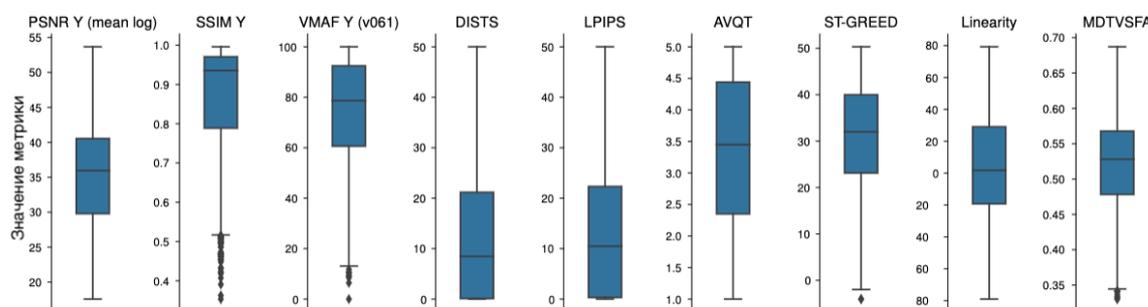


Рисунок 25 — Распределение значений методов в предложенном наборе данных.

2.4 Открытое онлайн сравнение

Результаты проведенного сравнения алгоритмов оценки качества изображений и видео были опубликованы онлайн <https://videoprocessing.ai/benchmarks/video-quality-metrics.html>. С момента публикации и на начало 2024 года было получено 15 заявок добавления новых алгоритмов в

опубликованное сравнение от различных команд разработчиков. На сайте также размещена открытая часть созданного набора данных, которую можно использовать для верификации результатов сравнения и обучения или настройки новых алгоритмов оценки качества видео.

2.5 Выводы

В данной главе предложена новая методика сравнения алгоритмов оценки качества видео. Предложенная методика состоит из алгоритма создания набора видео и получения эталонных оценок для закодированных видео. Получена вероятностная оценка для выполнения необходимого условия однозначного упорядочивания видео при разметке. Следуя предложенной методике, создан новый набор данных, содержащий 2435 видео, закодированных различными стандартами кодирования (AVC, HEVC, AV1, VVC и другие), а также для них получены эталонные оценки визуального качества от более чем 10 800 зрителей. Используя полученный набор данных была проанализирована точность работы современных методов оценки качества видео. Сравнение показало, что некоторые новые нереперенсные алгоритмы, такие как MDTVSFA, по точности близки к полнореференсным. Алгоритм VMAF показал самую высокую корреляцию с эталонными оценками. Открытая часть созданного набора данных опубликована и доступна онлайн. Предлагаемый набор данных будет полезен разработчикам и исследователям алгоритмов оценки качества изображений и видео для оценки искажений, вносимых современными видеокодеками. Набор данных может быть использован для обучения методов оценки качества и достижения более высокой корреляции с эталонными оценками в рамках решения задачи измерения качества закодированных видео.

Глава 3. Методы состязательных атак на алгоритмы оценки качества видео

Целью работы, описанной в данной главе, является разработка методов состязательных атак на алгоритмы оценки качества изображений и видео, а также разработка методики анализа их устойчивости.

3.1 Описание и постановка задачи

На вход метод состязательной атаки на неререференсный алгоритм оценки качества видео должен принимать видео $X = x_{i=1}^N$ — последовательность из N кадров, где $x_i \in \mathbb{R}^{W \times H \times 3}$, метод оценки качества $M(X) : X \rightarrow \mathbb{R}$ и параметры ε, q .

Выходом алгоритма является видео $X^* = x_{i=1}^{*N}$ такое, что:

$$M(X^*) > M(X), \|x_i - x_i^*\|_p \leq \varepsilon. \quad (3.1)$$

Требуется найти преобразование $f : x \rightarrow x^*$.

Постановка задачи отличается для методов атак на неререференсные и полнореференсные алгоритмы оценки качества видео. В случае полнореференсных методов, на вход метод состязательной атаки должен принимать видео $X = x_{i=1}^N$ — последовательность из N кадров, где $x_i \in \mathbb{R}^{W \times H \times 3}$, метод оценки качества $M(X, X) : (X, X) \rightarrow \mathbb{R}$ и параметры ε, q . Выходом алгоритма является видео $X^* = x_{i=1}^{*N}$ такое, что:

$$M(X, X^*) > M(X, X), \|x_i - x_i^*\|_p \leq \varepsilon. \quad (3.2)$$

Анализ полнореференсных алгоритмов оценки качества видео, проведенный в предыдущей главе, показал, что наилучшую корреляцию с визуальными оценками имеет метод Video Multimethod Assessment Fusion (VMAF) [20], разработанный компанией NETFLIX. Схема алгоритма представлена на рисунке 31. VMAF быстро стал популярен среди разработчиков видеокодеков благодаря техническому блогу NETFLIX на Medium [20] и большому количеству публикаций в других блогах. Согласно заявлениям авторов, алгоритм показывает высокую

корреляцию с эталонными оценками визуального качества по результатам множества сравнений и исследований. VMAF имеет открытый исходный код и имеет возможность подключения пользовательских моделей, обученных на различных наборах данных. Сложность разработки методов состязательных атак на алгоритм VMAF заключается в том, что он не является дифференцируемым [39]. Это усложняет разработку методов состязательных атак, однако он по-прежнему остается уязвим для методов, развернутых в режиме «черного ящика». В данной главе предложены методы состязательных атак на полнореференсные алгоритмы оценки качества видео на примере алгоритма VMAF.

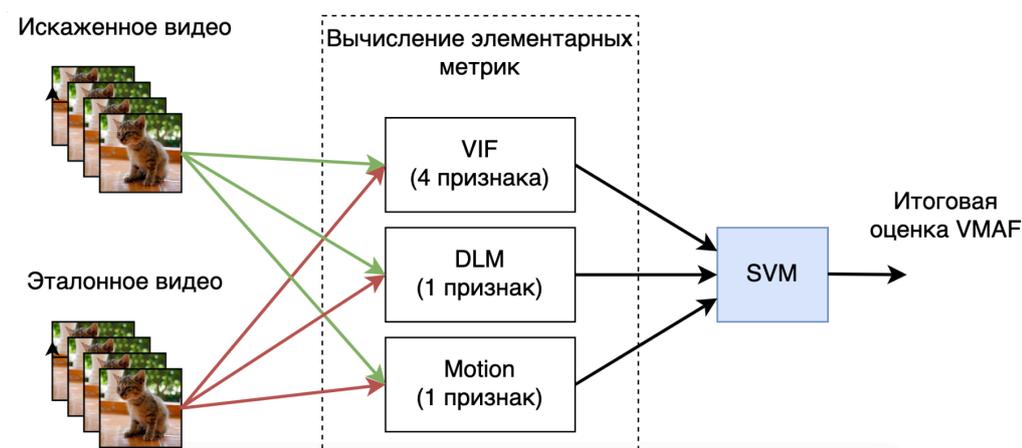


Рисунок 31 — Схема работы метода Video Multimethod Assessment Fusion (VMAF).

3.2 Обзор существующих методов

Rec. ITU-T J.149 [79] разделяет результаты сравнений с использованием алгоритмов оценки качества изображений и видео и визуального качества на четыре типа: false ordering, false differentiation, false tie и correct decision (таблица 3). Согласно данной классификации, существуют разные способы атак на алгоритмы оценки качества видео, например, увеличение или уменьшение их значений. В данной работе рассматриваются только атаки, повышающие оценки, выдаваемые алгоритмами измерения качества видео, поскольку такого рода атаки имеют большую область применения. Таким образом, рассматриваемые методы атак направлены на увеличение оценки качества видео, что потенциально

	Визуально хуже	Визуально одинаково	Визуально лучше
Объективно хуже	Correct decision	False differentiation	False differentiation
Объективно одинаково	False tie	Correct decision	False tie
Объективно лучше	False differentiation	False differentiation	Correct decision

Таблица 3 — Классификация ошибок методов оценки качества изображений и видео.

может привести к ошибкам типа false ordering (когда визуальное качество снижается, а оценка данная атакуемым методом повышается) или ошибкам типа false differentiation (когда визуальное качество остается неизменным, а оценка, выдаваемая алгоритмом, увеличивается).

Впервые идея разработки процедуры тестирования алгоритмов оценки качества изображений и видео к атакам путем генерации специальных примеров была представлена в работе Циарамелло и Рейбмана [80]. Авторы генерировали изображения с помощью бинарного поиска по набору искажений для генерации примеров атак на алгоритмы оценки качества. Они также представили метод тестирования алгоритмов оценки качества изображений в режиме черного ящика, оценивающий их уязвимости к возмущениям в исходных изображениях [81]. Авторы получили информацию о производительности метода для ограниченного набора искаженных изображений и затем выявили несоответствия оценок с эталонными на основе полученных результатов. Лю и Рейбман [82] представили программное обеспечение под названием “STIQE”, которое автоматически измеряет точность работы метода оценки качества изображения. Оно позволяет пользователям выполнить серию тестов, а затем построить отчеты о работе методов. Тестирование заключается в добавлении к изображениям нескольких искажений различной интенсивности и проверке того, монотонно ли растёт оценка качества, выдаваемая тестируемым методом, при увеличении силы применяемых искажений. Ванг и Симончелли [83] предложили метод сравнения алгоритмов оценки качества изображений, названный соревнованием максимальной дифференциации (maximum differentiation competition, или MADC). Он

позволяет сравнивать два метода оценки качества изображений путем построения пары примеров, которые повышают или снижают значения одного из методов при неизменном значении оценки другого. Затем эта процедура повторяется, но роли двух методов качества меняются местами. Чтобы выявить уязвимости в методе оценки качества, необходимо провести визуальное сравнение таких построенных пар примеров. Если на этих парах оценки одного метода лучше коррелируют с эталонными визуальными оценками, чем оценки другого метода, первый метод считается более надежным. Кеде Ма и др. [84] расширили идею MADС, предложив групповое соревнование максимальной дифференциации (group maximum differentiation competition, или gMADС). Эта техника использует группу моделей качества, названную «атакующие» модели, для автоматической генерации примеров, которые максимизируют или минимизируют их оценки, сохраняя при этом оценки другой устойчивой модели без изменения. Жан и др. [85] продемонстрировали устойчивость предложенного ими неререференсного метода оценки качества изображений с помощью gMADС.

Было опубликовано несколько исследований, посвященных состязательным атакам на алгоритмы оценки качества изображений и видео. Санг и др. [86] предложили итеративный алгоритм, основанный на адаптивных искажениях входного изображения, для атаки на алгоритмы оценки качества. Этот алгоритм основан на методе MI-FGSM [87] и для входного изображения x генерирует изображение x' с увеличенным значением метода оценки качества, такое что $\|x - x'\| \leq \varepsilon$. Параметр ε для каждого изображения выбирается с помощью метода NIQE [88]: чем ниже значения NIQE, тем больше ε , поскольку более низкие показатели NIQE соответствуют зашумленным и текстурированным изображениям. Корхонен и Ю [89] предложили метод итеративной атаки на неререференсные алгоритмы оценки качества изображений, который использует фильтр Собеля для добавления искажений на границы объектов. Такой подход помогает улучшить визуальное качество атакованных изображений. Для снижения визуальных искажений в атакованных изображениях Чжан и др. [90] предложили итеративную атаку, которая использует различные полнореференсные алгоритмы оценки качества, такие как расстояние Чебышева, SSIM [19], LPIPS [33] и DISTС [34]. Кеттунен и др. [45] показали, что полнореференсный алгоритм оценки качества изображений LPIPS также подвержен состязательным атакам. Так как LPIPS дифференцируем, авторы использовали градиентные методы и создали пары

изображений, которые отличаются визуально, но LPIPS считает их близкими по качеству.

Большинство существующих неререференсных алгоритмов оценки качества видео имеют нейросетевые архитектуры [24; 26—31], поэтому для создания атакующего изображения может быть применен метод градиентного спуска. Однако время обработки одного изображения таким методом может быть долгим, так как для вычисления точных градиентов требуется выполнить обратное распространение ошибки по нейронной сети один или несколько раз. Прямое применение градиентного спуска для создания атакующего изображения имеет низкую практическую применимость в алгоритмах обработки видео из-за высокой временной сложности. Более подходящими для состязательных атак на алгоритмы оценки качества видео являются методы, основанные на обучении универсальных возмущений. Моосави-Дезфули и др. [91] определили универсальное возмущение как изображение, которое может обмануть классификатор изображений, основанных на сверточной нейронной сети. Однако в отличие от классификаторов, при атаке на алгоритм оценки качества видео не существует однозначного определения успешности/неуспешности возмущения, поскольку добавление возмущения к исходному изображению изменяет оценку атакуемого метода непрерывно.

Описанные выше методы на основе белого ящика подходят для исследования устойчивости алгоритмов оценки качества изображений, дифференцируемых по входным параметрам. Для недифференцируемых методов оценки качества можно применить атаки в режиме «черного ящика». В основном они осуществляются путем поиска преобразования видео, которое увеличивает целевые значения атакуемого метода оценки качества. Поиск подходящих преобразований может быть ускорен с помощью бинарного поиска, генетических алгоритмов или машинного обучения. Атаки в режиме «черного ящика» проводятся без доступа к градиентам рассматриваемой оценки качества. Их можно также разделить на две группы:

1. Метод не аппроксимирует и не использует градиенты исходной функции. В эту категорию попадают такие методы, как алгоритм Монте-Карло и генетические алгоритмы.
2. Метод включает в себя создание (и, возможно, дообучение) дифференцируемой аппроксимации исходной оценки качества, исходного классификатора или аппроксимацию непосредственно значений градиентов рассматриваемой функции в точке.

Атаки в режиме «черного ящика» также в основном были использованы против моделей классификации изображений. Янг и др. [92] предложили метод состязательной атаки на основе запросов к неререференсным алгоритмам оценки качества. Они использовали концепцию оценки границ значений метода для генерации возмущения. Однако итеративные методы требуют большого количества запусков для успешности атаки.

3.3 Описание предложенных методов

3.3.1 Метод состязательной атаки на неререференсные алгоритмы оценки качества видео

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 — алгоритм оценки качества изображений или видео. Состязательная атака на алгоритм оценки качества видео с помощью универсального возмущения U формулируется следующим образом:

$$U = \underset{f}{\operatorname{argmax}} \frac{1}{N} \sum_{i=1}^N (M(x_i + U) - M(x_i)), \|U\|_{\infty} < \varepsilon. \quad (3.3)$$

Пусть $batch$ — это партия изображений для обучения и U — обучаемое универсальное возмущение. Тогда для обучения универсального возмущения функция потерь может быть записана в виде уравнения:

$$L = 1 - \frac{\frac{1}{batch} \sum_i \operatorname{metric}(I_i + U)}{a} \quad (3.4)$$

где a — коэффициент нормализации.

Алгоритм 1 описывает предложенный метод обучения универсального возмущения U .

На каждой итерации путем вычисления градиента атакуемого нейросетевого метода оценки качества по входному изображению определяется возмущение, необходимое для увеличения оценки целевого метода. Обновление возмущения происходит с помощью оператора проекции. Алгоритм работает до тех пор, пока средний показатель атакуемого метода оценки качества на обучающем наборе

Input: обучающий набор изображений D , атакуемый нереперенсный алгоритм оценки качества M , L_p -норма возмущения величины ε

Output: Универсальное возмущение U

Initialize $U \leftarrow 0$;

while *not converged* **do**

for *each image* x **in** D **do**

$x' \leftarrow x + U$;

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

 Compute the gradient of M with respect to U ;

 Update U using the gradient;

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

end

end

Algorithm 1: Генерация универсального возмущения предложенным методом

изображений не будет незначительно отличаться между соседними итерациями или пока не будет достигнуто максимальное количество итераций. Данный подход для атаки на алгоритмы оценки качества видео является обобщаемым: одно универсальное возмущение может применяться для атаки любого нового изображения или кадров видео, при этом время обработки каждого кадра фиксировано.

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:

$$\left| |U_{(u,v)}| - |G^k_{(u,v)}| \right| = \left| \left(\left| C_\varepsilon \sum_{i=1}^N G^i \right| - |G^k| \right)_{(u,v)} \right|. \quad (3.5)$$

Доказательство. Итеративное обновление U происходит по формуле $U^n = C_\varepsilon (G^n + U^{n-1})$. Обозначим $B = G^n_{(u,v)}$, $A = (G^{n-1} + U^{n-1})_{(u,v)}$. Из условия теоремы A и B имеют один знак. Тогда:

$$C_\varepsilon (C_\varepsilon (A) + B) = C_\varepsilon (A + B). \quad (3.6)$$

Докажем это равенство. Рассмотрим случай 1: $|A| < \varepsilon$. Тогда $C_\varepsilon (A) = A$ и равенство 3.5 верно. Теперь рассмотрим случай 2: $|A| \geq \varepsilon$. Тогда $C_\varepsilon (A) = \text{sign}(A) \varepsilon$.

Так как по условию $AB > 0$, $C_\varepsilon(C_\varepsilon(A) + B) = \text{sign}(A)\varepsilon = C_\varepsilon(A + B)$. Значит, в точке (u, v) :

$$U = C_\varepsilon(G^N + C_\varepsilon(G^{N-1} + \dots + C_\varepsilon(G^1))) = C_\varepsilon\left(\sum_{i=1}^N G^i\right) \quad (3.7)$$

и равенство 3.5 верно.

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

Доказательство. В силу инвариантности 3.5 от номера итерации метод сходится за одну итерацию.

Для генерации универсального возмущения можно использовать сверточную нейронную сеть, например, архитектуру U-Net [93].

3.3.2 Методы состязательных атак на полнореференсные алгоритмы оценки качества видео

Метод состязательной атаки на алгоритм оценки качества видео VMAF с помощью оптимизации параметров обработки видео генетическим алгоритмом

В данном разделе описана атака с использованием генетического алгоритма для оптимизации значений параметров алгоритмов обработки видео, которые будут повышать оценку качества, выдаваемую методом VMAF. Для предварительного поиска преобразований, которые могут повысить оценку VMAF, было выбрано четыре видео различного содержания и с разной пространственной и временной сложностью, яркостью и контрастностью. Все видео имели разрешение FullHD и высокий битрейт. Для преобразования цвета и контрастности были выбраны два известных и широко распространенных алгоритма обработки изображений: нерезкое маскирование и выравнивание гистограммы. Нерезкое маскирование имеет два параметра, которые влияют на уровень изображения: radius (радиус гауссова размытия) и amount (количество добавленного контраста по краям). Для обработки видео путем эквализации гистограммы был проанализирован параметр clipping limit. Для поиска оптимальных конфигураций параметров

эквализации использовался алгоритм многоцелевой оптимизации NSGA-II [94]. Генетическому алгоритму задавались только ограничения на параметры, и он применялся для поиска оптимальных параметров для каждого тестового видео.

Для некоторых конфигураций метода выравнивания гистограммы оценка качества алгоритмом VMAF становится выше (увеличивается с 68 до 74), при этом SSIM практически не изменяется (уменьшается с 0,88 до 0,86). Для видео Crowd Run значение VMAF не было увеличено при помощи алгоритма нерезкого маскирования, однако оно было увеличено при помощи выравнивания гистограммы. Для видео Red Kayak и Speed Bag с помощью нерезкого маскирования были значительно увеличены оценки VMAF и при этом значения SSIM практически не уменьшились. На рисунке 32 представлены результаты применения нерезкого маскирования с разными настройками на одном из тестовых видео. Таким образом, оба протестированных метода обработки видео подходят для моделирования состязательной атаки на VMAF.

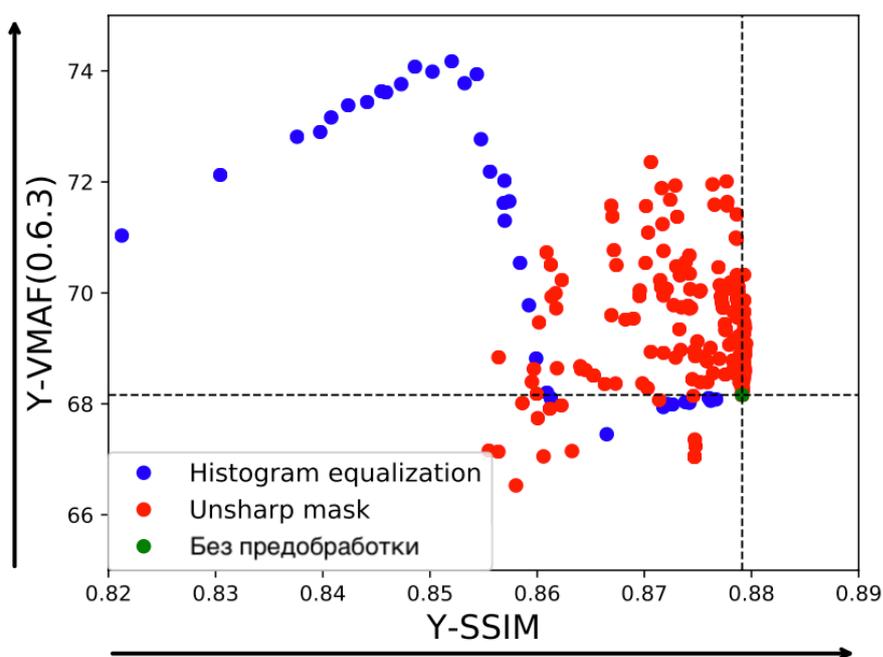


Рисунок 32 — Сравнение оценок VMAF и SSIM для различных конфигураций нерезкого маскирования и выравнивания гистограммы на видео Bay timelapse. Важны результаты во втором квадранте, где значения SSIM не изменились, а значения VMAF увеличились.

Предложенный метод атаки на VMAF основан на оптимизации параметров расширенного набора алгоритмов обработки видео. В него входят алгоритмы тонового картирования (метод тонового картирования Драго [95], метод тонового картирования Мانتюка [96], Retinex и метод тонового картирования

Рейнхарда [97]), адаптивная гистограммная эквализация с ограничением контраста (CLAHE), гистограммная эквализация, преобразования на основе контраста (гамма-коррекция, трехградусное полиномиальное 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 ($\mu + \lambda$) [98]. Схема метода представлена на рисунке 33. Так как сжатие видео сглаживает добавленные возмущения, его применение к атакованным видео снижает эффективность атаки. Поэтому в процессе подбора параметров алгоритмов обработки видео предлагается использовать сжатие видеокодеком x264.

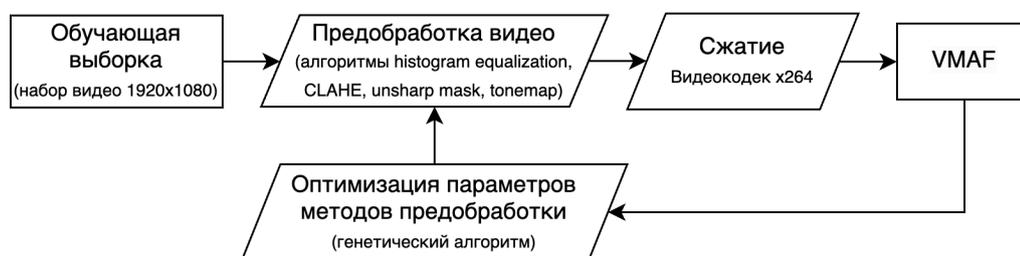


Рисунок 33 — Схема предлагаемого метода состязательной атаки на VMAF.

Метод состязательной атаки на алгоритм оценки качества видео VMAF с помощью аппроксимации

В качестве базового метода был взят подход, предложенный в статье Чена и др. [99]. В ней авторы рассматривают задачу обучения нейросетевого кодировщика для сжатия видео с потерями. Авторы ставят задачу максимизации оценки качества по методу VMAF при сохранении качества кодирования, обучая нейросетевой кодек. В силу того, что оценка качества VMAF не является дифференцируемой, невозможно напрямую использовать ее в ходе оптимизации, и авторы параллельно с кодировщиком обучают промежуточную сеть, которая осуществляет предсказание значения VMAF для закодированного видео по отношению к исходному. Такая нейронная сеть позволяет осуществлять диффе-

ренцируемую аппроксимацию оценки качества VMAF и может напрямую быть использована в функции потерь в ходе обучения основной нейронной сети.

Архитектура нейронной сети, осуществляющей обработку изображения, показана на рисунке 34. Архитектура нейронной сети, осуществляющей аппроксимацию VMAF, была взята из статьи [99] без изменений. Первым шагом при разработке предложенного метода было обучение покадровой модели повышения VMAF для изображений. В этой схеме обучающей выборкой для метода является множество изображений I_n размером по меньшей мере 256×256 и кратным 16 по каждому измерению.

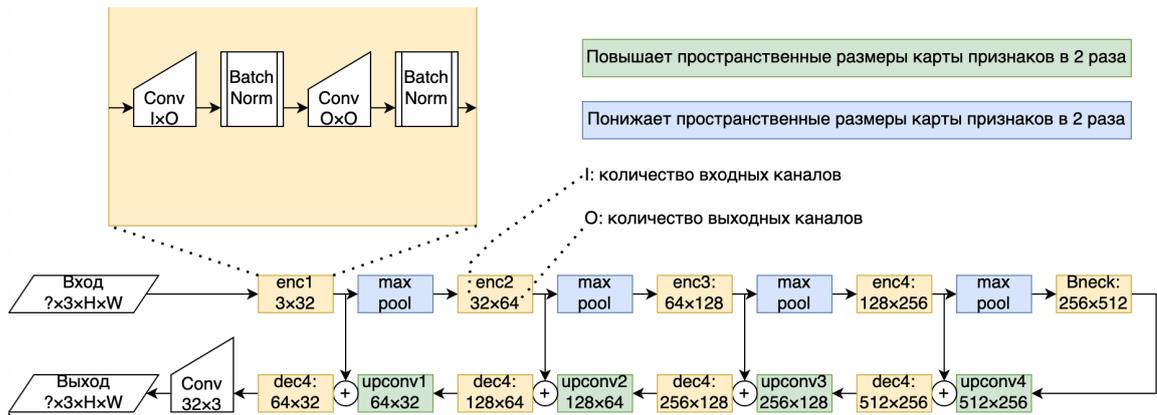


Рисунок 34 — Используемая архитектура на базе U-Net.

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 , $G(B_{in}) = B_{out}$. G_{frozen} обозначает, что в ходе применения параметры нейронной сети заморожены и не накапливают градиенты функции потерь по себе. Аппроксимирующая оценки VMAF сеть принимает на вход набор пар трехканальных изображений $B_{in} \in R_{[0..1]}^{batch \times 6 \times W \times H}$ и возвращает $Q_{pred} \in R_{[0..1]}^{batch}$ — спрогнозированное 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 трех компонент.

$$L = L_{vmaf} + \lambda_{proxy} L_{proxy} + \lambda_{stable} L_{stable} \quad (3.8)$$

Вспомогательная сеть P решает задачу аппроксимации VMAF для искаженного и исходного изображения. В общую функцию потерь она вносит следующий вклад:

$$L_{\text{proxy}} = \text{mean} \left(\left(P(B_{\text{out}}, B_{\text{in}}) - \text{VMAF}(B_{\text{out}}, B_{\text{in}}) \right)^2 \right), \quad (3.9)$$

$$B_{\text{out}} = G_{\text{frozen}}(B_{\text{in}}) \quad (3.10)$$

Основная сеть обучалась обрабатывать изображение так, чтобы максимизировать оценку VMAF получившегося 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. Также в ходе обучения использовался факт, что изображение на выходе основной нейронной сети зачастую визуально похоже на изображение на входе. Использование меры близости в евклидовом пространстве изображений в качестве стабилизирующей добавки к функции потерь позволяет нейронной сети быстрее обучаться на ранних этапах работы, а так же штрафует сильно отличающиеся от исходного выходные изображения.

$$L_{\text{vmf}} = M - P_{\text{frozen}}(G(B_{\text{in}}), B_{\text{in}}), \quad (3.11)$$

где M — произвольная константа, принятая равной 1.

$$L_{\text{stable}} = \|G(B_{\text{in}}) - B_{\text{in}}\|_2 \quad (3.12)$$

$$L_{\text{SSIM}} = \|\text{SSIM}(G(B_{\text{in}})) - \text{SSIM}(B_{\text{in}})\|_2 \quad (3.13)$$

3.3.3 Метод тестирования устойчивости алгоритмов оценки качества видео к атакам

Чтобы сравнить устойчивость различных методов оценки качества до и после атаки, необходимо соотнести повышение значений целевого метода и снижение оценок прокси-метода для разных алгоритмов оценки качества. Для этого предлагается использовать следующий подход. Считается, что тестовые видео

были атакованы перед сжатием видео, и уже проведено кодирование видео с различными битрейтами, а также вычислены значения целевых и прокси-методов с атакой и без нее. Для оценки устойчивости целевого метода:

1. Вычисляются максимальные и минимальные оценки целевого метода на RD-кривых для N видео ($K + 1$ кривая для каждого видео, $(K + 1) * N$ кривых всего). Результат — диапазон оценок целевых методов для тестового набора данных.
2. Выполняется нормализация всех RD-кривых по оценкам целевого метода к диапазону от 0 до 1 путем вычитания минимума и последующего деления на разницу между максимумом и минимумом. Независимо от метода, в результате получаются RD-кривые, где оценки по целевым методам находятся в диапазоне от 0 до 1.
3. Выполняется аналогичная шагам 1 и 2 нормализация для всех RD-кривых прокси-методов. Минимальное и максимальное значение прокси-метода вычисляется с помощью $M * N * (K + 1)$ RD-кривых.
4. Для целевого метода вычисляется значение площади между RD-кривой, соответствующей атакованному видео и RD-кривой без предварительной обработки. Площади вычисляются для области, где определены обе кривые. Прирост значений целевого метода показывает процентное увеличение диапазона целевого метода после предварительной обработки (атаки).
5. Как и на шаге 4, вычисляется площадь между RD-кривой с предварительной обработкой и без предварительной обработки как оценка потери качества с помощью прокси-метода.
6. Значения целевого метода и прокси-метода усредняются по всем тестовым видео и для всех уровней амплитуды атаки.
7. Строится график зависимости повышения значений целевого метода от снижения значений прокси-метода, содержащий K точек, соответствующих K уровням силы применения атаки.

На рисунке 35 представлена схема алгоритма, описанного выше. На нем также показана зависимость повышения значений целевого метода от снижения прокси-метода, которая строится для разных уровней амплитуд атаки. Учитывая эту зависимость можно сравнить все целевые методы друг с другом. Линия у устойчивого метода имеет нулевой или отрицательный наклон, то есть универсальное возмущение должно либо оставить оценки атакуемого метода

неизменными, либо уменьшить их. Для уязвимых методов линия имеет положительный наклон, так как обученные универсальные возмущения повышают оценку метода в этих случаях.

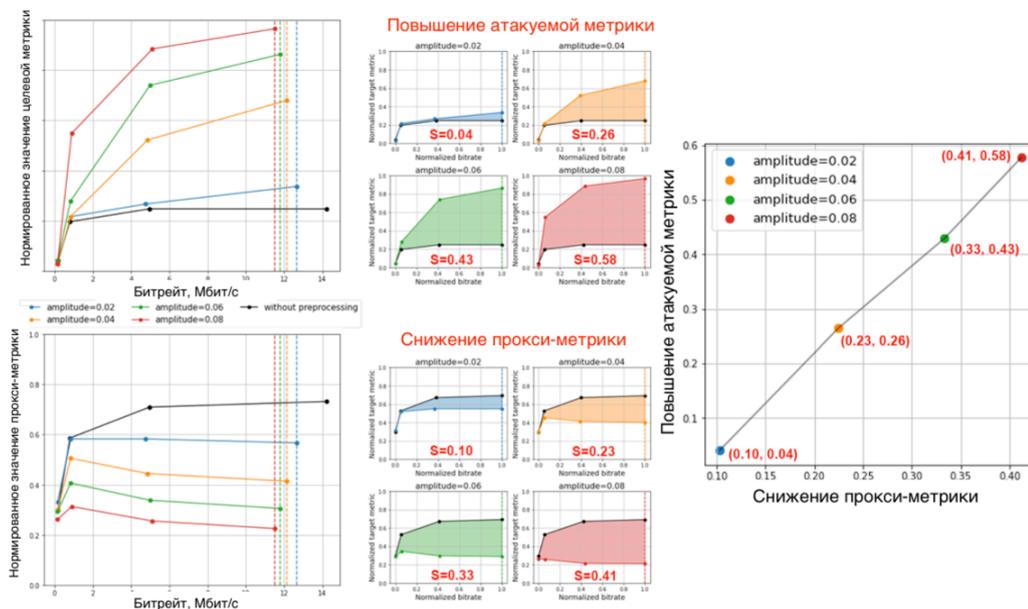


Рисунок 35 — Предложенный метод вычисления повышения целевого метода и потерь прокси-метода с помощью нормализованных RD-кривых для видео на четырех уровнях амплитуды атаки. Правый график показывает зависимость повышения целевого метода и снижения прокси-метода для этих уровней амплитуды.

3.4 Экспериментальная оценка

3.4.1 Оценка методов состязательных атак на нерепрезентативные алгоритмы оценки качества на основе построения универсальных возмущений

Объективная оценка

Для обучения универсального возмущения использовался набор из 10000 изображений из набора данных COCO [100]. Размер возмущения совпадает с размером изображений в обучающей выборке и равен 256×256 пикселей. Во время

обучения значения возмущения ограничивались диапазоном от $-0,1$ до $0,1$. Оптимизация проводилась с помощью алгоритма Adam [101] с шагом $0,001$. Размер партии изображений был равен четырем.

Состязательные атаки предложенным методом обучения универсального возмущения были проведены для семи дифференцируемых нереперенсных методов оценки качества изображений и видео (PaQ-2-PiQ [27], Linearity [30], VSFA [24], MDTVSA [26], KonCept512 [28], Nima [31] и SPAQ [29]) на $10\,000$ изображений из набора данных COCO [100]. На рисунке 36 показана визуализация обученных универсальных возмущений.

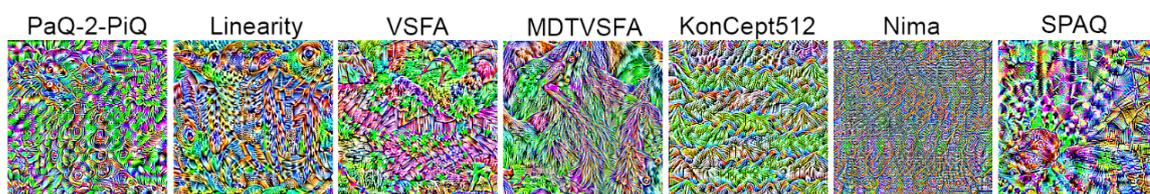


Рисунок 36 — Обученные универсальные возмущения для всех семи тестируемых методов.

Полученные универсальные возмущения для всех целевых методов были применены к 200 тестовым изображениям размером 256×256 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 $0,6$). Предложенный метод атаки был сопоставлен с градиентным методом состязательной атаки MADC [83] на метод PaQ-2-PiQ. В первой строке Таблицы 4 представлены результаты применения обученного на повышение значений PaQ-2-PiQ универсального возмущения. Вторая строка содержит результаты применения к изображениям атаки MADC с теми же уровнями амплитуды, где PaQ-2-PiQ увеличивается, а MSE остается ниже уровня амплитуды. Для каждого изображения проведено 1000 шагов градиентного спуска. Метод состязательной атаки MADC способен еще выше поднимать оценку качества, но проведение вычислений занимает большое количество времени, так как требует нескольких шагов обратного распространения градиента. В проведенных экспериментах состязательная атака предложенным методом универсального возмущения повышает оценку, выдаваемую атакуемым методом, на 70% от повышения, достигнутого атакой MADC, но делает это всего за 3% времени выполнения атаки MADC.

	Сила атаки							
	0.02	0.04	0.06	0.08	0.1	0.2	0.4	0.6
Предложенный метод	4.2	17.1	27.7	35.8	42.3	60.9	74.2	79.1
MADC [83]	28.2	39.9	47.3	52.7	56.8	69.1	76.8	77.8

Таблица 4 — Средние различия PaQ-2-PiQ до и после применения предложенной атаки и MADC к 200 тестовым изображениям. Оценки методов для исходных изображений вычитаются из оценок после атаки.

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 несжатых видео разрешения FullHD из набора данных Xiph.org [102], четыре уровня амплитуды атаки (0,02, 0,04, 0,06 и 0,08) и сжатие с использованием видеокодека x264 стандарта H.264 с четырьмя битрейтами (200 кбит/с, 1 Мбит/с, 5 Мбит/с и 12 Мбит/с). На рисунке 8 показаны атакованные видеокadres после сжатия. Зависимость повышения атакуемых методов от снижения значений прокси-методов проиллюстрирован на рисунке 37.

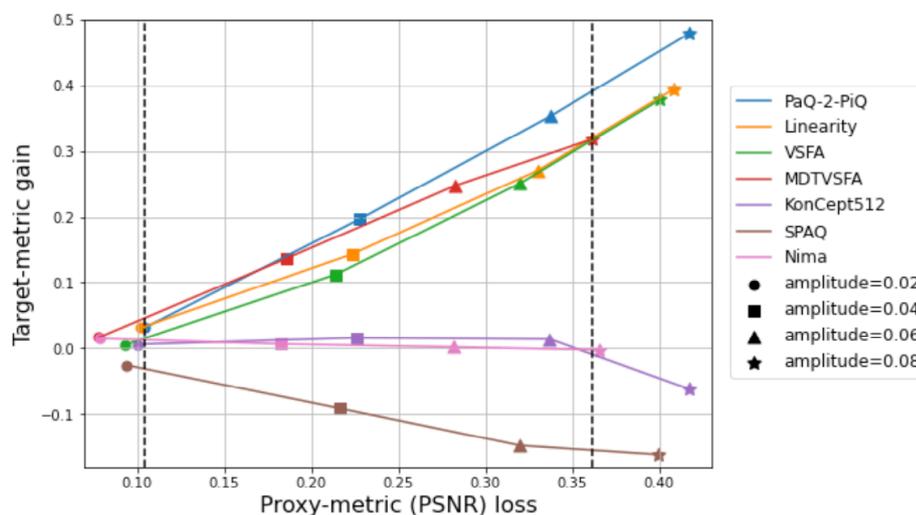


Рисунок 37 — Зависимости между приростом значений целевого метода и потерями прокси-методов для всех протестированных нереперенсных методов. Пунктирными линиями выделена область, где определены зависимости для всех методов. Чем меньше прирост значений целевого метода и выше потери PSNR, тем стабильнее метод.

Была вычислена оценка устойчивости для семи популярных методов оценки качества изображений. Согласно предложенному алгоритму вычисления устойчивости, отрицательный балл стабильности соответствует уязвимым методам, а положительный балл стабильности — стабильным методам. Большой балл стабильности лучше, поскольку метод правильно реагирует на добавление универсального возмущения, которое снижает качество исходных изображений и видео. В таблице 5 представлен рейтинг протестированных нереперенсных методов по предложенной оценке стабильности. В первом столбце таблицы представлен рейтинг методов по предложенному показателю стабильности. Согласно полученным результатам, четыре протестированные метода оценки качества изображений (PaQ-2-PiQ, Linearity, VSFA и MDTVSFA) оказались уязвимыми к предложенной атаке методом универсального возмущения, в то время как три (Koncept512, Nima и SPAQ) оказались устойчивыми. Во втором и третьем столбцах приведены корреляции SROCC, полученные в результате визуальных сравнений, проведенных с использованием эталонов MSU и BVQA. Балл стабильности характеризует надежность метода оценки качества; корреляция SROCC характеризует точность метода на данном наборе данных. MDTVSFA показал лучшую корреляцию, но он уязвим к предложенной атаке.

	Оценка стабильности	SROCC	
		MSU benchmark	BVQA benchmark
PaQ-2-PiQ	-5.3	0.89	0.61
Linearity	-4.2	0.92	-
VSFA	-3.8	0.91	0.75
MDTVSFA	-4.8	0.94	0.78
Koncept512	-0.3	0.86	0.73
SPAQ	2.6	-	-
Nima	-0.1	0.85	-

Таблица 5 — Оценки устойчивости и корреляции SROCC для протестированных методов.

Визуальная оценка

На рисунке 38 представлены изображения до и после добавления универсального возмущения, обученного для повышения значений PaQ-2-PiQ с тремя уровнями амплитуды.



Рисунок 38 — Различные амплитуды универсального возмущения PaQ-2-PiQ, примененные к изображениям. Чем выше амплитуда, тем выше оценка PaQ-2-PiQ, но при этом визуальные искажения также становятся более заметными.

Проведенный визуальный анализ изображений и видео, полученных в результате универсальных атак с возмущением, показал, что вносимые искажения распространены равномерно по кадру и достаточно заметны при высоких уровнях амплитуды. Однако существуют способы уменьшить видимость такой атаки. Например, можно применить универсальное возмущение пропорционально контрасту, поскольку люди меньше замечают искажения в высококонтрастных областях. Для демонстрации была вычислена маска функции контрастной чувствительности (CSF) (рисунок 39 (a)), маска была умножена на универсальное возмущение, повышающее PaQ-2-PiQ, и полученное возмущение было добавлено к исходному изображению (рисунок 39 (b)). Как видно из рисунка, такое преобразование все еще может увеличить значения PaQ-2-PiQ, но в то же время оно практически незаметно.

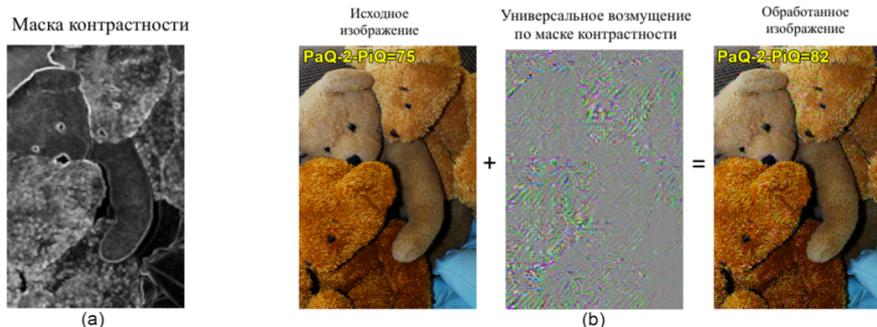


Рисунок 39 — CSF маска и ее применение к универсальному возмущению на примере атаки PaQ-2-PiQ.

3.4.2 Оценка методов состязательных атак на полнореференсные алгоритмы оценки качества видео

Объективная оценка

Для оценки эффективности предложенных методов состязательных атак на полнореференсные недифференцируемые алгоритмы оценки качества видео, предложенные методы были применены для атак алгоритма VMAF.

Для оценки эффективности метода на основе генетического алгоритма был создан специальный набор данных, содержащий 618 видео различных жанров разрешения FullHD. Предложенный метод был применен к данным видео и видео были закодированы с использованием видекодека x265 и набором битрейтов: 2, 4, 6, 8 и 10 Мбит/с.

Набор алгоритмов обработки видео, используемый для атаки, включал комбинацию гамма-коррекции и нерезкого маскирования, а также CLANE. Результаты показали, что метод CLANE обеспечил наилучшее увеличение VMAF: среднее повышение значений метода составило 86,8%, медианное значение — 50,2%; самое большое повышение — 218,8%. Для VMAF NEG комбинация гамма-коррекции и нерезкого маскирования показала наибольшее повышение значений метода после обработки. Она увеличила VMAF NEG в среднем на 6%. На рисунке 310 проиллюстрирована эффективность протестированных алгоритмов обработки видео с подобранными параметрами для атаки VMAF.

В таблице 6 представлено сравнение предложенных методов повышения VMAF и VMAF NEG с методами, описанными в других работах. Стоит заметить,

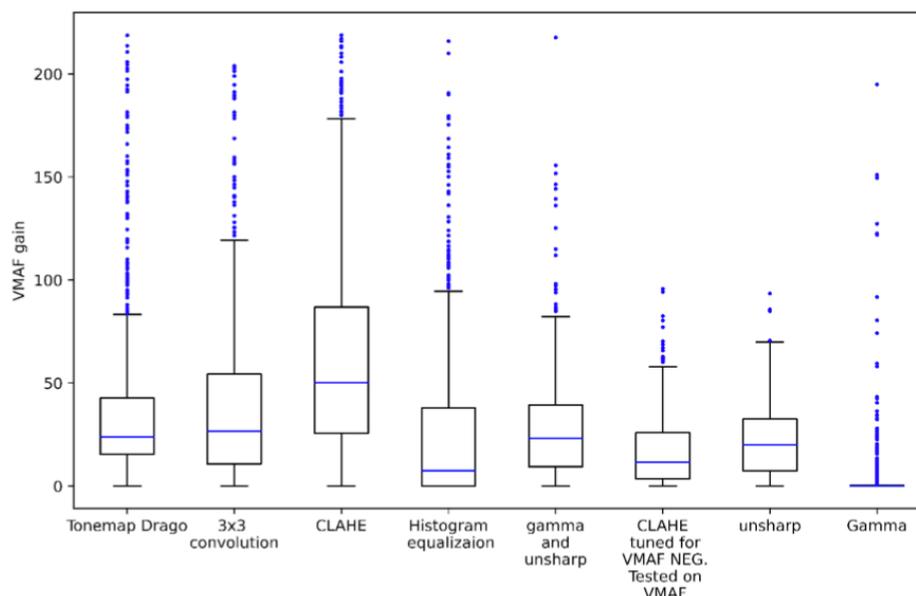


Рисунок 310 — Повышение значений алгоритма VMAF с помощью протестированных методов обработки видео.

что самая первая публикация по данной теме (Hacking VMAF with Video Color and Contrast Distortion) [14] была сделана автором работы с соавторами, остальные работы были опубликованы после нее для проверки полученных автором результатов. Так, в работе VMAF Based Rate-Distortion Optimization for Video Coding [42] применяется метод, предложенный в данной диссертации, практически без изменений. В нем добавляется дополнительная оптимизация сжатия блоков видео, что было возможным, так как авторы рассматривали модификацию конкретного видекодека libaom для повышения VMAF. В публикации VMAF is Hackable: What Now? [103] автор подтверждает работоспособность предложенного метода на независимом наборе данных. В работе On VMAF's property in the presence of image enhancement operations [104] авторами самого метода VMAF из Netflix проверяется работоспособность предложенных методов состязательных атак и предлагается улучшенная версия VMAF NEG. Видно, что новая версия алгоритма VMAF NEG все еще может быть атакована с помощью предложенного метода.

Для предложенного метода на основе создания аппроксимации была рассмотрена процедура обучения на двух различных наборах данных. Первый из них создан по описанной авторами работы [99] методологии: случайным выбором и масштабированием изображений из обширного разнообразного набора данных для классификации. В качестве такого набора был выбран Pascal VOC [105], размер получившегося набора данных составил 5000 изображений в обучающем и 1000 изображений в валидационном наборах. Исходные изображения были

Метод	Методы обработки	Макс. повышение VMAF	Макс. повышение VMAF NEG
VMAF Based Rate-Distortion Optimization for Video Coding	Обработка (unsharp mask) и оптимизация кодирования	6,6	-
Hacking VMAF with Video Color and Contrast Distortion	Выравнивание гистограммы	6	-
	Unsharp mask	5	-
Toward a Better Quality Metric for the Video Community	Повышение резкости	2,58	-
On VMAF's property in the presence of image enhancement operations	Выравнивание гистограммы	46,58	18,72
	Повышение резкости	14,56	12,09
	tune=vmaf	2,58	4,01
VMAF is Hackable: What Now?	Преобразование контраста	2,58	-
Предложенный метод на основе генетических алгоритмов	CLAFE	218,81	16,02
	Гамма-коррекция + unsharp mask	217,57	-
	Гамма-коррекция	194,89	-
	Unsharp mask	93,40	-
	Tonemap Drago	218,63	17,62
	Выравнивание гистограммы	215,82	-

Таблица 6 — Сравнение различных методов атаки на метод VMAF с помощью вариантов предложенного метода с различными методами обработки видео.

взяты из обучающей и валидационной части набора данных Pascal VOC; соответственно, эти множества изображений не пересекаются, и не происходит утечки информации. В качестве дополнительного набора данных был использован Vimeo 90К [106], содержащий 91,701 видео длиной 7 кадров. Из него случайным образом был выбран набор, состоящий из 5000 видео в обучающей и 1000 видео в валидационной выборке.

В качестве тестового набора видео использовался набор Derf's collection [102], который содержит видео длиной 60 кадров. Для этого видео с и без обученной обработки были закодированы видеокодеком h264 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 [6]. Дан-

ный метод оценивает две RD-кривые. Для фиксированного видео V первая RD-кривая получается кодированием видео V с набором битрейтов. Для каждой закодированной версии видео вычисляется VMAF и фактический битрейт. Вторая RD-кривая получается обработкой видео V с последующим кодированием с набором битрейтов. Для каждой закодированной версии видео вычисляется VMAF и фактический битрейт.

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 сжатия на 19.07 и 17.03 единицы в среднем на двух наборах данных.

Визуальная оценка

В большинстве случаев, когда значения оценки качества видео алгоритмом VMAF увеличивались более чем на 50%, визуальное качество значительно снижалось. Это проиллюстрировано на рисунке 311. Видео, обработанные предложенным методом для повышения оценок VMAF NEG не содержат видимых искажений (рисунок 312).



Рисунок 311 — Сравнение кадров исходного обработанного видео с помощью метода CLANE. VMAF увеличился на 181,22 на левом видео, на 147,30 на среднем видео, на 114,75 на правом видео.

Чтобы детальнее проанализировать, как предварительная обработка видео, подобранная с помощью генетических алгоритмов, влияет на качество воспри-



Рисунок 312 — Примеры кадров с различными методами предварительной обработки, настроенными на увеличение VMAF NEG. Повышение оценок метода рассчитывается как разница между значением VMAF NEG для атакующего перед сжатием видео и сжатого видео без атаки.

ятия, было проведено краудсорсинговое визуальное сравнение. Для этого было использовано шесть изображений с различной временной, пространственной сложностью и яркостью. Каждое изображение было обработано протестированными методами (сочетание гамма-коррекции и unsharp mask, метод тонового картирования Драго, CLAHE, гамма-коррекция и unsharp mask) с подобранными параметрами, увеличивающими оценку VMAF. С помощью проведения попарных сравнений на краудсорсинговом сервисе Subjectify.us были получены оценки сравнения от 114 участников. Участники должны были просмотреть пары изображений, обработанные разными методами, а затем выбрать из каждой пары изображение с наилучшим визуальным качеством. Голоса тех, кто не смог правильно ответить на проверочные вопросы (которые содержали очевидные различия в качестве), были исключены. Общее число полученных ответов составило 1 584. На рисунке 313 показаны результаты сравнения.

Сравнение показало, что в 58,3% случаев оригинальное изображение превосходит остальные по визуальному качеству, то есть в большинстве случаев пары были визуальными неразличимы. Однако во всех изображениях оценки метода VMAF были повышены с помощью предложенного метода состязательной атаки в среднем на 46%. Комбинация обработки видео с гамма-коррекцией и нерезким маскированием показало наихудшее визуальное качество.

При проведении визуального сравнения видео, обработанного с применением предложенного метода на основе аппроксимации, оказалось, что модели,

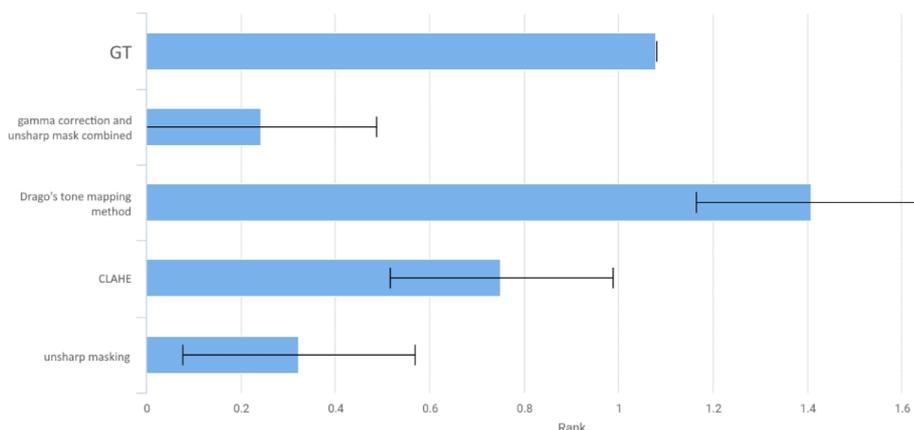


Рисунок 313 — Результаты визуальной оценки предложенных методов с помощью краудсорсингового подхода.

обученные на наборе данных Pascal VOC*, оставляют на изображениях или видео искажения нескольких типов. Наиболее заметным является добавление фиолетовых областей в темных фрагментах изображений 314. Так как VMAF использует для оценки исключительно яркостную компоненту входного видео, отбрасывая цветные компоненты целиком, данные искажения остаются незаметными для ее аппроксимации.

Так как наличие фиолетовых областей является наиболее заметным искажением и метод SSIM является дифференцируемым по входным параметрам, использование SSIM в качестве компоненты регуляризации помогло улучшить визуальное качество атакованных видео, которые можно видеть на рисунке 314 и рисунке 315. Модель с использованием SSIM в качестве регуляризации не производит цветных пятен и ложных цветных границ, однако проводит визуально заметную цветовую коррекцию и повышает резкость кадров видео.

3.5 Открытое онлайн сравнение

С помощью предложенного метода состязательной атаки универсальным возмущением и других атак на алгоритмы оценки качества изображений и видео было проведено расширенное исследование устойчивости современных методов оценки качества. Оно включало в себя анализ уязвимости 15 нереперенсных методов к 7 состязательным атакам. Результаты сравнения устойчивости алгоритмов оценки качества изображений и видео были опубликованы онлайн <https://>

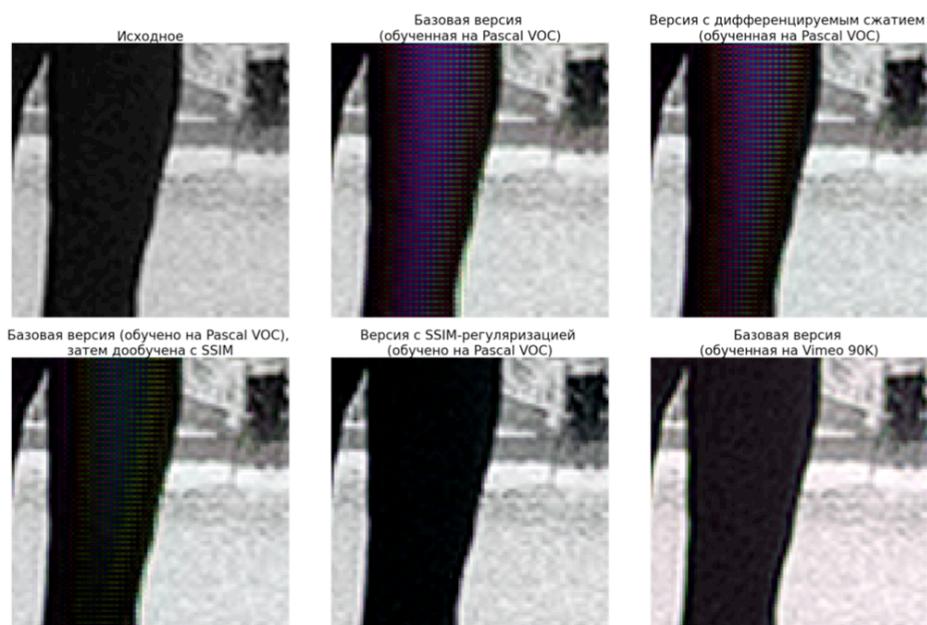


Рисунок 314 — Визуальные артефакты: фиолетовые области.

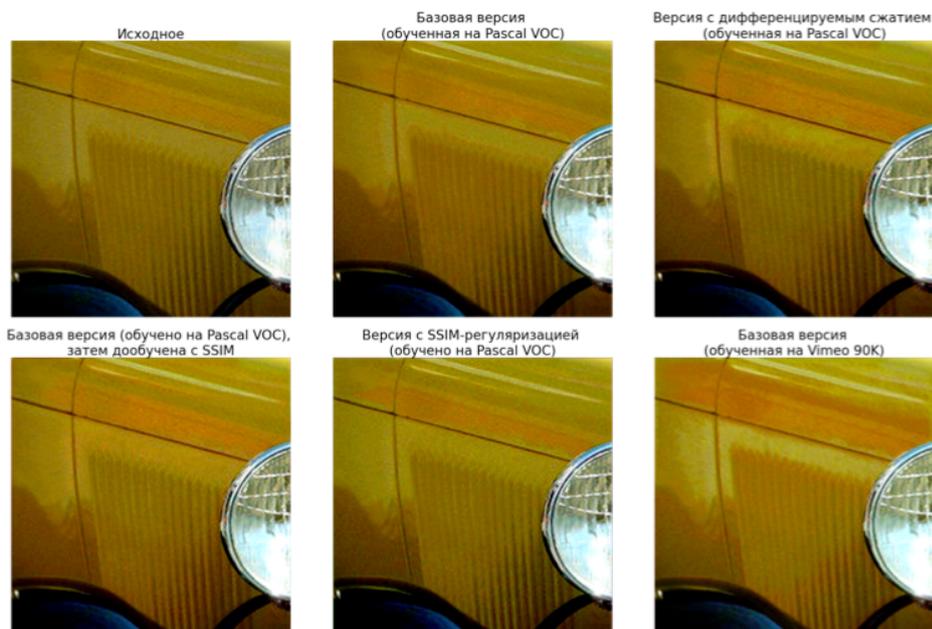


Рисунок 315 — Изменение границ, производимых различными вариантами предложенного метода.

[//videoprocessing.ai/benchmarks/metrics-robustness.html](http://videoprocessing.ai/benchmarks/metrics-robustness.html).

Сайт содержит интерактивные графики и таблицы, а также инструкцию для авторов, желающих добавить свой метод на страницу сравнения.

3.6 Выводы

В данной главе предложены подходы для состязательных атак на алгоритмы оценки качества видео. Проведенное исследование показало, что современные неререференсные алгоритмы оценки качества уязвимы к атакам методом универсального возмущения. В проведенных экспериментах атака методом универсального возмущения работает значительно быстрее, чем атака методом градиента MADC [83]. В сравнении с градиентной атакой, предложенный метод на основе обучения универсальных возмущений обеспечивает 70% повышения значения атакуемого метода всего за 3% времени обработки. В рамках данной главы предложен метод тестирования устойчивости неререференсных методов качества и проведено сравнение устойчивости 15 методов.

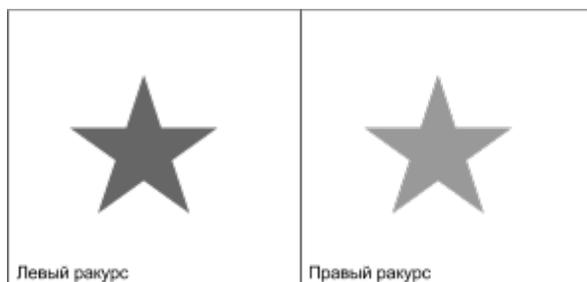
Так как метод на основе универсального возмущения применим только к дифференцируемым алгоритмам оценки качества видео, в данной главе предложены методы атак на недифференцируемые алгоритмы на примере атаки на алгоритм VMAF. Предложенный метод оптимизации параметров обработки видео, использующий генетический алгоритм, обеспечивает более сильную атаку, чем существующие методы атаки на VMAF. Среди протестированных методов обработки видео метод CLANE показал наибольший прирост оценок VMAF (63,17%), нерезкое маскирование в сочетании с гамма-коррекцией увеличили более устойчивую версию метода VMAF от NETFLIX под названием VMAF NEG в среднем на 5,93%. Однако данный метод вычислительно более сложный, чем другой предложенный метод на основе создания аппроксимации. Он обеспечивает повышение VMAF в среднем на 18-20%. Полученные результаты подтверждают, что методы VMAF, VMAF NEG, а также неререференсных методов не должны быть единственными критериями измерения качества видео в сравнениях, поскольку разработчики видеокодеков и других алгоритмов обработки видео могут их увеличить с помощью атак.

Глава 4. Метод оценки качества стереоскопического видео

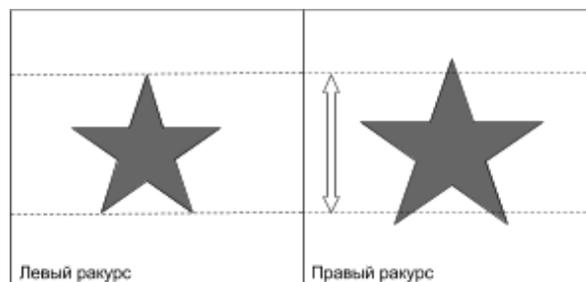
4.1 Постановка задачи

Поставленная задача формулируется следующим образом: по значениям стереоскопических искажений необходимо дать оценку качества стереоскопического видео, отражающую качество стереофильма. Метод на вход должен принимать значения $\{X_i^k\}_{i=1,N;k=1,S}$ стереоскопических искажений и других признаков стереофильма, а на выходе выдавать оценку качества для сцен стереофильма $\{Y_k\}_{k=1,S}$, где N — количество используемых показателей (искажений и признаков видео), S — количество сцен фильма.

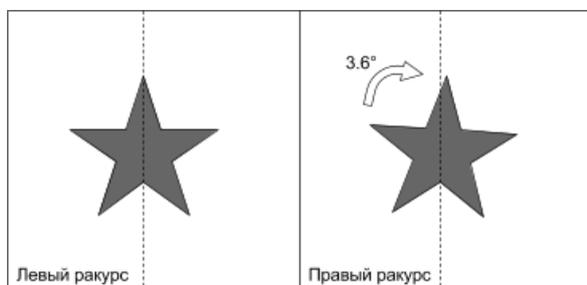
В работе рассматриваются следующие типы стереоскопических искажений: геометрические (поворот одного из ракурсов относительно другого, несоответствие масштаба ракурсов), несоответствие ракурсов по цвету, временной сдвиг между ракурсами (рисунок 41). Однако предлагаемый метод может быть применен для оценки других стереоскопических искажений.



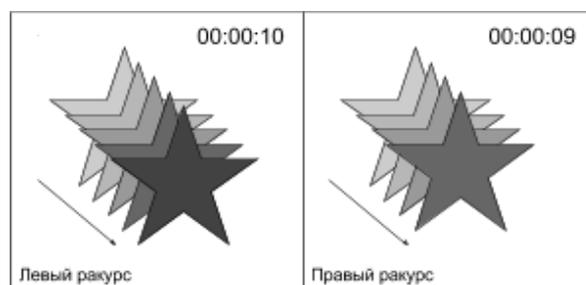
(а) Несоответствие цвета



(б) Несоответствие масштаба



(в) Поворот одного из ракурсов



(г) Временной сдвиг

Рисунок 41 — Схематическая визуализация стереоскопических искажений.

4.2 Обзор существующих методов

Задаче оценки качества стереоскопического видео посвящено множество исследований, однако лишь в небольшом количестве работ были проведены эксперименты для анализа искажений, характерных для фильмов, произведенных методом стереокиносъемки. Например, в работе [107] рассматривается влияние геометрических несоответствий в стереопаре на качество видео, но эксперименты проводились только с использованием стереоизображений. В работе [108] рассматривается 22 типа искажений, в том числе, геометрические и цветовые. Однако авторы не исследовали сдвиг во времени между ракурсами, а характер исследованных в эксперименте геометрических искажений редко встречается при съемке художественных стереофильмов.

В большинстве существующих работ для получения от зрителей ответов для эталонной разметки уровня дискомфорта использовались электронные или бумажные анкеты. Существуют методы, позволяющие участникам непрерывно сообщать об испытываемом дискомфорте, например, в работе [108] используется программный интерфейс, с помощью которого участники могут перемещать слайдеры в процессе просмотра. В работе [109] предлагается улучшение данного подхода, например, добавление напоминаний зрителю, в случае, если он длительное время не перемещал слайдер. Преимуществом данных методов является возможность получения поккадровых оценок дискомфорта, однако, они имеют низкую корреляцию (в среднем 0,2) с оценками, полученными после просмотра с помощью анкет. Также, существуют исследования, в которых используются другие методы для получения эталонных оценок дискомфорта; например, запись электроэнцефалограммы или движения глаз. Вэй Чен провел масштабное исследование [110], посвященное анализу влияния различных стереоскопических искажений на визуальное качество и дискомфорт зрителей, с использованием анкетирования и анализа электроэнцефалограммы. В его работе, как и во многих других (например, [111; 112]), были проведены эксперименты по сравнению восприятия зрителем обычной и стереоскопической версий видео.

В таблице 7 представлены параметры проведенных в мире экспериментов в рамках исследования причин дискомфорта при просмотре стереоскопического видео. В большинстве работ участникам демонстрируются либо синтетические видеопоследовательности, либо видео, снятые на непрофессиональные стерео-

видеокамеры. Лишь в нескольких исследованиях эксперименты проводились с использованием фрагментов из реальных фильмов, однако, в них оценивалось только общее состояние участников до и после просмотра видео. В последней строке представлена информация об экспериментах, проведенных в рамках данной работы.

Таблица 7 — Эксперименты, проведенные в рамках исследований причин появления дискомфорта у зрителей при просмотре стереоскопического видео.

Университет	Кол-во участников	Длит.	Год	Метод показа 3D	Тип искажений
Catholic University of Korea [113]	20	18 мин	2011	Пассивная	Снятые S3D послед., разл. параллакс и движение
Telecom Innovation Labs [114]	24	50 мин	2011	Активная	База S3D видео, разл. параллакс
Philips Research Labs [115]	24	24 мин	2011	Авто-стереоскопич.	Стереofilm (конвертир.), разл. параллакс
Beijing Institute of Ophthal. [116]	30	30 мин	2012	Активная, пассивная	Стереofilm (снятый)
LUNAM University [117]	29	28 мин	2012	Активная	База S3D видео, 10 типов деградаций
Tampere University of Tech. [118]	10	45 мин	2013	Пассивная	Снятые S3D послед., разл. кодирование
LUNAM and Yonsei U., Acreo Inst. [119]	29, 28, 48	29 мин	2013	Активная, пассивная	База S3D видео, 10 типов деградаций
Yonsei University [109]	56	10,5 мин	2014	Авто-стереоскопич.	База S3D видео, разл. SI/TI
Roma Tre University [120]	854	~90 мин	2015	Пассивная	Опрос после просмотра стереofilmов в кинотеатрах
University College Dublin [121]	4 для кажд. видео	5.5 мин	2015	Активная	База S3D видео, имитирование потерь при трансляции
University of British Columbia [122]	88	35 мин	2016	Пассивная	База S3D видео, разл. SI/TI, кодирование
University North [108]	18	8 мин	2016	Активная	База S3D видео, 22 типа деградаций (кодир., размер, геометрич. и др.)
Предложенный	302	40 мин	2017	Пассивная	Сцены из снятых стереofilmов, 4 типа искажений, 5 уровней

4.3 Разработка метода оценки качества стереоскопических видео

На рисунке 42 представлена схема предложенного подхода для разработки метода оценки качества стереоскопического видео. Данный подход может использоваться для разработки иных специализированных методов оценки качества видео, для которых доступно малое количество обучающих данных.

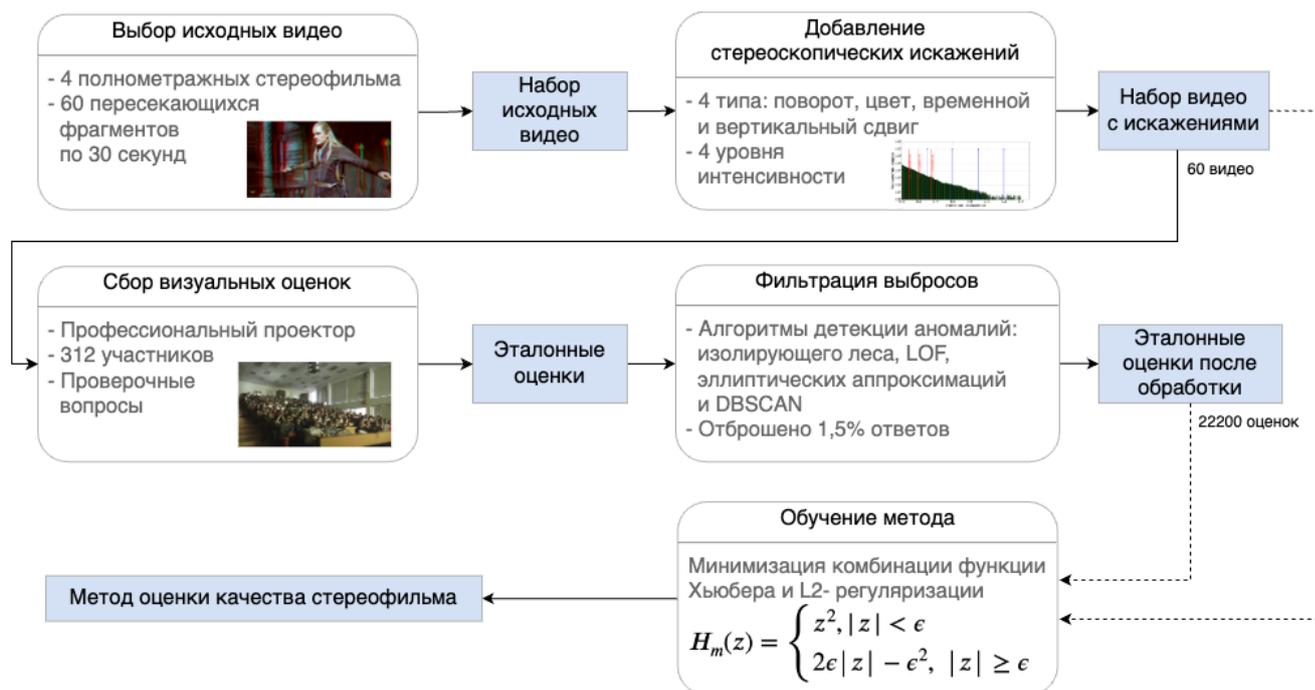


Рисунок 42 — Схема разработки предложенного метода оценки качества стереоскопического видео.

4.3.1 Выбор признаков

В качестве показателей, отражающих искажения в стереоскопического видео, предлагается использовать следующие признаки:

- Среднее значение и отклонение яркости для левого и правого ракурсов
- Средняя длина вектора движения
- Средняя длина проекции вектора движения на вертикальную ось
- Средняя длина вектора диспаратности
- Значения положительной и отрицательной диспаратности

- Значения геометрических искажений между правым и левым ракурсами
 - масштаба, поворота, вертикального сдвига, искажения перспективы
- Величина несоответствия по цвету между ракурсами
- Величина несоответствия по резкости между ракурсами
- Величина временного сдвига между ракурсами
- Вероятность перепутанных местами правого и левого ракурсов
- Также некоторые из перечисленных алгоритмов помимо величины искажения выдают доверие к вычисленному значению. Поэтому дополнительно были получены следующие показатели:
 - Достоверность вычисленных значений геометрических искажений
 - Достоверность вычисленных значений искажения перспективы
 - Достоверность вычисленных значений несоответствия по цвету
 - Достоверность вычисленных значений диспаратности
 - Достоверность вычисленных значений несоответствия по резкости
 - Достоверность вычисленных значений сдвига во времени между ракурсами

Все значения, за исключением величины временного сдвига, вычисляются для каждого кадра фильма, временной сдвиг и достоверность значения временного сдвига вычисляются для сцен фильма.

Так как некоторые признаки вычисляются для левого правого ракурсов стереокадра, их значения коррелированы, как и значения многих показателей достоверности результатов работы алгоритмов, которые часто основаны на точности компенсации левого и правого ракурсов при построении карты диспаратности. В связи с коррелированностью признаков, использование всех 23 показателей избыточно и может снизить точность работы метода. Выбор k признаков в предложенном методе проводится голосованием с помощью метода оценки взаимной информации и оценки лучшей точности при обучении ансамбля деревьев решений на разном количестве признаков.

Рассматривая признаки видео как случайные величины, их взаимную информацию можно определить как функцию, описывающую количество информации, содержащееся в одной случайной величине относительно другой. Пусть X, Y — признаки стереоскопического видео, их совместное распределение $P_{(X,Y)}$, а маргинальные распределения каждого признака — P_X, P_Y . Взаимная информация $I(X; Y)$ определяется через расстояние Кульбака-Лейблера

$D_{KL}(P_{(X,Y)} || P_X \times P_Y)$. Признаки, имеющие наименьшую взаимную информацию, используются для обучения метода.

Для выбора признаков на основе обучения ансамбля деревьев решений, используется модель на основе градиентного бустинга. Она обучается с меняющимся набором и количеством признаков, и модель с признаками, показавшая лучшие результаты на кросс-валидации, определяет признаки, которые следует использовать в дальнейшем для разработки метода.

4.3.2 Исключение аномалий в данных

Так как на оценки качества влияет не только субъективное предпочтение эксперта, но и условия просмотра, острота зрения и другие факторы, необходимо проводить процедуру исключения аномалий для полученных эталонных оценок. Из полученных ответов исключаются ответы участников, которые сильно отличаются от остальных. Для определения таких респондентов предлагается применить следующие методы:

- метод эллипсоидальной аппроксимации [123];
- метод изолирующего леса [124];
- метод обнаружения локальных аномалий [125];
- кластеризация с помощью метода DBScan [126], использующего информацию о плотности данных.

Первые три метода после обучения на входных данных вычисляют значения аномальности наблюдаемых точек. Чем это значение меньше, тем более вероятно, что наблюдение является выбросом в имеющихся данных. Четвертый метод присваивает метки выделенных кластеров и метку «-1» для остальных точек. Кластеризация проводилась на основе следующих признаков: среднего значения и отклонения оценок, выданных экспертом за все время эксперимента (mean opinion score, MOS) и разница между оценкой участника для видео без искажения и для видео с искажением (differential mean opinion score, DMOS). Последний признак характеризует оценку качества для сцен с искажениями, нормированными на исходную сцену без добавления искажения, и вычисляется для похожих видеофрагментов, которые были взяты из одной сцены стереофильма.

Исключение аномальных ответов было проведено отдельно для групп зрителей, посмотревших продемонстрированное видео с прямым и обратным порядком следования сцен.

Выбор способа исключения аномалий сделан при помощи сравнения среднего значения доверительного интервала оценок дискомфорта в сценах. По результатам сравнения (таблица 8), наибольшее уменьшение величины доверительного интервала для оценки качества сцен было обеспечено алгоритмом изолирующего леса. Он основан на разбиении пространства точек на области, при построении которого аномальные точки становятся листьями дерева разбиения, имеющими наименьшую глубину. С помощью метода изолирующих деревьев были исключены ответы 5% участников (6 человек), посмотревших тестовое видео с прямым порядком следования сцен, и также 5% участников (13 человек), посмотревших видео с обратным порядком следования сцен.

Таблица 8 — Сравнение среднего значения доверительных интервалов ответов участников.

	Без исключения аномалий	Elliptic Envelope (EE)	Isolation Forest (IF)	Local Outlier Factor (LOF)	EE+IF+ +LOF	DBScan
Средняя величина доверительного интервала	0,0948	0,0942	0,0933	0,0934	0,0939	0,0959

4.3.3 Описание предложенного метода

При решении задачи линейной регрессии:

$$Y_k = \alpha + \sum_{i=1}^N X_i^k \omega_i + \varepsilon,$$

где α , ω — параметры регрессии, определяемые при решении задачи, ε — аддитивная случайная величина, имеющая нормальное распределение с нулевым средним и дисперсией σ^2 , N — количество признаков, с помощью метода наименьших квадратов [127], решается задача оптимизации для обучающей выборки

размера S

$$\min_{\omega} \sum_{k=1}^S (X^k \omega - Y_k)^2.$$

Функция Хьюбера [128] $H_m(z)$ накладывает меньший штраф на точки с большим отклонением, и при решении задачи регрессии производится оптимизация по параметрам ω (коэффициенты вклада признаков) и σ (масштабирующий параметр, позволяющий задавать пороговое значение ε функции потерь Хьюбера):

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

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

Для снижения переобучения дополнительно предлагается использовать L2-регуляризацию, накладывающую ограничение на параметры. Итоговая предлагаемая оптимизируемая функция потерь принимает следующий вид:

$$\min_{\omega, \sigma} \sum_{k=1}^S \left(\sigma + H_m \left(\frac{X^k \omega - Y_k}{\sigma} \right) \sigma \right)^2 + \alpha \|\omega\|_2^2, \text{ где } H_m - \text{функция Хьюбера.}$$

Оценка параметров ω производится с помощью метода наименьших квадратов. Для решения поставленной задачи регрессии были применены методы машинного обучения. Разработка метода оценки качества стереовидео проводилась с использованием данных, полученных в результате проведенных экспериментов. Так как во время эксперимента зрители давали оценку качества для каждой короткой сцены тестового стереофильма, то для покадровых значений признаков стереоскопических искажений были вычислены средние значения по сценам. Для обучения метода разбиение полученного набора данных проводилось с учетом распределения всех оценок на группы по процентиллям. Это было необходимо для попадания более редких оценок в кросс-валидационные подвыборки.

$$\text{label}(Y_i) = \begin{cases} 0, & 0 \leq Y_i < \text{percentile}(Y, 33); \\ 1, & \text{percentile}(Y, 33) \leq Y_i < \text{percentile}(Y, 66); \\ 2, & Y_i \geq \text{percentile}(Y, 66) \end{cases},$$

где Y_i — эталонная оценка качества сцены i , $\text{percentile}(Y, k)$ — k -й перцентиль полученных оценок Y .

4.3.4 Формулы построенных моделей

Для оценки качества стереофильмов предложенным методом используются следующие признаки видео:

- *GC* — значение доверия (от 0 до 255) к вычисленным значениям геометрических искажений
- *NP* — значение отрицательной диспаратности
- *PP* — значение положительной диспаратности
- *RD* — значение поворота в градусах одного из ракурсов относительно другого
- *ScD* — значение несоответствия масштаба ракурсов как доля от ширины кадра
- *SM* — значение несоответствия резкости ракурсов как параметр
- *TS* — значение временного сдвига одного из ракурсов относительно другого в кадрах

Оценка качества стереофильма выражается через перечисленные признаки по следующей формуле:

$$Y = 1,084 - 0,003 \cdot GC + 0,145 \cdot NP - 0,232 \cdot PP + \\ + 0,314 \cdot RD + 9,844 \cdot ScD + 0,076 \cdot SM + 1,938 \cdot TS$$

4.4 Экспериментальная оценка

4.4.1 Подготовка набора данных

Для получения эталонных оценок качества стереоскопического видео с искажениями стереосъемки была проведена серия субъективных экспериментов, в которой участникам предлагалось оценивать уровень дискомфорта во время просмотра специально подготовленного стереоскопического видео. Постановка задания заключалась в оценке дискомфорта, а не естественности демонстрируемого видео, так как в области качества стереоскопического видео эти понятия

связаны. При этом, возникновение дискомфорта оказывает более негативное влияние на качество восприятия видео зрителем.

Используемые стереоскопические видео и оборудование

Подготовленное для экспериментов видео состояло из 60 тридцатисекундных видеофрагментов из художественных стереофильмов «Сталинград» (2013), «47 ронинов» (2013), «Прометей» (2012), «Хоббит: Нежданное путешествие» (2012). Были выбраны сцены с различной интенсивностью движения, глубиной и яркостью (рисунок 43). Выбранные видеофрагменты из каждого фильма частично пересекались, то есть изначально из каждого фильма было выбрано по 3 длинных сцены, которые далее были разделены на 30-секундные фрагменты с пересечением. Этот подход применен для того, чтобы не показывать зрителям несколько раз один и тот же видеофрагмент во избежание потери внимания.

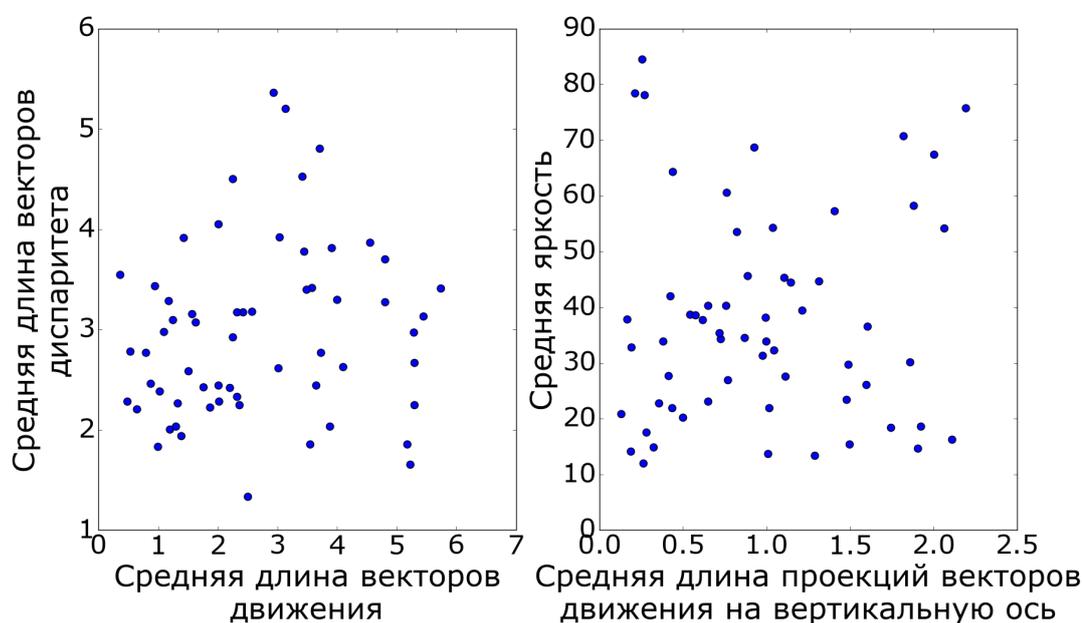


Рисунок 43 — Временная и пространственная сложность выбранных видеофрагментов из фильмов. Левый график показывает зависимость средней длины вектора движения между кадрами в пикселях от средней длины вектора диспаратности в процентах от ширины кадра. Правый график отражает зависимость между средней длиной проекции вектора движения на вертикальную ось в пикселях и средней яркостью кадра по шкале от 0 до 255.

Видеофрагменты были модифицированы путем добавления одного из 4 исследуемых типов искажений: поворот одного из ракурсов относительно другого, несоответствие масштаба ракурсов, цветовое несоответствие и временное несоответствие между ракурсами, 20 видеофрагментов содержали комбинацию двух различных искажений. Каждое искажение присутствовало с одним из 5 уровней интенсивности от 0 до 4, где 0 — отсутствие искажения, 4 — сильное искажение. Уровни интенсивности были выбраны в соответствии с распределением значений данного искажения в художественных фильмах, но во всех случаях имели достаточную величину для того, чтобы искажение можно было заметить при просмотре коротких видеофрагментов.

Общая длительность составленного тестового стереофильма составила 40 минут.

Получение эталонных оценок качества

В экспериментах приняли участие 302 человека, из них 68 участников — дважды, посмотрев стереофильм с прямым и обратным порядком следования видеофрагментов. Таким образом, было получено 370 ответов для каждой сцены, и всего — 22200 оценок от зрителей. В основном в эксперименте приняли участие студенты факультета ВМК МГУ, среди которых 30% женского пола, возраст респондентов от 18 до 42, средний возраст — 20 лет. 31% участников отметили, что в кинотеатрах предпочитают не смотреть стереофильмы, 59% смотрят и 3D и 2D фильмы, и лишь 10% ответили, что предпочитают смотреть 3D-версии фильмов.

Во время просмотра участникам необходимо было отмечать уровень дискомфорта после просмотра каждого 30-секундного видеофрагмента. Для записи ответов использовалась электронная анкета, в которой уровень дискомфорта предлагалось оценить по дискретной шкале 0 до 4, где 0 — отсутствие дискомфорта, 4 — сильный дискомфорт. В начале эксперимента участникам были продемонстрированы короткие примеры видеофрагментов без искажения и с сильным искажением, вызывающим дискомфорт. Для уменьшения влияния накопления дискомфорта от просмотра предыдущих видеофрагментов на полученную среднюю оценку, в части экспериментов подготовленный стереофильм был

продемонстрирован с обратным порядком следования видеофрагментов, сохраняющим порядок сцен, взятых из одного фильма.

Эксперимент проводился в аудитории П-13 факультета ВМК МГУ, оборудованной двумя профессиональными проекторами Digital Projection Titan 1080p-700 с линейной поляризацией, разрешение каждого проектора 1920×1080 , экран с серебряным покрытием диагональю 9 метров. На рисунке 44 изображена фотография участников во время одного из экспериментов.



Рисунок 44 — Процесс проведения эксперимента (яркость изображения искусственно повышена). Аудитория П-13 факультета ВМК МГУ.

Анализ результатов эксперимента

До и после просмотра всего стереофильма участникам предлагалось оценить свое состояние и отметить наличие негативных симптомов. В таблице 9 приведено количество ответов участников о наличии симптомов. Можно заметить, что вдвое меньше участников имели нормальное самочувствие после просмотра тестового видео, заметный процент участников отметили появление головной боли и боли глаз.

На рисунке 45 показано распределение среднего уровня оценки дискомфорта участников за все время эксперимента. На графике видно, что лишь небольшая

Таблица 9 — Количество участников, отметивших наличие негативных симптомов до и после эксперимента.

	Нормальное самочувствие	Сонливость	Усталость	Головная боль	Боль глаз	Другие симптомы
До просмотра	203 (70%)	70 (24%)	51 (18%)	8 (3%)	11 (4%)	8 (3%)
После просмотра	114 (40%)	74 (26%)	112 (39%)	45 (16%)	94 (33%)	17 (6%)

часть людей ставила высокие и низкие оценки уровню дискомфорта, остальные участники в среднем оценивали уровень дискомфорта от 1 до 1,5. Привлечение большого количества участников способствовало получению данных о дискомфорте от зрителей с различной остротой стереоскопического зрения.

На рисунке 46 показано распределение оценок дискомфорта для показанных видео. На графике видно, что несмотря на наличие в тестовом видео равного количество сцен с каждой из 4 степеней искажений, оценки дискомфорта для сцен неравномерны — лишь 20% сцен привело к появлению среднего или сильного дискомфорта при просмотре.

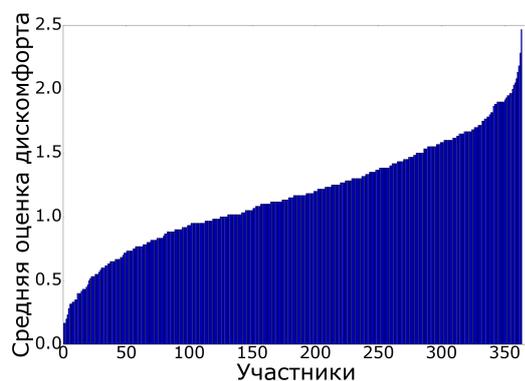


Рисунок 45 — Упорядоченные по возрастанию усредненные оценки дискомфорта участников.

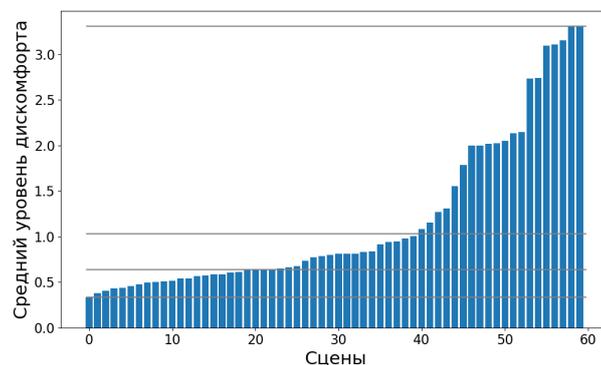


Рисунок 46 — Средние значения оценок дискомфорта для сцен. Линиями отмечены 30%-перцентили.

Ответы, полученные при просмотре стереоскопического видео с прямым и обратным порядком следования сцен, были объединены, и эталонная оценка качества вычислена как среднее значение ответов зрителей для каждой сцены, взвешенное на количество посмотревших каждый из порядков следования сцен участников. Для каждой сцены k вычислены доверительные интервалы для оцен-

ки качества μ , где $\alpha = 0,05$:

$$P \left(\bar{x}_k - t_{1-\alpha, N-1} \frac{\sqrt{\frac{\sum_{i=1}^N (x_k^i - \bar{x}_k)^2}{N-1}}}{\sqrt{N}} \leq \mu \leq \bar{x}_k + t_{1-\alpha, N-1} \frac{\sqrt{\frac{\sum_{i=1}^N (x_k^i - \bar{x}_k)^2}{N-1}}}{\sqrt{N}} \right) = 1 - \alpha,$$

где x_k^i — ответ i -го участника для сцены с номером k , \bar{x}_k — средний зарегистрированный уровень дискомфорта среди всех участников на сцене k , $t_{1-\alpha, N-1}$ — α -квантиль распределения Стьюдента с $N - 1$ степенями свободы, N — количество участников. Полученные доверительные интервалы c_k и средние значения дискомфорта \bar{z}_k отображены на рисунке 47, где каждый столбец соответствует оценке качества данной сцены, сцены располагаются в том порядке, в котором они следовали в продемонстрированном стереоскопическом видео. Сцены, отмеченные красным цветом, были наиболее болезненны для зрителей, зеленый цвет столбцов соответствует высокому визуальному качеству и отсутствию дискомфорта. Наиболее болезненными были видео с добавленным временным несоответствием между ракурсами, а также видео, в которых присутствовало два различных искажения.

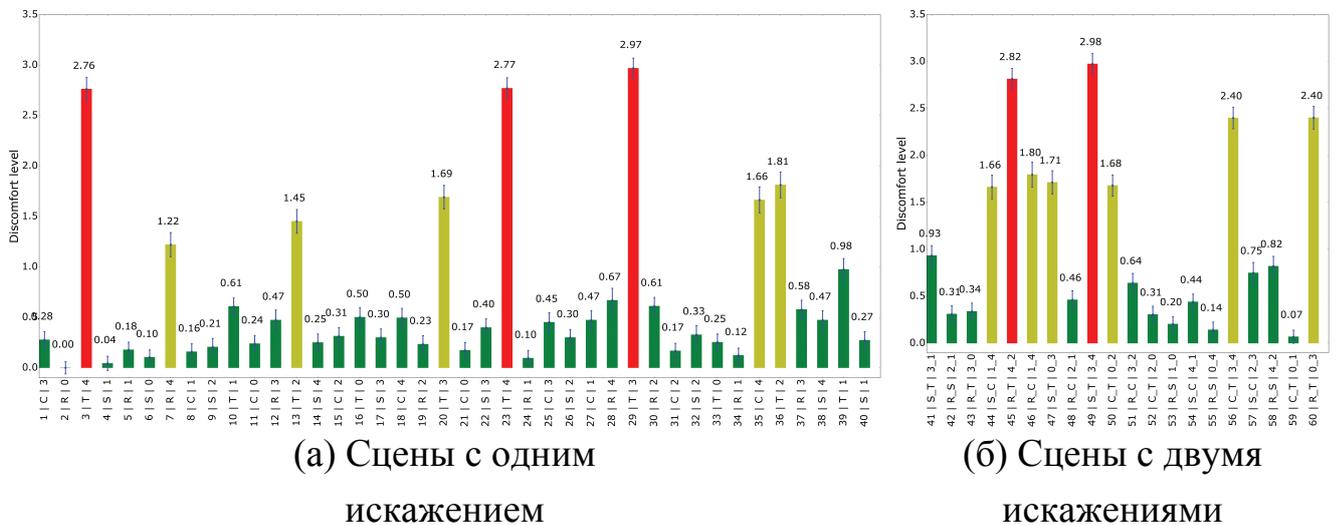


Рисунок 47 — Оценка качества для сцен с различными типами искажениями и их интенсивностью. Здесь С обозначает несоответствие цвета ракурсов, R — поворот одного из ракурсов, T — временной сдвиг и S — несоответствие масштаба ракурсов. Интенсивность искажения отмечена цифрами от 0 до 4, где 0 соответствует отсутствию искажения (была продемонстрирована исходная сцена).

Так как в эксперименте приняли участие больше трехсот человек, что значительно превышает количество участников в ранее проведенных исследованиях в данной области, это позволяет оценить необходимое количество участников

в подобных экспериментах. Для анализа были построены выборки из исходных данных, размером от 1 до общего количества полученных ответов для каждой сцены. Для каждого анализируемого количества респондентов было случайным образом сгенерировано 1000 bootstrap-наборов. На рисунке 48 представлены средние значения оценок дискомфорта в зависимости от количества респондентов для двух сцен, одна из которых не содержала искажения, а другая — содержала сильное искажение, а также 95%-доверительный интервал к полученным средним значениям оценок. На графике видно, что оценки с разной интенсивностью стремятся к среднему значению для всех респондентов, что означает разное необходимое количество респондентов для сцен различной сложности.

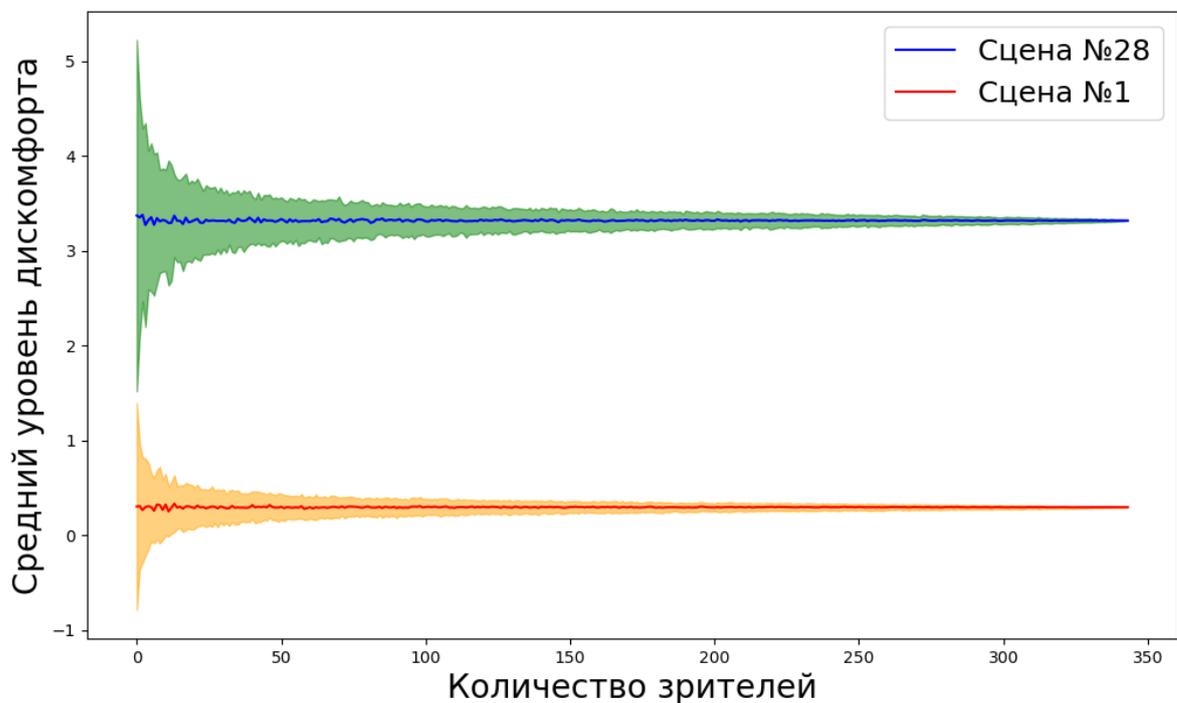


Рисунок 48 — Среднее значение оценки дискомфорта для двух сцен в bootstrap-выборках с разным количеством респондентов.

На рисунке 49 отражен размер 95% доверительного интервала в зависимости от количества респондентов. Таким образом, для оценки видео с различными уровнями искажения требуется разное количество респондентов, но в среднем это количество не должно быть меньше 200.

В таблице 10 указаны полученные оценки качества сцен с различными несоответствиями между ракурсами стереофильма. В верхней части таблицы представлены численные величины исследуемых несоответствий, в нижней части таблицы — средний уровень дискомфорта участников при просмотре сцен

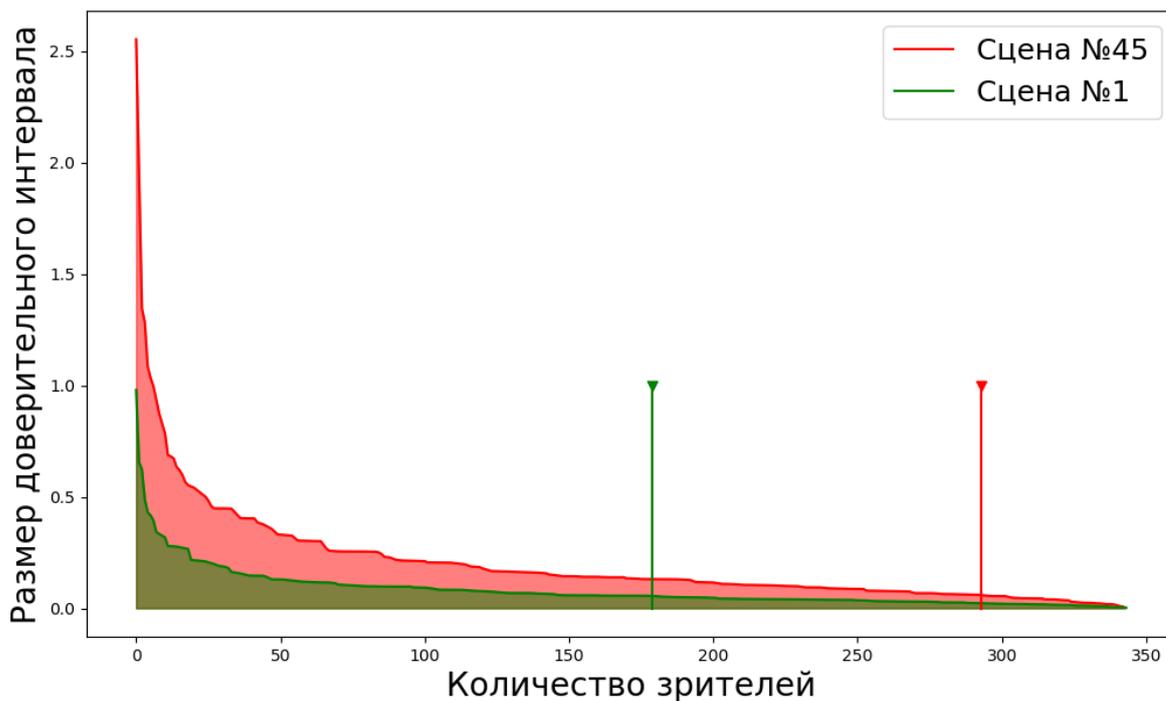


Рисунок 49 — Размер 95%-доверительного интервала к среднему значению дискомфорта в bootstrap-выборках с разным количеством респондентов.

с данным искажением. Наибольший уровень дискомфорта был отмечен на сценах с временным сдвигом между ракурсами. Сопоставимый дискомфорт зрители испытывали на сценах с геометрическими искажениями, однако при больших значениях несоответствия масштаба ракурсов уровень дискомфорта не увеличивался. Это объясняется достижением величины искажения, при которой большие значения несоответствия не отличимы зрителями. Для некоторых искажений при увеличении силы искажения уровень дискомфорта изменяется не монотонно. Такой эффект обусловлен наличием других факторов, повлиявших на ответ респондента, таких как физические параметры сцены, например, яркость или глубина. Таблица содержит информацию о перцентилях интенсивности искажений, добавленных к видео. В данном случае, например, 95 перцентиль означает, что 5% кадров полнометражных художественных стереофильмов содержат искажения с большей величиной несоответствия, и соответственно зритель при просмотре испытает такой же или больший дискомфорт.

Таблица 10 — Уровень дискомфорта зрителей при просмотре сцен с различными искажениями.

Величины искажений в тестовых видео				
Тип искажения	Интенсивность искажения			
	1	2	3	4
Цветовое искажение (MSU-3DColor-2017)	2,5	5	8	10
Процентиль	82,06	95,00	99,15	99,55
Поворот одного из ракурсов	0,3	0,6	0,9	1,2
Процентиль	98,38	99,85	99,98	99,99
Несоответствие по масштабу	100,5%	100,8%	101%	101,3%
Процентиль	93,34	97,87	98,87	99,55
Временной сдвиг	0,1 кадра	0,3 кадра	0,5 кадра	0,6 кадра
Процентиль	97,61	99,38	99,69	99,75
Среднее значение дискомфорта зрителей				
Тип искажения	Интенсивность искажения			
	1	2	3	4
Цветовое искажение	0,65±0.03	0,58±0.03	0,70±0.03	1,42±0.05
Поворот одного из ракурсов	0,47±0.01	0,75±0.03	0,86±0.03	1,28±0.04
Несоответствие по масштабу	0,49±0.03	0,61±0.02	0,69±0.03	0,70±0.03
Временной сдвиг	1,13±0.04	1,97±0.04	2,67±0.05	3,10±0.04

Сравнение предложенного метода с аналогами

В таблице 11 представлены результаты сравнения предложенного метода с другими методами и в комбинации с параметрами метода выбора признаков.

Лучший результат показала предложенная модель, основанная на решении задачи линейной регрессии с функцией потерь Хьюбера и L2-регуляризацией. Сравнение проведено на обучающей выборке из 40 сцен, выбранных из предложенного набора данных.

Таблица 11 — Точность работы различных моделей для оценки качества стереоскопического видео при обучении и валидации на полученном наборе данных.

Название модели	Коэффициент детерминации при обучении	Коэффициент детерминации при валидации	Выбор признаков
Регрессия с функцией потерь Хьюбера и L2-регуляризацией (предложенный)	$0,925 \pm 0,008$	$0,846 \pm 0,066$	10 лучших признаков по взаимной информации
Ридж-регрессия	$0,936 \pm 0,008$	$0,842 \pm 0,073$	12 лучших признаков по взаимной информации
Регрессия с оценочной функцией Тейла-Сена [129; 130]	$0,917 \pm 0,033$	$0,783 \pm 0,105$	10 лучших признаков по взаимной информации
Линейная регрессия	$0,945 \pm 0,009$	$0,771 \pm 0,139$	10 признаков на основе градиентного бустинга над деревьями
Лассо	$0,844 \pm 0,019$	$0,767 \pm 0,159$	12 признаков на основе градиентного бустинга над деревьями
Регрессия с L1- и L2-регуляризацией (ElasticNet)	$0,801 \pm 0,208$	$0,726 \pm 0,158$	8 признаков на основе градиентного бустинга над деревьями
Метод опорных векторов	$0,985 \pm 0,002$	$0,633 \pm 0,111$	8 лучших признаков по взаимной информации
Градиентный бустинг над деревьями решений	$0,989 \pm 0,001$	$0,517 \pm 0,427$	12 признаков на основе кластеризации
Метод ближайших соседей	$0,639 \pm 0,088$	$0,495 \pm 0,159$	4 лучших признака по взаимной информации

В таблице 12 представлены результаты оценки моделей Ридж-регрессии и линейной регрессии с функцией потерь Хьюбера и L2-регуляризацией на тестовой выборке из полученных данных экспериментов, отделенной в начале анализа моделей и содержащей треть продемонстрированных во время эксперимента сцен.

Таблица 12 — Точность работы методов с регуляризацией для оценки качества видео на тестовой выборке.

Метод	Коэфф. детерминации на обучающей выборке	Коэфф. детерминации на тестовой выборке	Среднеквадр. ошибка на тестовой выборке
Предложенный	0,921	0,881	0,091
Ридж-регрессия ($\alpha = 5$)	0,908	0,856	0,110

Таблица 13 содержит результаты сравнения предложенной модели и нейросетевой модели, опубликованной в работе [131] на других наборах данных. Большинство открытых наборов данных, состоящих из стереоскопических видео

Метод	Набор данных		
	3DVQA (30 видео, разная диспаратность)	NAMA3DS1-COSPAD1 (100 видео, разное кодир.)	Предложенный (60 видео, искажения стереокиносъемки)
Xu J. et al. [131] (с дообучением)	0,19	0,13	0,26
Предложенный (без дообучения)	0,61	-0,23	0,88

Таблица 13 — Сравнение различных методов повышения VMAF с помощью лучших полученных методов предварительной обработки видео.

и оценок визуального качества для них, нацелены на оценку качества кодирования и не содержат стереоскопические артефакты. Корреляция полученных оценок качества предложенной моделью на наиболее близком по тематике наборе данных 3DVQA [132], содержащем видео с разной диспаратностью, составила 0.61, что превосходит результаты метода [131].

Апробация предложенного метода для оценки качества полнометражных стереофильмов

Предложенная модель была применена для оценки 60 художественных стереофильмов. Проанализированные стереофильмы были произведены методом съемки и включают в себя различные жанры, бюджет, год выпуска. Для всех стереофильмов были вычислены покадровые значения технических данных качества и физических признаков. Фильмы были разбиты на сцены длиной 720 ± 200 кадров с учетом границ имеющихся сцен в фильме. Для каждой сцены были вычислены средние значения признаков и применены предложенные модели. Визуализация результатов сравнения представлены в приложении. Анализ показал, что большинство из потенциально болезненных для просмотра фильмов имеют низкий бюджет производства, однако среди них также присутствуют фильмы со средним бюджетом (например, «Сумасшедшая езда» (Drive Angry)). Стоит отметить, что среди фильмов низкого качества есть не только высокобюджетные фильмы, такие как «Жизнь Пи» (Life Of Pi), «Оз: Великий и Ужасный» (Oz The Great

and Powerful), «Джек — покоритель великанов» (Jack the Giant Slayer), но также фильмы со средним бюджетом производства, например, «Сталинград», «История дельфина» (Dolphin Tale).

При валидации полученных результатов было обнаружено, что в некоторых фильмах сцены, обладающие высоким качеством, были конвертированы из двумерного формата в трехмерный. Например, в фильмах «Кэти Перри: Частька меня» (Katy Perry: Part of Me) и «One Direction: Это мы» (One Direction: This is Us) большая часть сцен, содержащих записи концертов, были конвертированы из 2D в 3D. Из-за этого значения, вычисленные алгоритмами оценки технического качества, которые направлены на работу со снятым стереоконтентом, могли не соответствовать действительности. Появление конвертированных сцен в снятых стереофильмах происходит из-за вынужденного исправления испорченного во время съемки ракурса, или, как в случае с документальными фильмами, из-за необходимости создания дополнительных ракурсов из имеющихся одноракурсных видеозаписей.

4.5 Программная реализация

Программная реализация предложенного метода выполнена на языке Python в виде консольного приложения. Объем программной части, посвященной анализу и интерпретации результатов проведенных экспериментов, составил 2400 строк. Реализация метода построения модели и ее работы на новых данных насчитывает 2000 строк. Для визуализации полученных данных, а также результатов анализа фильмов, были созданы интерактивные графики, программный код для которых, написанный на HTML и jQuery, содержит 980 строк.

Видеопоследовательности, продемонстрированные зрителям во время эксперимента, содержали искажения, добавленные в исходные стереопары. Для преобразования ракурсов стереокадра использовался язык AviSynth [133], средствами которого можно осуществить необходимые преобразования: поворот любого из ракурсов на заданный угол; изменение масштаба ракурса; изменение частоты кадров, с помощью которого имитируется сдвиг во времени; добавление цветового искажения (с помощью модуля, разработанного видеогруппой лаборатории «Компьютерной графики и мультимедиа»). Для автоматической генерации

текста программ на языке AviSynth была написана программа на языке Python объемом 250 строк. Итоговые сгенерированные файлы с программами на языке AviSynth для создания видео суммарно содержат более 12000 строк.

Для вычисления значений анализируемых искажений, а также других показателей (таких, как яркость, глубина, скорость движения объектов в сцене) и доверия к ним, использовались алгоритмы, реализованные видеогруппой лаборатории «Компьютерной графики и мультимедиа» в программном комплексе VQMT3D [51]. Вычисление всех необходимых признаков для 60 стереофильмов заняло 3 недели счета в 96 потоков на 13 компьютерах. Вычисленные результаты в текстовом формате занимают 6,3 Гб.

4.6 Выводы

В данной главе предложен метод оценки качества стереоскопического видео, использующий технические данные качества стереофильма. Проведенная в процессе исследования серия экспериментов позволила получить уникальную базу оценок визуального дискомфорта 302 участников для 40-минутного стереоскопического видео. Из особенностей проведенного исследования можно выделить:

- Использование фрагментов из реальных снятых художественных стереофильмов с добавленными к ним стереоскопическими искажениями, характерными для процесса съемки, которые часто встречаются в полнометражных стереофильмах;
- привлечение большого количества участников эксперимента для получения эталонных оценок качества и уровня дискомфорта зрителей;
- использование объективных методов оценки качества стереоскопического видео;
- создание программной реализации, применимой на практике для анализа полнометражных стереофильмов.

В результате анализа полученных данных были выявлены зависимости между интенсивностью стереоскопических искажений и визуальным дискомфортом, возникающим при просмотре таких стереоскопических видео с искажениями. Среди исследованных искажений несоответствие во времени меж-

ду ракурсами вызывает наибольший дискомфорт при просмотре. На основе полученных данных был предложен метод оценки качества стереоскопического видео. Точность метода при тестировании составила 0,881. Предложенные методы были применены для анализа 60 стереофильмов, произведенных методом стереосъемки. Разработанный программный комплекс может быть востребован стереографами, как автоматическое средство контроля качества стереофильма, учитывающее восприятие стереоскопических искажений зрителями; также предложенный метод может быть использован для разработки стандартов производства стереоскопических видео и создания рекомендаций к просмотру стереофильмов.

Заключение

Основные результаты работы заключаются в следующем.

1. Разработана методика тестирования алгоритмов оценки качества видео, основанный на новом наборе видео. Предложенный набор состоит из 2486 видео, закодированных 47 видеокодеками 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 прогнозирования дискомфорта зрителей от просмотра стереофильма, основанный на оценке стереоскопических искажений. Экспериментальная оценка метода на новом наборе данных, содержащем 22200 оценок дискомфорта от зрителей, 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.

В заключение автор выражает благодарность и большую признательность научному руководителю Ватолину Д. С. за поддержку, помощь, обсуждение результатов и ценные советы. Также автор благодарит Шумицкую Е. А., Соловьева А. В., Гущина А. Е., Смирнова М. И., Синюкова М. В., Звездакова С. В., Паутова М. А. за содействие в получении результатов и коллег из лаборатории компьютерной графики и мультимедиа ВМК МГУ за обсуждение результатов. Автор также благодарит всех, кто сделал настоящую работу автора возможной.

Список литературы

1. *Antsiferova, A.* The influence of 3D video artifacts on discomfort of 302 viewers [Text] / A. Antsiferova, D. Vatolin // 2017 International Conference on 3D Immersion (IC3D). — IEEE. 2017. — P. 1—8.
2. *Анциферова, А.* Исследование влияния геометрических, цветовых и временных искажений стереоскопических видео на дискомфорт зрителей [Текст] / А. Анциферова, Д. Ватолин, С. Звездаков // 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.
3. *Анциферова, А.* Прогнозируемая оценка степени дискомфорта зрителей, вызываемого геометрическими, цветовыми и временными искажениями стереоскопического видео, для 60 снятых стереофильмов [Текст] / А. Анциферова, Д. Ватолин // 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.
4. *Анциферова, А.* Автоматический метод оценки степени усталости от просмотра 3D-видео [Текст] / А. Анциферова, Д. Ватолин // 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.
5. Video compression dataset and benchmark of learning-based video-quality metrics [Text] / A. Antsiferova [et al.] // Advances in Neural Information Processing Systems. Vol. 35. — 2022. — P. 13814—13825.
6. BSQ-rate: a new approach for video-codec performance comparison and drawbacks of current solutions [Text] / A. V. Zvezdakova (Antsiferova) [et al.] // Programming and computer software. — 2020. — Vol. 46, no. 3. — P. 183—194.
7. Applying Objective Quality Metrics to Video-Codec Comparisons: Choosing the Best Metric for Subjective Quality Estimation [Text] / A. Antsiferova [et al.] // Graphicon-Conference on Computer Graphics and Vision. Vol. 31. — 2021. — P. 199—210.

8. Barriers towards no-reference metrics application to compressed video quality analysis: on the example of no-reference metric NIQE [Text] / A. Zvezdakova (Antsiferova) [et al.] // Graphicon-Conference on Computer Graphics and Vision. Vol. 2485. — 2019. — P. 22—27.
9. BSQ-rate: новый подход к сравнению производительности видеокодеков и недостатки существующих решений [Текст] / А. В. Звездакова (Анциферова) [и др.] // Труды Института системного программирования РАН. — 2020. — Т. 32, № 1. — С. 89—108.
10. Comparing the robustness of modern no-reference image-and video-quality metrics to adversarial attacks [Text] / A. Antsiferova [et al.] // Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 38. — 2024. — P. 700—708.
11. *Shumitskaya, E.* Fast Adversarial CNN-based Perturbation Attack on No-Reference Image-and Video-Quality Metrics [Text] / E. Shumitskaya, A. Antsiferova, D. S. Vatolin // Tiny Papers @ ICLR 2023, Kigali, Rwanda, May 5, 2023. — 2023.
12. *Shumitskaya, E.* Towards adversarial robustness verification of no-reference image- and video-quality metrics [Text] / E. Shumitskaya, A. Antsiferova, D. Vatolin // Computer Vision and Image Understanding. — 2024. — Vol. 240. — P. 103913.
13. *Shumitskaya, E.* Universal Perturbation Attack on Differentiable No-Reference Image- and Video-Quality Metrics [Text] / E. Shumitskaya, A. Antsiferova, D. S. Vatolin // 33rd British Machine Vision Conference 2022, BMVC 2022, London, UK, November 21-24, 2022. — 2022.
14. Hacking VMAF with video color and contrast distortion [Text] / A. Zvezdakova (Antsiferova) [et al.] // Graphicon-Conference on Computer Graphics and Vision. Vol. 2485. — 2019. — P. 53—57.
15. Hacking VMAF and VMAF NEG: vulnerability to different preprocessing methods [Text] / M. Siniukov [et al.] // 2021 4th Artificial Intelligence and Cloud Computing Conference. — 2021. — P. 89—96.
16. Development of neural network-based video preprocessing method to increase the VMAF score relative to source video using distillation [Текст] / А. Solovev [и др.] // Keldysh Institute preprints. — Moscow, 2023. — № 66. — С. 1—11.

17. *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).*
18. Objective Video Quality Assessment Methods: A Classification, Review, and Performance Comparison [Текст] / S. Chikkerur [и др.] // IEEE Transactions on Broadcasting. — 2011. — Т. 57, № 2. — С. 165—182.
19. Image quality assessment: from error visibility to structural similarity [Текст] / Z. Wang [и др.] // IEEE Transactions on Image Processing. — 2004. — Т. 13, № 4. — С. 600—612.
20. Toward A Practical Perceptual Video Quality Metric [Текст] / Z. Li [и др.] // Netflix Technology Blog. — 2016.
21. No-reference image and video quality assessment: a classification and review of recent approaches [Текст] / M. Shahid [и др.] // EURASIP Journal on image and Video Processing. — 2014. — Т. 2014, № 1. — С. 1—32.
22. Video Quality Assessment of User Generated Content: A Benchmark Study and a New Model [Текст] / Z. Tu [и др.] // 2021 IEEE International Conference on Image Processing (ICIP). — 2021. — С. 1409—1413.
23. End-to-End Blind Image Quality Assessment Using Deep Neural Networks [Текст] / K. Ma [и др.] // IEEE Transactions on Image Processing. — 2018. — Т. 27, № 3. — С. 1202—1213.
24. *Li, D. Quality Assessment of In-the-Wild Videos [Текст] / D. Li, T. Jiang, M. Jiang // Proceedings of the 27th ACM International Conference on Multimedia. — Nice, France : Association for Computing Machinery, 2019. — С. 2351—2359. — (MM '19).*
25. Deep Residual Learning for Image Recognition [Текст] / K. He [и др.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 06.2016.
26. *Li, D. Unified quality assessment of in-the-wild videos with mixed datasets training [Текст] / D. Li, T. Jiang, M. Jiang // International Journal of Computer Vision. — 2021. — Т. 129, № 4. — С. 1238—1257.*

27. From Patches to Pictures (PaQ-2-PiQ): Mapping the Perceptual Space of Picture Quality [Текст] / Z. Ying [и др.] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). — 06.2020.
28. KonIQ-10k: An Ecologically Valid Database for Deep Learning of Blind Image Quality Assessment [Текст] / V. Hosu [и др.] // IEEE Transactions on Image Processing. — 2020. — Т. 29. — С. 4041—4056.
29. Perceptual Quality Assessment of Smartphone Photography [Текст] / Y. Fang [и др.] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). — 06.2020.
30. *Li, D.* Norm-in-norm loss with faster convergence and better performance for image quality assessment [Текст] / D. Li, T. Jiang, M. Jiang // Proceedings of the 28th ACM International Conference on Multimedia. — 2020. — С. 789—797.
31. *Talebi, H.* NIMA: Neural image assessment [Текст] / H. Talebi, P. Milanfar // IEEE transactions on image processing. — 2018. — Т. 27, № 8. — С. 3998—4011.
32. *Wang, Z.* Multiscale structural similarity for image quality assessment [Текст] / Z. Wang, E. P. Simoncelli, A. C. Bovik // The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, 2003. Т. 2. — Ieee. 2003. — С. 1398—1402.
33. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric [Текст] / R. Zhang [и др.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 06.2018.
34. Image quality assessment: Unifying structure and texture similarity [Текст] / K. Ding [и др.] // IEEE transactions on pattern analysis and machine intelligence. — 2020.
35. *Tencent.* Dvqa — deep learning-based video quality assessment. [Текст] / Tencent. — 2022.
36. C3DVQA: Full-reference video quality assessment with 3d convolutional neural network [Текст] / M. Xu [и др.] // ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). — IEEE. 2020. — С. 4447—4451.
37. Fovvideovdp: A visible difference predictor for wide field-of-view video [Текст] / R. K. Mantiuk [и др.] // ACM Transactions on Graphics (TOG). — 2021. — Т. 40, № 4. — С. 1—19.

38. ST-GREED: Space-time generalized entropic differences for frame rate dependent video quality prediction [Текст] / P. C. Madhusudana [и др.] // IEEE Transactions on Image Processing. — 2021. — Т. 30. — С. 7446—7457.
39. VMAF: The journey continues [Текст] / Z. Li [и др.] // Netflix Technology Blog. — 2018. — Т. 25.
40. Image quality assessment by separately evaluating detail losses and additive impairments [Текст] / S. Li [и др.] // IEEE Transactions on Multimedia. — 2011. — Т. 13, № 5. — С. 935—949.
41. *Sheikh, H. R.* An information fidelity criterion for image quality assessment using natural scene statistics [Текст] / H. R. Sheikh, A. C. Bovik, G. De Veciana // IEEE Transactions on image processing. — 2005. — Т. 14, № 12. — С. 2117—2128.
42. *Deng, S.* VMAF based rate-distortion optimization for video coding [Текст] / S. Deng, J. Han, Y. Xu // 2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP). — IEEE. 2020. — С. 1—6.
43. *V-Nova.* FFmpeg with LCEVC [Текст] / V-Nova. — 2023. — <https://docs.v-nova.com/>.
44. Comparison of full-reference image quality models for optimization of image processing systems [Текст] / K. Ding [и др.] // International Journal of Computer Vision. — 2021. — Т. 129. — С. 1258—1281.
45. *Kettunen, M.* E-LPIPS: robust perceptual image similarity via random transformation ensembles [Текст] / M. Kettunen, E. Härkönen, J. Lehtinen // arXiv preprint arXiv:1906.03973. — 2019.
46. *Comparison, M. V. C.* MSU Video Codecs Comparison 2021 Part 2: Subjective [Текст] / M. V. C. Comparison. — 2021. — http://www.compression.ru/video/codec_comparison/2021/subjective_report.html.
47. Benchmarking ultra-high-definition image super-resolution [Текст] / K. Zhang [и др.] // Proceedings of the IEEE/CVF international conference on computer vision. — 2021. — С. 14769—14778.
48. *Bing, M.* A Behind the Scenes Look at How Bing is Improving Image Search Quality [Текст] / M. Bing. — 2013. — <https://blogs.bing.com/search-quality-insights/2013/08/23/a-behind-the-scenes-look-at-how-bing-is-improving-image-search-quality>.

49. *MediaEval*. Pixel Privacy: Quality Camouflage for Social Images [Текст] / *MediaEval*. — 2020. — <https://multimediaeval.github.io/editions/2020/tasks/pixelprivacy/>.
50. Рожкова, Г. Теоретическая оценка дефектности стереокадров и реальный зрительный дискомфорт [Текст] / Г. Рожкова // VIII Международная научно-практическая конференция "Запись и воспроизведение объёмных изображений в кинематографе и других отраслях". — ВГИК. 2016. — С. 85—98.
51. Methodology for stereoscopic motion-picture quality assessment [Текст] / A. Voronov [и др.] // *Proc. SPIE*. Т. 8648. — 2013. — С. 864810.
52. Sinno, Z. Large-Scale Study of Perceptual Video Quality [Текст] / Z. Sinno, A. C. Bovik // *IEEE Transactions on Image Processing*. — 2019. — Т. 28, № 2. — С. 612—627.
53. Video Quality Assessment of User Generated Content: A Benchmark Study and a New Model [Текст] / Z. Tu [и др.] // 2021 IEEE International Conference on Image Processing (ICIP). — 2021. — С. 1409—1413.
54. User-generated Video Quality Assessment: A Subjective and Objective Study [Текст] / Y. Li [и др.] // *IEEE Transactions on Multimedia*. — 2021. — С. 1—1.
55. MCL-JCV: A JND-based H.264/AVC video quality assessment dataset [Текст] / H. Wang [и др.] // 2016 IEEE International Conference on Image Processing (ICIP). — 2016. — С. 1509—1513.
56. A H.264/AVC video database for the evaluation of quality metrics [Текст] / F. De Simone [и др.] // 2010 IEEE International Conference on Acoustics, Speech and Signal Processing. — 2010. — С. 2430—2433.
57. Study of Temporal Effects on Subjective Video Quality of Experience [Текст] / C. G. Vampis [и др.] // *IEEE Transactions on Image Processing*. — 2017. — Т. 26, № 11. — С. 5217—5231.
58. Study of Subjective and Objective Quality Assessment of Audio-Visual Signals [Текст] / X. Min [и др.] // *IEEE Transactions on Image Processing*. — 2020. — Т. 29. — С. 6054—6068.
59. MCL-V: A streaming video quality assessment database [Текст] / J. Y. Lin [и др.] // *Journal of Visual Communication and Image Representation*. — 2015. — Т. 30. — С. 1—9.

60. *Paudyal, P.* A study on the effects of quality of service parameters on perceived video quality [Текст] / P. Paudyal, F. Battisti, M. Carli // 2014 5th European Workshop on Visual Information Processing (EUVIP). — 2014. — С. 1—6.
61. *Keimel, C.* The TUM high definition video datasets [Текст] / C. Keimel, A. Redl, K. Diepold // 2012 Fourth International Workshop on Quality of Multimedia Experience. — 2012. — С. 97—102.
62. In-Capture Mobile Video Distortions: A Study of Subjective Behavior and Objective Algorithms [Текст] / D. Ghadiyaram [и др.] // IEEE Transactions on Circuits and Systems for Video Technology. — 2018. — Т. 28, № 9. — С. 2061—2077.
63. CVD2014—A Database for Evaluating No-Reference Video Quality Assessment Algorithms [Текст] / M. Nuutinen [и др.] // IEEE Transactions on Image Processing. — 2016. — Т. 25, № 7. — С. 3073—3086.
64. The Konstanz natural video database (KoNViD-1k) [Текст] / V. Hosu [и др.] // 2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX). — 2017. — С. 1—6.
65. *Wang, Y.* YouTube UGC Dataset for Video Compression Research [Текст] / Y. Wang, S. Inguva, B. Adsumilli // 2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP). — 2019. — С. 1—5.
66. Patch-VQ: 'Patching Up' the Video Quality Problem [Текст] / Z. Ying [и др.] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). — 06.2021. — С. 14019—14029.
67. VideoSet: A large-scale compressed video quality dataset based on JND measurement [Текст] / H. Wang [и др.] // Journal of Visual Communication and Image Representation. — 2017. — Т. 46. — С. 292—302.
68. UGC-VIDEO: perceptual quality assessment of user-generated videos [Текст] / Y. Li [и др.] // 2020 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). — IEEE. 2020. — С. 35—38.
69. GamingVideoSET: a dataset for gaming video streaming applications [Текст] / N. Barman [и др.] // 2018 16th Annual Workshop on Network and Systems Support for Games (NetGames). — IEEE. 2018. — С. 1—6.

70. No-reference video quality estimation based on machine learning for passive gaming video streaming applications [Текст] / N. Barman [и др.] // IEEE Access. — 2019. — Т. 7. — С. 74511—74527.
71. A subjective study for the design of multi-resolution ABR video streams with the VP9 codec [Текст] / C. Chen [и др.] // Electronic Imaging. — 2016. — Т. 2016, № 2. — С. 1—5.
72. *Lloyd, S.* Least squares quantization in PCM [Текст] / S. Lloyd // IEEE Transactions on Information Theory. — 1982. — Т. 28, № 2. — С. 129—137.
73. *IBM.* Recommended encoding settings for IBM Watson Media [Текст] / IBM.
74. *Twitch.* Recommended encoding settings for twitch streaming [Текст] / Twitch.
75. *YouTube.* Recommended encoding settings for youtube [Текст] / YouTube.
76. The relation between MOS and pairwise comparisons and the importance of cross-content comparisons [Текст] / E. Zerman [и др.] // Human Vision and Electronic Imaging Conference, IS&T International Symposium on Electronic Imaging (EI 2018). — 2018.
77. *Corey, D. M.* Averaging correlations: Expected values and bias in combined Pearson r s and Fisher's z transformations [Текст] / D. M. Corey, W. P. Dunlap, M. J. Burke // The Journal of general psychology. — 1998. — Т. 125, № 3. — С. 245—261.
78. *Winkler, S.* Analysis of Public Image and Video Databases for Quality Assessment [Текст] / S. Winkler // IEEE Journal of Selected Topics in Signal Processing. — 2012. — Т. 6, № 6. — С. 616—625.
79. *ITU.* Method for specifying accuracy and cross- calibration of Video Quality Metrics (VQM) [Текст] / ITU. — 2004.
80. *Ciaramello, F. M.* Supplemental subjective testing to evaluate the performance of image and video quality estimators [Текст] / F. M. Ciaramello, A. R. Reibman // Human Vision and Electronic Imaging XVI. Т. 7865. — SPIE. 2011. — С. 249—257.
81. *Ciaramello, F. M.* Systematic stress testing of image quality estimators [Текст] / F. M. Ciaramello, A. R. Reibman // 2011 18th IEEE International Conference on Image Processing. — IEEE. 2011. — С. 3101—3104.

82. *Liu, H.* Software to stress test image quality estimators [Текст] / H. Liu, A. R. Reibman // 2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX). — IEEE. 2016. — С. 1—6.
83. *Wang, Z.* Maximum differentiation (MAD) competition: A methodology for comparing computational models of perceptual quantities [Текст] / Z. Wang, E. P. Simoncelli // Journal of Vision. — 2008. — Т. 8, № 12. — С. 8—8.
84. Group mad competition—a new methodology to compare objective image quality models [Текст] / К. Ма [и др.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2016. — С. 1664—1673.
85. Blind image quality assessment using a deep bilinear convolutional neural network [Текст] / W. Zhang [и др.] // IEEE Transactions on Circuits and Systems for Video Technology. — 2018. — Т. 30, № 1. — С. 36—47.
86. On the Generation of Adversarial Samples for Image Quality Assessment [Текст] / Q. Sang [и др.] // Available at SSRN 4112969. —
87. Boosting adversarial attacks with momentum [Текст] / Y. Dong [и др.] // Proceedings of the IEEE conference on computer vision and pattern recognition. — 2018. — С. 9185—9193.
88. *Mittal, A.* Making a “completely blind” image quality analyzer [Текст] / A. Mittal, R. Soundararajan, A. C. Bovik // IEEE Signal processing letters. — 2012. — Т. 20, № 3. — С. 209—212.
89. *Korhonen, J.* Adversarial Attacks against Blind Image Quality Assessment Models [Текст] / J. Korhonen, J. You // Proceedings of the 2nd Workshop on Quality of Experience in Visual Multimedia Applications. — 2022. — С. 3—11.
90. Perceptual Attacks of No-Reference Image Quality Models with Human-in-the-Loop [Текст] / W. Zhang [и др.] // Advances in Neural Information Processing Systems. Т. 35 / под ред. S. Koyejo [и др.]. — Curran Associates, Inc., 2022. — С. 2916—2929.
91. Universal adversarial perturbations [Текст] / S.-M. Moosavi-Dezfooli [и др.] // Proceedings of the IEEE conference on computer vision and pattern recognition. — 2017. — С. 1765—1773.
92. Exploring Vulnerabilities of No-Reference Image Quality Assessment Models: A Query-Based Black-Box Method [Текст] / C. Yang, Y. Liu, D. Li [и др.] // arXiv preprint arXiv:2401.05217. — 2024.

93. *Ronneberger, O.* U-net: Convolutional networks for biomedical image segmentation [Текст] / O. Ronneberger, P. Fischer, T. Brox // International Conference on Medical image computing and computer-assisted intervention. — Springer. 2015. — С. 234—241.
94. A fast and elitist multiobjective genetic algorithm: NSGA-II [Текст] / K. Deb [и др.] // IEEE transactions on evolutionary computation. — 2002. — Т. 6, № 2. — С. 182—197.
95. Adaptive logarithmic mapping for displaying high contrast scenes [Текст] / F. Drago [и др.] // Computer graphics forum. Т. 22. — Wiley Online Library. 2003. — С. 419—426.
96. *Mantiuk, R.* A perceptual framework for contrast processing of high dynamic range images [Текст] / R. Mantiuk, K. Myszkowski, H.-P. Seidel // ACM Transactions on Applied Perception (TAP). — 2006. — Т. 3, № 3. — С. 286—308.
97. Photographic tone reproduction for digital images [Текст] / E. Reinhard [и др.] // Proceedings of the 29th annual conference on Computer graphics and interactive techniques. — 2002. — С. 267—276.
98. *Beyer, H.-G.* Evolution strategies—a comprehensive introduction [Текст] / H.-G. Beyer, H.-P. Schwefel // Natural computing. — 2002. — Т. 1, № 1. — С. 3—52.
99. ProxIQA: A proxy approach to perceptual optimization of learned image compression [Текст] / L.-H. Chen [и др.] // IEEE Transactions on Image Processing. — 2020. — Т. 30. — С. 360—373.
100. Microsoft coco: Common objects in context [Текст] / T.-Y. Lin [и др.] // European conference on computer vision. — Springer. 2014. — С. 740—755.
101. *Kingma, D. P.* Adam: A method for stochastic optimization [Текст] / D. P. Kingma, J. Ba // arXiv preprint arXiv:1412.6980. — 2014.
102. Xiph.org video test media (derf's collection) [Текст].
103. *Ozer, J.* VMAF is Hackable: What Now? [Текст] / J. Ozer. — 2020.
104. *Li, Z.* On VMAF's property in the presence of image enhancement operations [Текст] / Z. Li. — 2021.

105. The pascal visual object classes (voc) challenge [Текст] / M. Everingham [и др.] // International journal of computer vision. — 2010. — Т. 88, № 2. — С. 303—338.
106. Video enhancement with task-oriented flow [Текст] / Т. Xue [и др.] // International Journal of Computer Vision. — 2019. — Т. 127, № 8. — С. 1106—1125.
107. An objective method for 3D quality prediction using visual annoyance and acceptability level [Текст] / D. Khaustova [и др.] // SPIE/IS&T Electronic Imaging. — International Society for Optics, Photonics. 2015. — 93910P.
108. 3D video subjective quality: a new database and grade comparison study [Текст] / E. Dumić [и др.] // Multimedia tools and applications. — 2017. — Т. 76, № 2. — С. 2087—2109.
109. Multimodal interactive continuous scoring of subjective 3D video quality of experience [Текст] / Т. Kim [и др.] // IEEE Transactions on Multimedia. — 2014. — Т. 16, № 2. — С. 387—402.
110. *Chen, W.* Multidimensional characterization of quality of experience of stereoscopic 3d tv [Текст] : дис. ... канд. / Chen Wei. — Université de Nantes Angers Le Mans, 2012.
111. EEG-based detection and evaluation of fatigue caused by watching 3DTV [Текст] / С. Chen [и др.] // Displays. — 2013. — Т. 34, № 2. — С. 81—88.
112. *Kim, Y.-J.* EEG based comparative measurement of visual fatigue caused by 2D and 3D displays [Текст] / Y.-J. Kim, E. C. Lee // International Conference on Human-Computer Interaction. — Springer. 2011. — С. 289—292.
113. *Cho, S.-H.* An Assessment of Visual Discomfort Caused by Motion-in-Depth in Stereoscopic 3D Video. [Текст] / S.-H. Cho, H.-B. Kang // BMVC. — 2012. — С. 1—10.
114. Evaluating depth perception of 3D stereoscopic videos [Текст] / P. Lebreton [и др.] // IEEE Journal of Selected Topics in Signal Processing. — 2012. — Т. 6, № 6. — С. 710—720.
115. *Lambooi, M.* Visual discomfort of 3D TV: Assessment methods and modeling [Текст] / M. Lambooi, W. A. IJsselsteijn, I. Heynderickx // Displays. — 2011. — Т. 32, № 4. — С. 209—218.

116. Visual fatigue and discomfort after stereoscopic display viewing [Текст] / L. Zhang [и др.] // *Acta ophthalmologica*. — 2013. — Т. 91, № 2.
117. NAMA3DS1-COSPAD1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences [Текст] / M. Urvoy [и др.] // *Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on*. — IEEE. 2012. — С. 109—114.
118. *Aflaki, P.* Subjective quality assessment of asymmetric stereoscopic 3D video [Текст] / P. Aflaki, M. M. Hannuksela, M. Gabbouj // *Signal, Image and Video Processing*. — 2015. — Т. 9, № 2. — С. 331—345.
119. Towards standardized 3DTV QoE assessment: Cross-lab study on display technology and viewing environment parameters [Текст] / M. Barkowsky [и др.] // *Stereoscopic Displays and Applications XXIV*. Т. 8648. — International Society for Optics, Photonics. 2013. — С. 864809.
120. *Zeri, F.* Visual discomfort while watching stereoscopic three-dimensional movies at the cinema [Текст] / F. Zeri, S. Livi // *Ophthalmic and Physiological Optics*. — 2015. — Т. 35, № 3. — С. 271—282.
121. *Han, Y.* Extended no reference objective quality metric for stereoscopic 3D video [Текст] / Y. Han, Z. Yuan, G.-M. Muntean // *Communication Workshop (ICCW), 2015 IEEE International Conference on*. — IEEE. 2015. — С. 1729—1734.
122. *Banitalebi-Dehkordi, A.* An efficient human visual system based quality metric for 3D video [Текст] / A. Banitalebi-Dehkordi, M. T. Pourazad, P. Nasiopoulos // *Multimedia Tools and Applications*. — 2016. — Т. 75, № 8. — С. 4187—4215.
123. *Rousseeuw, P. J.* A fast algorithm for the minimum covariance determinant estimator [Текст] / P. J. Rousseeuw, K. V. Driessen // *Technometrics*. — 1999. — Т. 41, № 3. — С. 212—223.
124. *Liu, F. T.* Isolation forest [Текст] / F. T. Liu, K. M. Ting, Z.-H. Zhou // *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*. — IEEE. 2008. — С. 413—422.
125. LOF: identifying density-based local outliers [Текст] / M. M. Breunig [и др.] // *ACM sigmod record*. Т. 29. — ACM. 2000. — С. 93—104.

126. A density-based algorithm for discovering clusters in large spatial databases with noise. [Текст] / М. Ester [и др.] // Kdd. Т. 96. — 1996. — С. 226—231.
127. *Линник, Ю. В.* Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений [Текст] / Ю. В. Линник. — 1958.
128. *Huber, P. J.* Robust estimation of a location parameter [Текст] / P. J. Huber // The annals of mathematical statistics. — 1964. — С. 73—101.
129. *Theil, H.* A rank-invariant method of linear and polynomial regression analysis, 3; confidence regions for the parameters of polynomial regression equations [Текст] / H. Theil // Indagationes Mathematicae. — 1950. — Т. 1, № 2. — С. 467—482.
130. *Sen, P. K.* Estimates of the regression coefficient based on Kendall's tau [Текст] / P. K. Sen // Journal of the American statistical association. — 1968. — Т. 63, № 324. — С. 1379—1389.
131. Binocular Rivalry Oriented Predictive Autoencoding Network for Blind Stereoscopic Image Quality Measurement [Текст] / J. Xu [и др.] // IEEE Transactions on Instrumentation and Measurement. — 2020. — Т. 70. — С. 1—13.
132. *Goldmann, L.* A comprehensive database and subjective evaluation methodology for quality of experience in stereoscopic video [Текст] / L. Goldmann, F. De Simone, T. Ebrahimi // Three-Dimensional Image Processing (3DIP) and Applications. Т. 7526. — SPIE. 2010. — С. 242—252.
133. *AviSynthFAQ, I.* Avisynth, 5 pages [online],[retrieved on Dec. 7, 2007] [Текст] / I. AviSynthFAQ // Retrieved from the Internet URL: <http://avisynth.org/medawiki/index.php>. —
134. *Xiao, F.* DCT-based Video Quality Evaluation [Текст] / F. Xiao. — 2000.
135. *Apple.* Evaluate videos with the advanced video quality tool. [Текст] / Apple. — 2022.

Список рисунков

11	Фрагменты кадров из видео, закодированных видеокодеком стандарта H.265 и видеокодеком нового стандарта LCEVC-H.265. LCEVC использует повышение разрешения, которое позволяет восстановить больше деталей и приводит к появлению искажений нового типа.	19
12	Атака на метод оценки качества изображений PaQ-2-PiQ.	20
21	Схема предложенной методики тестирования алгоритмов оценки качества видео.	27
22	Сравнение пространственной и временной сложности 18 418 пользовательских видео с ресурса vimeo.com с популярными открытыми коллекциями видео media.xiph.org.	28
23	Распределение битрейта видео в открытой и скрытой частях набора видео.	33
24	Сравнение PSNR (диапазон и равномерность) предложенного набора видео с другими наборами видео.	34
25	Распределение значений методов в предложенном наборе данных.	37
31	Схема работы метода Video Multimethod Assessment Fusion (VMAF).	40
32	Сравнение оценок VMAF и SSIM для различных конфигураций нерезкого маскирования и выравнивания гистограммы на видео Bay timelapse. Важны результаты во втором квадранте, где значения SSIM не изменились, а значения VMAF увеличились.	47
33	Схема предлагаемого метода состязательной атаки на VMAF.	48
34	Использованная архитектура на базе U-Net.	49
35	Предложенный метод вычисления повышения целевого метода и потерь прокси-метода с помощью нормализованных RD-кривых для видео на четырех уровнях амплитуды атаки. Правый график показывает зависимость повышения целевого метода и снижения прокси-метода для этих уровней амплитуды.	52
36	Обученные универсальные возмущения для всех семи тестируемых методов.	53

37	Зависимости между приростом значений целевого метода и потерями прокси-методов для всех протестированных неререференсных методов. Пунктирными линиями выделена область, где определены зависимости для всех методов. Чем меньше прирост значений целевого метода и выше потери PSNR, тем стабильнее метод.	54
38	Различные амплитуды универсального возмущения PaQ-2-PiQ, примененные к изображениям. Чем выше амплитуда, тем выше оценка PaQ-2-PiQ, но при этом визуальные искажения также становятся более заметными.	56
39	CSF маска и ее применение к универсальному возмущению на примере атаки PaQ-2-PiQ.	57
310	Повышение значений алгоритма VMAF с помощью протестированных методов обработки видео.	58
311	Сравнение кадров исходного обработанного видео с помощью метода CLANE. VMAF увеличился на 181,22 на левом видео, на 147,30 на среднем видео, на 114,75 на правом видео.	60
312	Примеры кадров с различными методами предварительной обработки, настроенными на увеличение VMAF NEG. Повышение оценок метода рассчитывается как разница между значением VMAF NEG для атакованного перед сжатием видео и сжатого видео без атаки.	61
313	Результаты визуальной оценки предложенных методов с помощью краудсорсингового подхода.	62
314	Визуальные артефакты: фиолетовые области.	63
315	Изменение границ, производимых различными вариантами предложенного метода.	63
41	Схематическая визуализация стереоскопических искажений.	65
42	Схема разработки предложенного метода оценки качества стереоскопического видео.	68

43	Временная и пространственная сложность выбранных видеофрагментов из фильмов. Левый график показывает зависимость средней длины вектора движения между кадрами в пикселях от средней длины вектора диспаратности в процентах от ширины кадра. Правый график отражает зависимость между средней длиной проекции вектора движения на вертикальную ось в пикселях и средней яркостью кадра по шкале от 0 до 255.	74
44	Процесс проведения эксперимента (яркость изображения искусственно повышена). Аудитория П-13 факультета ВМК МГУ.	76
45	Упорядоченные по возрастанию усредненные оценки дискомфорта участников.	77
46	Средние значения оценок дискомфорта для сцен. Линиями отмечены 30%-процентили.	77
47	Оценка качества для сцен с различными типами искажениями и их интенсивностью. Здесь С обозначает несоответствие цвета ракурсов, R — поворот одного из ракурсов, T — временной сдвиг и S — несоответствие масштаба ракурсов. Интенсивность искажения отмечена цифрами от 0 до 4, где 0 соответствует отсутствию искажения (была продемонстрирована исходная сцена).	78
48	Среднее значение оценки дискомфорта для двух сцен в bootstrap-выборках с разным количеством респондентов.	79
49	Размер 95%-доверительного интервала к среднему значению дискомфорта в bootstrap-выборках с разным количеством респондентов.	80
A1	Прогнозируемый уровень дискомфорта, который может возникнуть у зрителей при просмотре проанализированных фильмов.	111
A2	Прогнозируемый уровень дискомфорта в зависимости от бюджета производства.	112
A3	Прогнозируемый уровень дискомфорта в зависимости от года выхода фильма в прокат. 2009 год — год, когда в прокат вышел культовый для стереокинематографа фильм «Аватар».	112
A4	Ранжированные по степени значимости в задаче прогнозирования уровня дискомфорта зрителей признаки технического качества стереоскопического видео.	113

- A5 Взаимная корреляция признаков, вычисленных для стереоскопического видео, показанного во время эксперимента. Темные цвета соответствуют более низкому значению коэффициента корреляции Спирмена. 114
- A6 Точность работы линейной модели с функцией потерь Хьюбера в зависимости от значения параметра L2-регуляризации. 114
- A7 Точность работы линейной модели с функцией потерь Хьюбера в зависимости от значения параметра L2-регуляризации и количества используемых признаков. 115

Список таблиц

1	Обзор наборов видео с эталонными оценками и предложенный набора данных.	26
2	Обзор существующих сравнений алгоритмов оценки качества видео.	36
3	Классификация ошибок методов оценки качества изображений и видео.	41
4	Средние различия PaQ-2-PiQ до и после применения предложенной атаки и MADС к 200 тестовым изображениям. Оценки методов для исходных изображений вычитаются из оценок после атаки.	54
5	Оценки устойчивости и корреляции SROCC для протестированных методов.	55
6	Сравнение различных методов атаки на метод VMAF с помощью вариантов предложенного метода с различными методами обработки видео.	59
7	Эксперименты, проведенные в рамках исследований причин появления дискомфорта у зрителей при просмотре стереоскопического видео.	67
8	Сравнение среднего значения доверительных интервалов ответов участников.	71
9	Количество участников, отметивших наличие негативных симптомов до и после эксперимента.	77
10	Уровень дискомфорта зрителей при просмотре сцен с различными искажениями.	81
11	Точность работы различных моделей для оценки качества стереоскопического видео при обучении и валидации на полученном наборе данных.	82
12	Точность работы методов с регуляризацией для оценки качества видео на тестовой выборке.	82
13	Сравнение различных методов повышения VMAF с помощью лучших полученных методов предварительной обработки видео.	83

- 14 Коэффициенты корреляции Спирмена и Кендалла нереперенсных методов оценки качества видео для созданного полного набора данных и его открытой части. 109
- 15 Коэффициенты корреляции Спирмена и Кендалла полнореференсных методов оценки качества видео для созданного полного набора данных и его открытой части. 110

Приложение А

Приложение

А.1 Глава 1: зависимость прогнозируемого уровня дискомфорта зрителей при просмотре стереофильма от бюджета его производства и года выхода в прокат.

На рисунке **A2** отражена зависимость прогнозируемого уровня дискомфорта зрителей при просмотре стереофильма от бюджета его производства, на рисунке **A3** — от года выхода фильма в прокат.

А.2 Глава 1: визуализации выбора признаков для обучения метода прогнозирования дискомфорта.

Упорядочивание признаков по степени значимости показано на рисунке **A4**. Оба метода отмечают значения несоответствия по резкости, временного сдвига и достоверности значений геометрических искажений как одни из значимых признаков, в то время как позиции других признаков отличаются. Взаимная коррелированность признаков визуализирована на рисунке **A5**.

А.3 Глава 1: исследование влияния коэффициента регуляризации на точность предложенной модели.

На рисунке **A6** представлены результаты работы данной модели с различным значением параметра регуляризации α на валидационном наборе. На первом графике представлен коэффициент детерминации измеренных и прогнозированных моделью значений. Видно, что при увеличении коэффициента α , приводящего к большему штрафу на размер параметров регрессии, точность ра-

боты сначала повышается за счет уменьшения эффекта переобучения, а затем постепенно снижается из-за недообучения при ограничении на параметры.

Дополнительно было проведено исследование влияния на точность модели количества признаков совместно с параметром регуляризации. Рисунок **A7** отражает трехмерный график данной зависимости, на котором видно, что наилучшие результаты модели соответствуют использованию 7 и 13 признаков. Для сравнения моделей с разным количеством параметров использовался скорректированный на количество признаков коэффициент детерминации. Таким образом, лучшее значение параметра регуляризации для метода регрессии с функцией потерь Хьюбера, при котором на валидационной выборке наблюдается лучшая точность, равно 5.

А.4 Глава 2: результаты сравнения методов оценки качества видео с эталонными оценками, полученными с помощью предложенной методики.

В таблицах **14** и **15** показаны коэффициенты корреляции Спирмена и Кендалла значений анализируемых методов с эталонными оценками качества. На всем наборе данных наибольшую корреляцию показал метод VMAF и его вариации, рассчитанные при различных соотношениях компонентов цветности. В оригинальной статье VMAF должен вычисляться только по яркостной компоненте изображения, но по результатам анализа, при вычислении его для всех компонент полученный метод YUV-VMAF работает лучше. Также, VMAF NEG показал более низкую корреляцию с эталонными оценками, чем оригинальная версия VMAF. Нереперенсные методы MDTVSAFA и LINEARITY показали самые высокие корреляции среди всех нереперенсных методов (около 0,91), что очень близко к лучшим результатам полнореференсных методов (0,94).

Набор данных	Весь набор (Открытый+Скрытый) 2486 видео		Открытый набор 1022 видео	
	Метод	SROCC	KROCC	SROCC
MEON [23]	0.507 (0.495, 0.518)	0.375 (0.366, 0.384)	0.554 (0.534, 0.574)	0.347 (0.330, 0.363)
Y-NIQE [88]	0.599 (0.586, 0.611)	0.421 (0.411, 0.431)	0.701 (0.679, 0.721)	0.552 (0.536, 0.568)
VIDEVAL [53]	0.729 (0.719, 0.738)	0.541 (0.531, 0.550)	0.719 (0.700, 0.737)	0.558 (0.540, 0.576)
KonCept512 [28]	0.836 (0.831, 0.841)	0.660 (0.655, 0.666)	0.861 (0.853, 0.868)	0.694 (0.686, 0.701)
NIMA [31]	0.849 (0.844, 0.854)	0.674 (0.668, 0.680)	0.868 (0.860, 0.875)	0.726 (0.716, 0.735)
PaQ-2-PiQ [27]	0.871 (0.866, 0.875)	0.707 (0.701, 0.713)	0.901 (0.894, 0.908)	0.750 (0.740, 0.759)
SPAQ MT-A [29]	0.879 (0.875, 0.884)	0.714 (0.708, 0.720)	0.912 (0.905, 0.919)	0.794 (0.785, 0.804)
SPAQ BL [29]	0.880 (0.875, 0.884)	0.710 (0.704, 0.716)	0.912 (0.905, 0.918)	0.789 (0.779, 0.798)
SPAQ MT-S [29]	0.882 (0.878, 0.886)	0.718 (0.712, 0.723)	0.912 (0.906, 0.918)	0.785 (0.776, 0.794)
VSFA [24]	0.905 (0.901, 0.908)	0.748 (0.743, 0.753)	0.891 (0.886, 0.897)	0.757 (0.749, 0.766)
Linearity [30]	0.910 (0.907, 0.913)	0.758 (0.754, 0.763)	0.905 (0.899, 0.911)	0.782 (0.773, 0.791)
MDTVSFA [26]	0.929 (0.927, 0.931)	0.788 (0.784, 0.792)	0.930 (0.927, 0.934)	0.811 (0.805, 0.818)

Таблица 14 — Коэффициенты корреляции Спирмена и Кендалла нереперенсных методов оценки качества видео для созданного полного набора данных и его открытой части.

Набор данных	Весь набор (Открытый+Скрытый) 2486 видео		Открытый набор 1022 видео	
	Метод	SROCC	KROCC	SROCC
FOV VIDEO [37]	0.527 (0.521, 0.534)	0.381 (0.377, 0.386)	0.565 (0.551, 0.579)	0.463 (0.445, 0.481)
LPIPS [33]	0.749 (0.742, 0.756)	0.566 (0.560, 0.573)	0.787 (0.774, 0.799)	0.660 (0.649, 0.672)
DVQA [35]	0.763 (0.756, 0.770)	0.578 (0.571, 0.585)	0.774 (0.762, 0.786)	0.685 (0.671, 0.698)
GREED [38]	0.764 (0.759, 0.769)	0.588 (0.583, 0.594)	0.790 (0.782, 0.797)	0.649 (0.639, 0.659)
Y-VQM [134]	0.821 (0.815, 0.827)	0.644 (0.637, 0.651)	0.881 (0.870, 0.890)	0.760 (0.749, 0.770)
DISTS [34]	0.847 (0.842, 0.851)	0.671 (0.666, 0.676)	0.873 (0.866, 0.879)	0.743 (0.735, 0.751)
AVQT [135]	0.876 (0.872, 0.879)	0.716 (0.711, 0.720)	0.889 (0.882, 0.896)	0.788 (0.780, 0.796)
YUV-PSNR	0.883 (0.879, 0.887)	0.725 (0.720, 0.730)	0.949 (0.946, 0.952)	0.836 (0.831, 0.842)
YUV-SSIM	0.906 (0.902, 0.909)	0.755 (0.750, 0.761)	0.948 (0.945, 0.951)	0.854 (0.848, 0.860)
Y-MS-SSIM [32]	0.909 (0.905, 0.912)	0.755 (0.750, 0.760)	0.946 (0.943, 0.949)	0.839 (0.834, 0.845)
Y-VMAF NEG [39]	0.914 (0.911, 0.917)	0.764 (0.760, 0.769)	0.945 (0.942, 0.948)	0.841 (0.836, 0.847)
YUV-VMAF NEG [39]	0.917 (0.914, 0.920)	0.769 (0.764, 0.773)	0.947 (0.944, 0.950)	0.850 (0.844, 0.856)
Y-VMAF (v061) [39]	0.942 (0.940, 0.944)	0.808 (0.804, 0.812)	0.945 (0.942, 0.948)	0.842 (0.836, 0.848)
YUV-VMAF (v061)[39]	0.943 (0.941, 0.945)	0.810 (0.806, 0.814)	0.948 (0.945, 0.951)	0.852 (0.846, 0.858)

Таблица 15 — Коэффициенты корреляции Спирмена и Кендалла полнореференсных методов оценки качества видео для созданного полного набора данных и его открытой части.

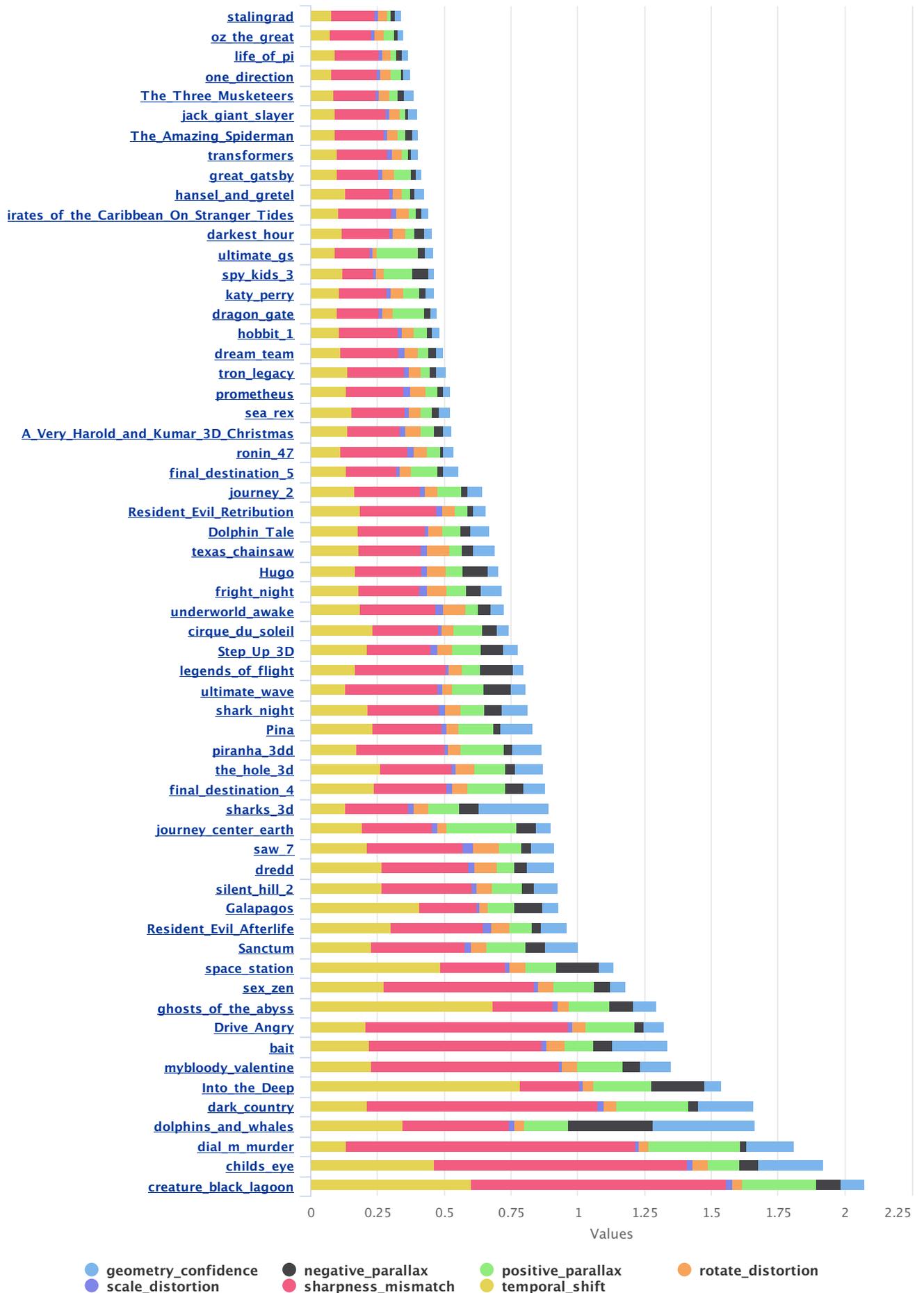


Рисунок А1 — Прогнозируемый уровень дискомфорта, который может возникнуть у зрителей при просмотре проанализированных фильмов.

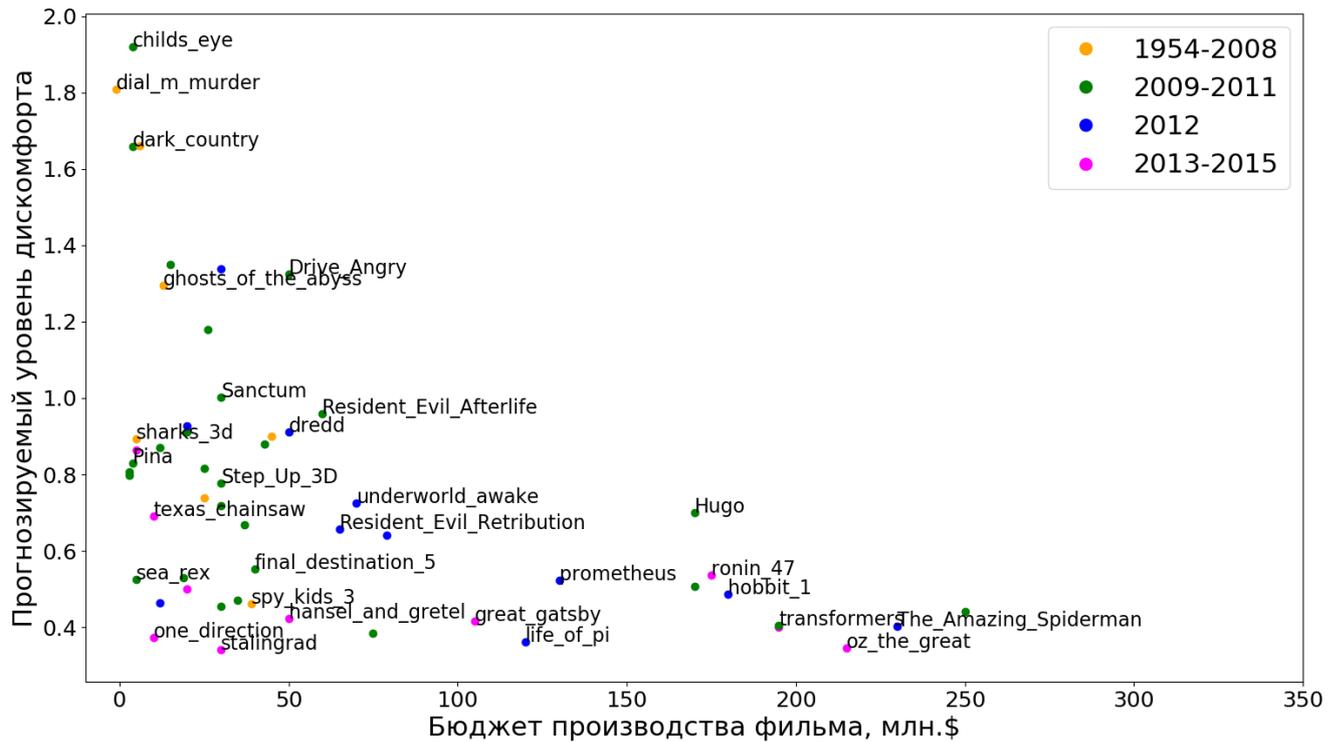


Рисунок А2 — Прогнозируемый уровень дискомфорта в зависимости от бюджета производства.

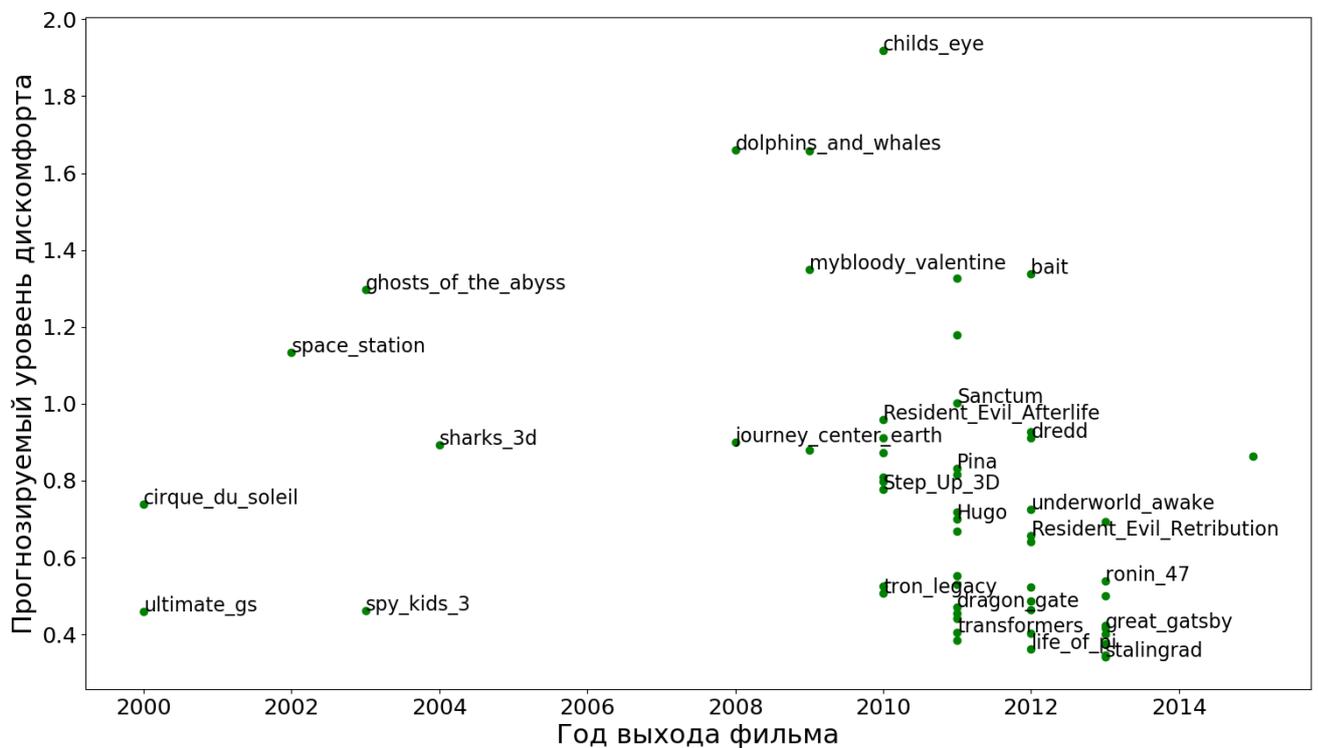
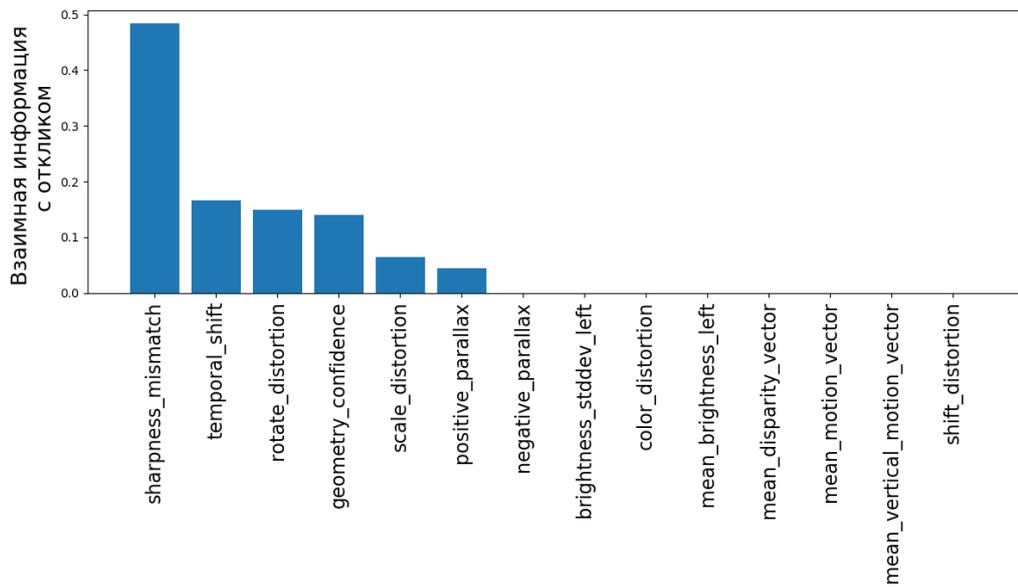
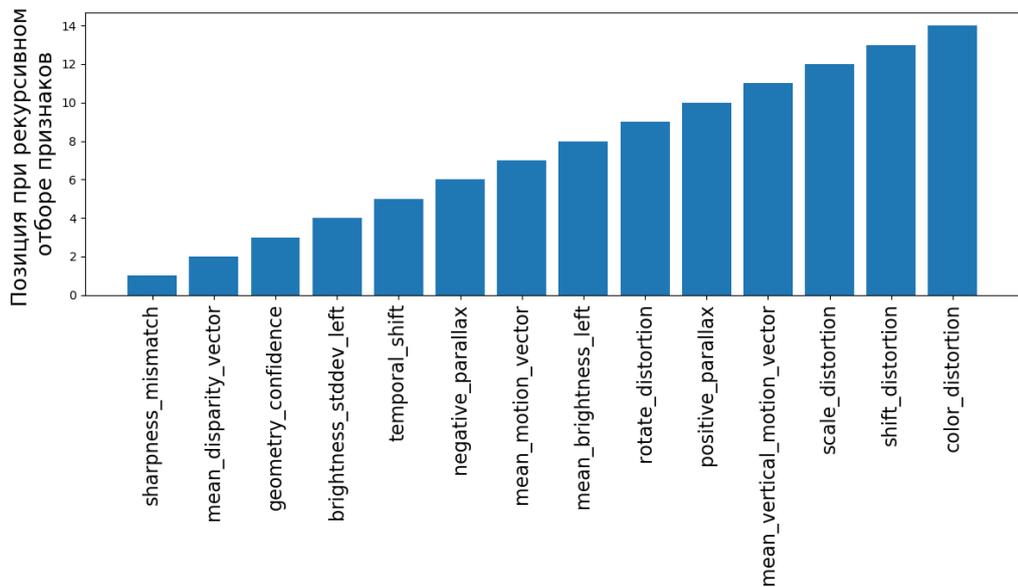


Рисунок А3 — Прогнозируемый уровень дискомфорта в зависимости от года выхода фильма в прокат. 2009 год — год, когда в прокат вышел культовый для стереокинематографа фильм «Аватар».

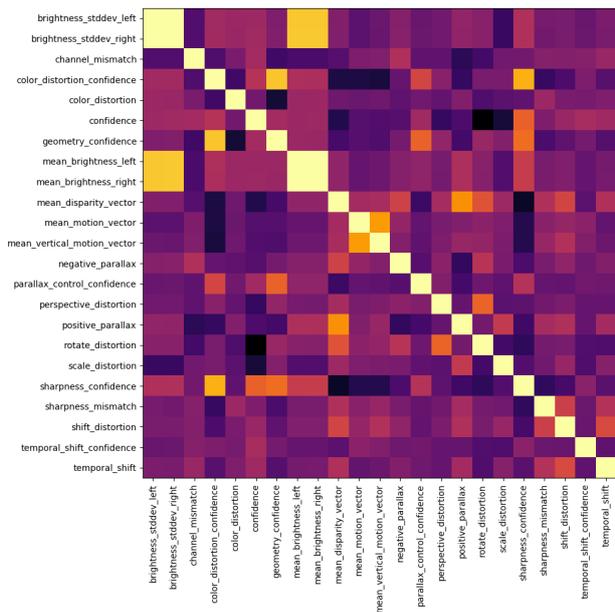


(а) Метод выбора k лучших на основе взаимной информации с ОТКЛИКОМ

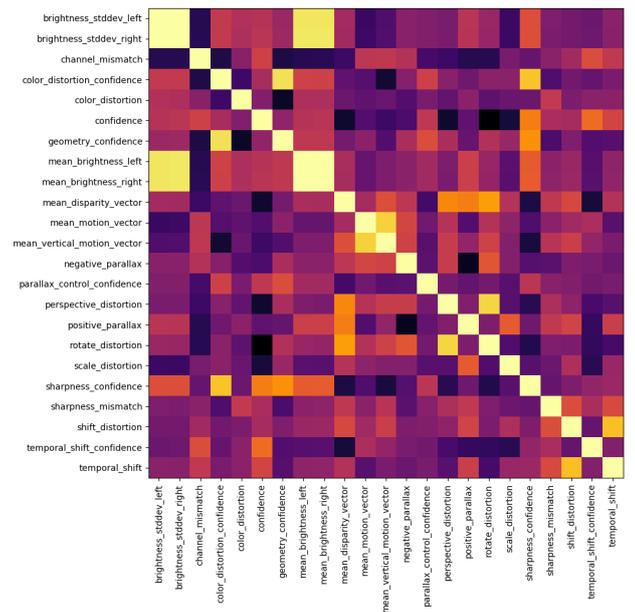


(б) Метод рекурсивного выбора лучших признаков на основе точности работы градиентного бустинга над деревьями

Рисунок А4 — Ранжированные по степени значимости в задаче прогнозирования уровня дискомфорта зрителей признаки технического качества стереоскопического видео.

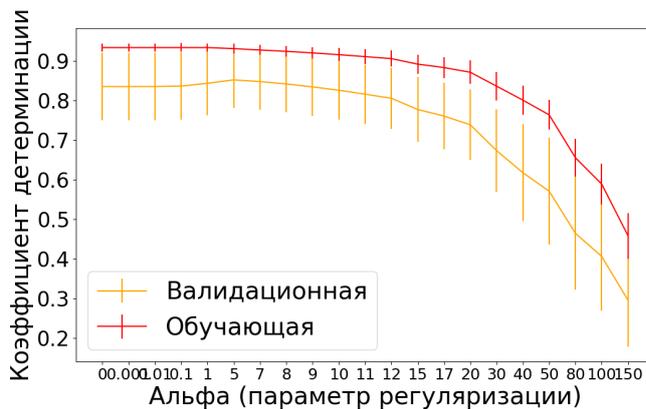


(а) Коррелированность покадровых значений признаков.

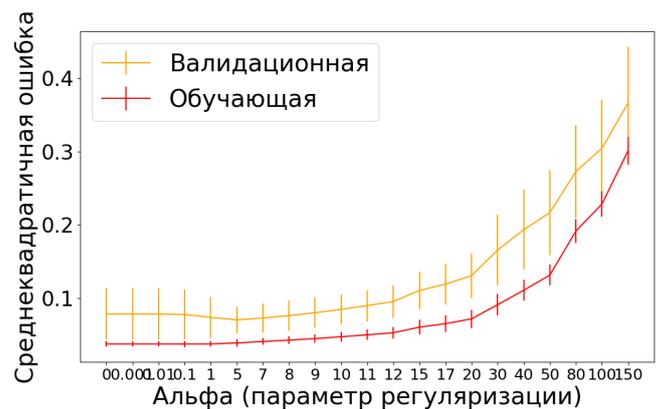


(б) Скоррелированность значений признаков, усредненных для сцен

Рисунок А5 — Взаимная корреляция признаков, вычисленных для стереоскопического видео, показанного во время эксперимента. Темные цвета соответствуют более низкому значению коэффициента корреляции Спирмена.



(а) Коэффициент детерминации



(б) Среднеквадратичная ошибка

Рисунок А6 — Точность работы линейной модели с функцией потерь Хьюбера в зависимости от значения параметра L2-регуляризации.

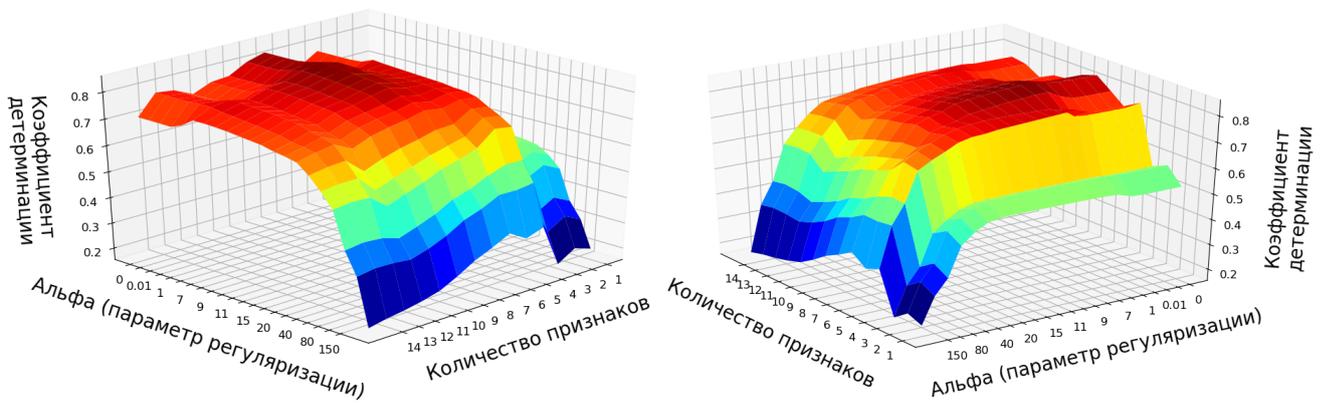


Рисунок А7 — Точность работы линейной модели с функцией потерь Хьюбера в зависимости от значения параметра L2-регуляризации и количества используемых признаков.