



ИПМ им.М.В.Келдыша РАН • Электронная библиотека

Препринты ИПМ • Препринт № 81 за 2022 г.



ISSN 2071-2898 (Print)
ISSN 2071-2901 (Online)

Е.Н. Богатырев, И.А. Молодецких,
Д.С. Ватолин, [В.А. Галактионов](#)

Исследование качества
сжатого видео после
повышения разрешения:
бенчмарк и метрика качества

Статья доступна по лицензии
[Creative Commons Attribution 4.0 International](#)



Рекомендуемая форма библиографической ссылки: Исследование качества сжатого видео после повышения разрешения: бенчмарк и метрика качества / Е.Н. Богатырев [и др.] // Препринты ИПМ им. М.В.Келдыша. 2022. № 81. 24 с. <https://doi.org/10.20948/prepr-2022-81>
<https://library.keldysh.ru/preprint.asp?id=2022-81>

О р д е н а Л е н и н а
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ
имени М.В.Келдыша
Р о с с и й с к о й а к а д е м и и н а у к

Е.Н. Богатырев, И.А. Молодецких,
Д.С. Ватолин, В.А. Галактионов

**Исследование качества сжатого видео
после повышения разрешения:
бенчмарк и метрика качества**

Москва — 2022

Богатырев Е.Н., Молодецких И.А., Ватолин Д.С., Галактионов В.А

**Исследование качества сжатого видео после повышения разрешения:
бенчмарк и метрика качества**

Разработан бенчмарк методов повышения разрешения для анализа возможностей методов по восстановлению сжатого видео. Набор данных для бенчмарка был собран с использованием видеокодеков 5 различных стандартов сжатия. Было проанализировано 17 современных моделей суперразрешения с помощью нашего бенчмарка и оценена их устойчивость к артефактам сжатия и способность сохранять контекст сцены. Чтобы получить точную субъективную оценку моделей суперразрешения, было проведено субъективное сравнение с помощью краудсорсинга.

Также были проанализированы результаты сравнения и разработана объективная метрика оценки качества, основанная на существующих объективных метриках, показавших лучшие результаты. Разработанная метрика превосходит другие метрики качества видео по корреляции Спирмена с субъективными оценками для задачи повышения разрешения сжатого видео.

Ключевые слова: суперразрешение, видеопроцессинг, датасет, видеокодек

Evgeney Bogatyrev, Ivan Molodetskikh, Dmitriy Vatolin, Vladimir Galaktionov

Compressed Video Quality Assessment for Super-Resolution: A Benchmark and a Quality Metric

We developed a super-resolution (SR) benchmark to analyze SR capabilities to upscale compressed videos. The dataset for the benchmark was collected using video codecs of 5 different compression standards. We assessed 17 state-of-the-art SR models using our benchmark and evaluated their ability to preserve scene context and their robustness to compression artifacts. To get an accurate perceptual ranking of SR models, we conducted a crowd-sourced side-by-side comparison of SR results.

We also analyzed the results of the benchmark and developed an objective quality assessment metric based on existing best-performing objective metrics. Our metric outperforms other video quality metrics by Spearman correlation with subjective scores for the task of upscaling compressed videos.

Key words: super-resolution, video processing, dataset, video compression

Исследование выполнено за счет гранта Российского научного фонда № 22-21-00478, <https://rscf.ru/project/22-21-00478/>

1 Введение

К 2022 году на видеотрафик приходится более 80% всего потребительского интернет-трафика [50], и, по прогнозам, эта цифра будет расти и дальше. Разрешение экранов мобильных устройств растет, и современные мобильные устройства поддерживают разрешение FullHD [39]. Несмотря на это, сотовые сети по-прежнему имеют низкую пропускную способность в большинстве регионов мира. Сжатие видео при минимальном изменении визуального качества видео сейчас как никогда актуально.

Последние работы показали, что новые методы сжатия видео с использованием глубокого обучения могут превзойти классические алгоритмы сжатия, однако эти методы значительно более ресурсоемкие [15, 9, 11]. В технологиях сжатия видео произошли некоторые прорывы, например, недавно разработанный стандарт сжатия MPEG-5 Part 2 Low Complexity Enhancement Video Coding (LCEVC) [30]. В архитектуре LCEVC используется дополнительный слой, который улучшает восстановленное после декодирования видео. Разрешение видео уменьшается перед кодированием, а затем увеличивается на устройстве пользователя. Такой подход приводит к значительному снижению нагрузки сети, однако на видео могут появляться артефакты сжатия, а некоторые детали могут быть потеряны после уменьшения разрешения видео.

Суперразрешение стало широко изучаемой темой в последние годы, и новые работы появляются ежемесячно [60, 63, 38, 29]. Суперразрешение может использоваться в различных задачах, включая улучшение, восстановление и даже сжатие изображений и видео [20]. Некоторые современные методы суперразрешения могут восстанавливать детали видео, которые были потеряны в исходном видео после понижения разрешения, работая при этом в реальном времени [60]. Соседние кадры могут быть использованы для восстановления деталей при увеличении разрешения, поскольку небольшое движение, вызванное дрожанием камеры, может дать достаточно информации о деталях видео, как это было продемонстрировано в камере Google Pixel 3 [59].

Некоторые современные видеокодеки [20, 7, 14] понижают разрешение видео перед сжатием, чтобы снизить битрейт, а затем повышают его до исходного разрешения с помощью методов суперразрешения. Такой подход снижает загрузку сети и сохраняет визуальное качество видео. Однако не все методы суперразрешения подходят для работы со сжатым видео. Также существует не так много моделей, которые могут работать в реальном времени и генерировать видео достойного качества.

Чтобы проанализировать, какие модели суперразрешения лучше работают с каждым из стандартов кодеков, и помочь исследователям найти лучшую модель для использования вместе с определенным кодеком, мы представляем наш бенчмарк Super-Resolution for Video Compression. Для разработки этого бенчмарка мы выбрали 17 современных моделей повышения разрешения с различными архитектурами и оценили их способности по восстановлению сжатого

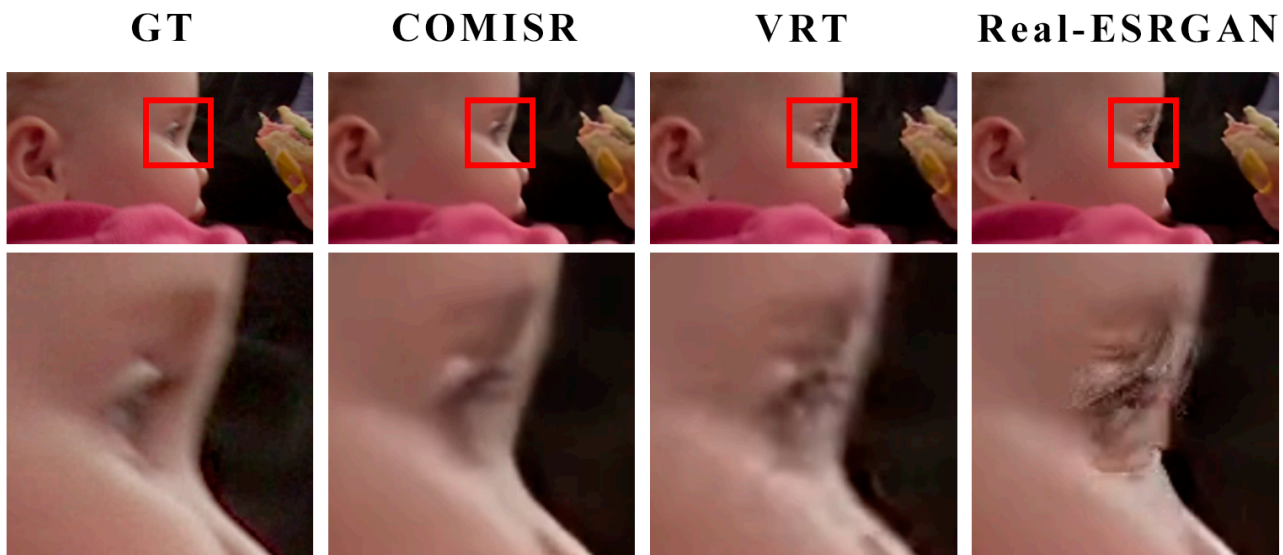


Рисунок 1: Сравнение трех моделей повышения разрешения на сжатом видео.

видео на нашем наборе данных, который включает видео, сжатые 5 различными видеокодеками. Для оценки качества моделей мы использовали объективные метрики и субъективную оценку. Анализируя корреляцию между объективными метриками и субъективными оценками, мы разработали объективную метрику для оценки качества сжатого видео.

Наш основной вклад заключается в следующем:

1. Мы разработали бенчмарк методов суперразрешения для проверки способности моделей повышать качество и восстанавливать видео, сжатое видеокодеками различных стандартов. Бенчмарк находится в открытом доступе:
<https://videoprocessing.ai/benchmarks/super-resolution-for-video-compression.html>.
2. Мы представили анализ 6 метрик качества видео по их корреляции с субъективными оценками на нашем наборе данных.
3. Мы разработали новую объективную метрику для оценки качества сжатого видео, которая имеет самую высокую корреляцию с субъективными оценками среди других метрик. Метрика находится в открытом доступе:
<https://github.com/EvgeneyBogatyrev/super-resolution-metric>.

2 Обзор области

В этом разделе мы представим обзор существующих методов суперразрешения, видеокодеков с понижением разрешения, метрик качества видео и бенчмарков методов повышения разрешения.

2.1 Методы повышения разрешения

Область суперразрешения привлекает внимание ученых с тех пор, как нейронные сети впервые были использованы для решения этой задачи, поэтому существует множество различных подходов. Сначала методы повышения разрешения в основном использовали только пространственную информацию. Ранние методы, такие как ESPCN [45], извлекали признаки из изображений с низким разрешением и использовали фильтры повышения разрешения для получения результатов высокого разрешения.

Некоторые алгоритмы повышения разрешения полагаются на временную избыточность видеок кадров, что позволяет им восстанавливать один кадр высокого разрешения из серии кадров низкого разрешения. DBVSR [37] оценивает оптический поток и размытие видео для компенсации движения между кадрами. RBPN [12] объединяет пространственный и временной контексты видео с помощью рекуррентного кодер-декодер модуля. COMISR [26] специализируется на повышении качества сжатого видео и использует двунаправленную рекуррентную архитектуру для оценки потока с сохранением деталей. Метод BasicVSR++ [5], который является улучшенной версией метода BasicVSR [4], также использует двунаправленное распространение и пространственное выравнивание.

Генеративные адверсарные сети (GAN) широко используются в глубоком обучении и особенно в повышении разрешения. По сравнению с базовой архитектурой SRGAN [24], архитектура ESRGAN [52] была изменена: добавлены Residual-in-Residual Dense блоки, улучшены функции потерь. Real-ESRGAN [53] улучшил этот подход, добавив процесс моделирования деградации высокого порядка для имитации деградации в реальном мире.

Существуют и другие различные подходы к задаче повышения разрешения. VRT [28] извлекает признаки видео, увеличивает их, а затем восстанавливает кадры высокого разрешения на основе этих признаков с помощью трансформаторной сети. TMNet [61] повышает не только разрешение, но и частоту кадров видео, используя временную информацию, интегрированную деформируемой сверткой с многокадровым входом.

2.2 Видеокодеки на основе суперразрешения

Основная идея LCEVC [30] заключается в сжатии видео низкого разрешения с помощью обычного видеокодека и объединении этого видео с признаками исходного видео высокого разрешения. Некоторые видеокодеки используют аналогичную идею реализации сети и применяют метод суперразрешения на стороне декодера.

SRVC [20] кодирует видео в два битовых потока: поток контента и поток модели. Поток контента создается путем сжатия уменьшенного видео низкого разрешения с помощью существующего кодека (например, x265). Поток модели

кодирует обновления для легковесной сети суперразрешения, которая используется для повышения разрешения видео на стороне декодера. Сеть обучается на локальных сегментах видео в процессе кодирования.

RR-DnCNN [14] решает проблему удаления артефактов сжатия во время кодирования видео. Прямой подход к применению методов удаления артефактов сжатия перед суперразрешением может привести к потере деталей видео. Авторы данной работы предложили сквозную глубокую нейронную сеть восстановления и реконструкции, использующую технику, учитывающую деградацию исходного видео.

LiveSRVC [7] кодирует кадры с уменьшением их разрешения, ориентируясь на ключевые кадры видео. В работе описан модуль сжатия кадров на основе суперразрешения и алгоритм адаптации битрейта. Первый применяет адаптивное сжатие для всех кадров, не являющихся ключевыми, второй учитывает динамически изменяющуюся пропускную способность канала для выбора битрейта и степени сжатия кадров. На стороне пользователя применяется повышение разрешения для восстановления уменьшенных кадров.

2.3 Методы оценки качества видео

Пиковое отношение сигнала к шуму (PSNR) и индекс структурного сходства (SSIM)[57] часто используются в различных областях обработки видео для оценки качества алгоритмов, и область суперразрешения не является исключением. PSNR использует среднеквадратичную ошибку и максимальные значения пикселей для расчета отношения сигнала к шуму. SSIM рассчитывает среднее значение, дисперсию и ковариацию пиксельных значений по областям изображения. Существует несколько версий этой метрики, в частности Multiscale SSIM (MS-SSIM) [56], которая проводится по нескольким масштабам с помощью нескольких этапов субдискретизации. Чтобы применить PSNR и SSIM к видео, мы можем усреднить их значения на каждом кадре видео. Эти метрики до сих пор часто используются в работах области повышения разрешения, несмотря на их простые подходы и низкую корреляцию с субъективными оценками [17, 16].

В последнее время в области оценки качества видео все большую популярность набирают подходы глубокого обучения. Примером этому является MDTVSFA [25]. Измерение этой метрики состоит из трех этапов: оценки относительного качества, нелинейного отображения и выравнивания перцептивной шкалы по набору данных для прогнозирования относительного качества, перцептивного качества и субъективного качества соответственно. LPIPS [66] – это метрика, использующая глубокие признаки различных нейронных сетей для сравнения изображений. Эта метрика используется в качестве функции потерь для некоторых моделей суперразрешения [1]. Другим примером является DISTS [8], которая использует сверточную нейронную сеть для извлечения текстовых представлений из изображений, а затем объединяет пространственные

особенности этих текстурных представлений с традиционным структурным сходством.

Существует множество других подходов к задаче оценки качества видео. SFSN [67] рассматривает изображения после повышения разрешения в двумерном пространстве структурной корректности и статистической естественности, чтобы учесть поведение конкретных подходов к повышению разрешения, таких как генеративные состязательные сети. NeuralSBS [21] была разработана специально для задачи повышения разрешения. Эта метрика использует сиамскую нейронную сеть, обученную на выровненных парах изображений. Метками для обучения являются вероятности выбора каждого изображения в паре, основанные на субъективной оценке изображений. Метрика может предсказать, какое из двух изображений предпочтет человек. ERQA [22] – это метрика, которая оценивает изображения с точки зрения восстановления деталей. Для этого определяются края объектов, которые сопоставляются с их аналогами на эталонном изображении.

Все эти метрики применимы к широкому кругу задач, включая оценку качества работы методов повышения разрешения.

2.4 Бенчмарки методов повышения разрешения

В бенчмарке NTIRE 2022 Challenge on Super-Resolution and Quality Enhancement of Compressed Video [64] представлены результаты улучшения качества сжатого видео (трек 1), улучшения качества и повышения разрешения в 2 и в 4 раза (трек 2 и трек 3 соответственно). Всего на всех треках было зарегистрировано более 600 участников. В этом испытании использовался разработанный датасет LDV 2.0 с 335 видео и применялась метрика PSNR для оценки результатов участников.

Vimeo-90k [62] - это большой набор данных высококачественных видео для низкоуровневой обработки видео. Авторы используют этот набор данных для оценки своего алгоритма TOFlow в трех задачах обработки видео: интерполяция кадров, денойзинг видео и повышение разрешения. Для оценки используются метрики PSNR и SSIM.

Датасет RealVSR [65] состоит из 500 реальных пар видео высокого и низкого разрешения, снятых с помощью многокамерной системы iPhone 11 Pro Max. Для борьбы с параллаксом и цветовыми различиями между изображениями, вызванными использованием двух отдельных камер, использовалась пирамида Лапласиана. Для объективной оценки используются метрики PSNR, SSIM, NIQE [32] и BRISQUE [31].

Наш бенчмарк отличается от всех представленных выше бенчмарков более широким набором метрик качества видео и разнообразием видеокодеков, используемых для создания искаженного видео.

3 Бенчмарк

В этом разделе описан наш бенчмарк Super-Resolution for Video Compression. Описывается методология бенчмарка и делается обзор результатов оценки моделей суперразрешения. Наш бенчмарк находится в открытом доступе: <https://videoprocessing.ai/benchmarks/super-resolution-for-video-compression.html>. Мы приглашаем исследователей методов суперразрешения принять участие в нашем бенчмарке.

3.1 Подготовка датасета

Чтобы убедиться, что датасет для бенчмарка достаточно разнообразен для тестирования различных аспектов моделей суперразрешения, мы собрали видео из разных источников:

- **Видео с Vimeo:** Мы собрали 50 видео разрешения FullHD с Vimeo. Эти видео включают как реальные, так и анимационные клипы. Мы разделили эти ролики на сцены с помощью плагина Scene Change Detector [33] для VQMT [34].
- **Видео с камеры:** Мы сняли несколько видео на камеру Canon EOS 7D. Настройки камеры были выставлены так, чтобы минимизировать размытие кадров и установить соответствующую яркость. Мы сняли 20 видео в помещении и 30 видео на открытом воздухе. Видео в помещении состоят из синтетически созданных сцен с предметами из повседневной жизни. Каждая сцена содержит либо движущиеся объекты, либо параллельное движение камеры.
- **Игровые видео:** Мы записали 20 клипов из различных 2D и 3D видеоигр.

Затем мы получили следующие характеристики для каждого видео: характеристики Google SI/TI [54], количество кадров в секунду (FPS), красочность [13] и максимальное количество лиц на видео [10]. На основе этих характеристик мы разделили все видео на 9 кластеров с помощью модели K-Means из scikit-learn и выбрали по одному видео из каждого кластера.

Чтобы гарантировать, что детали видео не потеряются после сжатия, мы предпочли видео с низкой пространственной и временной сложностью, однако движение камеры было обязательным. При таких ограничениях получается видео, которое не теряет важные детали при уменьшении масштаба. Движение камеры также может помочь моделям, работающим с видео, восстановить каждый кадр по соседним кадрам.

Мы будем называть эти 9 отобранных видео *исходными видео*. Кадры из этих видео можно увидеть на рисунке 2.

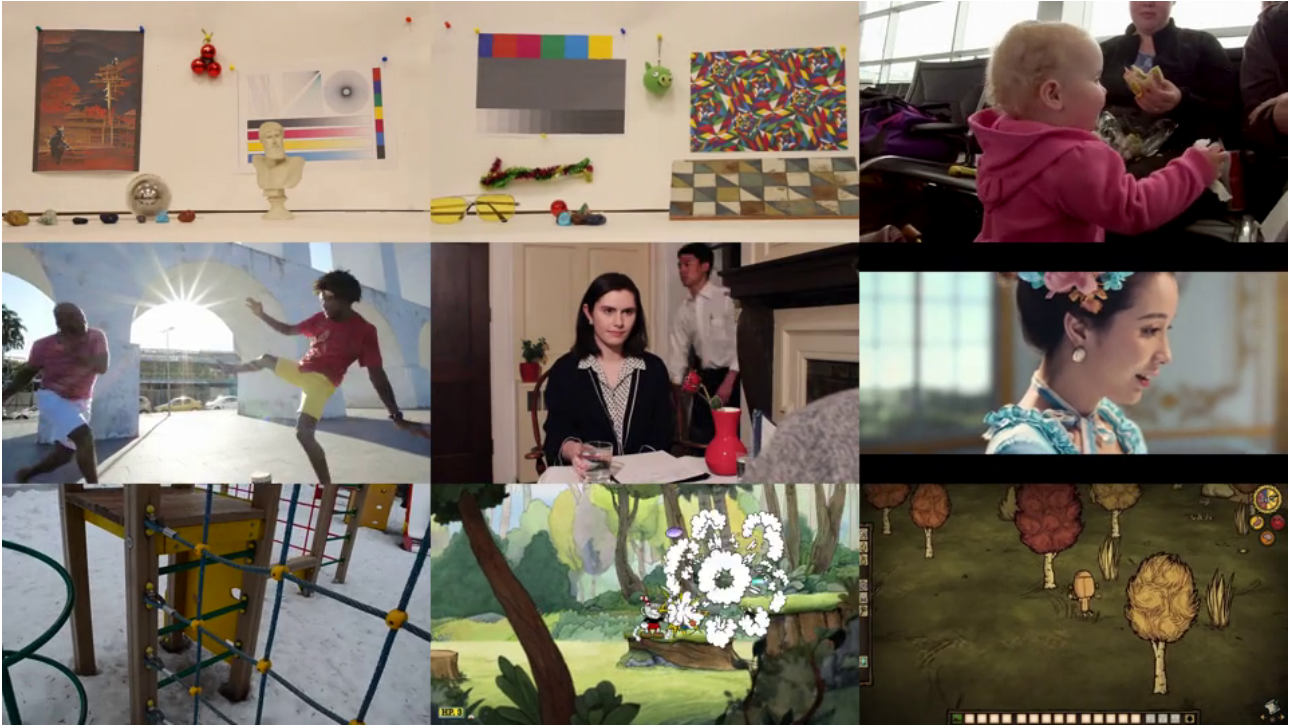


Рисунок 2: Примеры видео из датасета бенчмарка. Датасет включает в себя реалистичные видео, анимацию и клипы из видеоигр.

3.2 Модели суперразрешения

Чтобы выбрать модели суперразрешения для сравнения, мы использовали существующие бенчмарки, нацеленные на различные задачи: восстановление деталей [35] и качество восприятия [36]. Мы также исключили похожие методы суперразрешения, ориентируясь на схожие значения метрик.

Кодек	Стандарт кодирования
x264	H.264
x265	H.265
aomenc	AV1
vvenc [58]	H.266
uavs3e [48]	AVS3

Таблица 1: Видеокодеки, которые использовались при создании датасета бенчмарка. Мы использовали пресет *medium* для всех этих кодеков.

Мы отобрали следующие 17 методов: BasicVSR++ [5], COMISR [26], DBVSR [37], EGVSR [3], LGFN [46], RBPN [12], Real-ESRGAN [53], RealSR [19], RSDN [18], SOF-VSR-BD [51], SOF-VSR-BI [51], SwinIR [27], TMNet [61], VRT [28], ahq-11 [47], amq-12 [47] и бикубическая интерполяция.

Метрика	PLCC	SRCC
MS-SSIM [56]	0.146	0.151
PSNR	0.187	0.285
VMAF [49]	0.344	0.448
LPIPS [66]	0.414	0.431
ERQA [22]	0.582	0.624
MDTVSFA [25]	0.634	0.644
ERQAхMDTVSFA	0.770	0.801
Разработанная метрика	0.821	0.838

Таблица 2: Средние корреляции Пирсона (PLCC) и Спирмена (SRCC) метрик с субъективными оценками.

3.3 Объективная оценка качества

Сначала мы понизили разрешение исходного видео до 480×270 с помощью FFmpeg. Затем мы сжали видео низкого разрешения, используя каждый из 5 видеокодеков с 6 различными целевыми битрейтами: 0.1, 0.3, 0.6, 1, 2 и 4 Мбит/с. Мы выбрали эти битрейты, потому что хотели, чтобы они были относительно низкими, а построенные по ним Rate-Distortion (RD) кривые имели логарифмическую структуру. Выбранные кодеки показаны в таблице 1. Все кодеки использовались с пресетом *medium* во время сжатия. Сжатое видео транскодировалось в последовательности PNG кадров с помощью FFmpeg и подавалось на вход моделям повышения разрешения. Модели для изображений применялись к каждому кадру отдельно, а моделям для видео мы давали путь к директории с кадрами, расположенными в правильном порядке.

В нашем бенчмарке мы тестируем четырехкратное повышение разрешения, однако есть некоторые модели суперразрешения, которые могут повышать разрешение только в 2 раза. В этом случае мы применяли их дважды.

После повышения разрешения видео мы рассчитывали следующие объективные метрики качества видео: PSNR, MS-SSIM, VMAF, LPIPS и ERQA. Для ранжирования методов суперразрешения мы рассчитывали BSQ-rate (bitrate-for-the-same-quality rate) [68] каждой пары метода суперразрешения и кодека относительно базового кодека, где базовый кодек – это кодек, который использовался для сжатия видео низкого разрешения.

В ходе сравнения видеокодеков существует ряд случаев, в которых BD-rate невозможно точно посчитать. Наиболее распространенной является ситуация, когда две RD-кривые не пересекаются по значениям метрики для двух разных кодеков. Для решения этой проблемы используется BSQ-rate вместо BD-rate. Ключевая идея расчета BSQ-rate в том, чтобы инвертировать оси битрейта и качества перед расчетом площади под кривыми. Более низкий BSQ-rate соответствует большей экономии битрейта при том же качестве. Более подробную

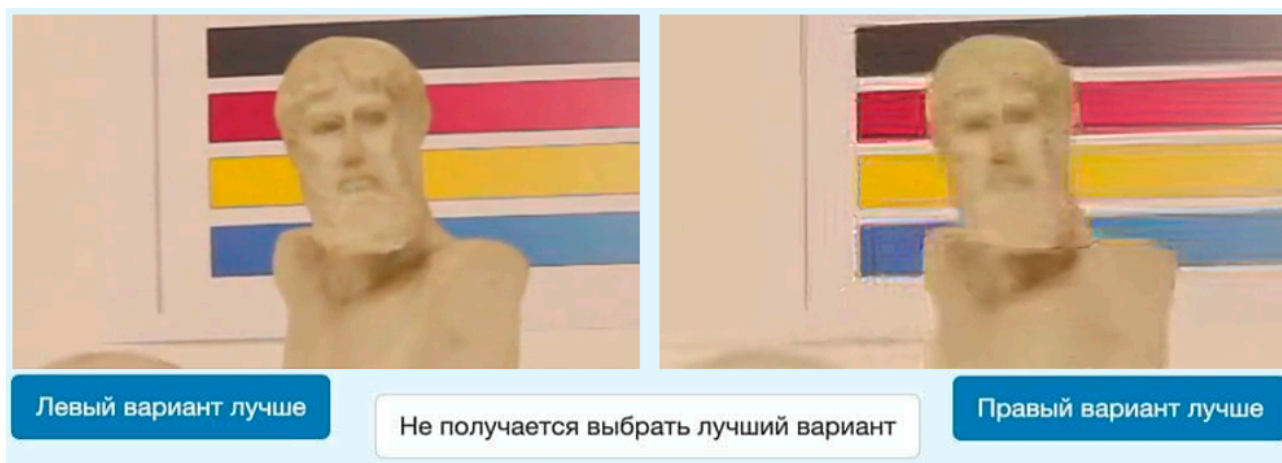


Рисунок 3: Интерфейс сервиса Subjectify.us. Участники могут выбрать левое или правое видео. Также есть опция ”Не получается выбрать лучший вариант”.

информацию о BSQ-rate вы можете найти в статье [68].

3.4 Субъективное сравнение

Для расчета субъективных оценок моделей суперразрешения мы провели субъективное сравнение с помощью сервиса Subjectify.us. Поскольку потерю деталей и артефакты сжатия трудно заметить на полном кадре, мы решили провести субъективное сравнение на сегментах видео. Сначала мы сгенерировали карты салиентности для каждого исходного видео, используя метод, предложенный в статье [23]. Затем мы усреднили карты салиентности по всем кадрам и применили ядро размытия Гаусса к результату, чтобы определить наиболее салиентную область видео. После этого мы взяли видео после применения методов суперразрешения и вырезали из каждого видео один сегмент размером 480×270 , используя наиболее салиентную область в качестве центра сегмента. Мы проводили сравнение на видео, сжатых с целевыми битрейтами: 0.6, 1 и 2 Мбит/с. Битрейты были выбраны таким образом, чтобы сохранить логарифмическую структуру RD кривой. Разрешение сегментов было установлено так, чтобы оно умещалось на экранах устройств, популярных среди участников субъективного сравнения. Для расчета корреляции с субъективными оценками объективные метрики были также посчитаны на выбранных сегментах.

Мы разделили наше сравнение на 5 сравнений для каждого кодека. В каждом сравнении мы использовали только 10 лучших моделей SR, выбранных по лучшим значениям метрики LPIPS: SwinIR [27], COMISR [26], RealSR [19], Real-ESRGAN [53], VRT [28], BasicVSR++ [5], ahq-11[47], SOF-VSR [51] и RBPN [12].

Мы показывали каждому участнику пару видео с двумя случайными моделями сверхразрешения и просили выбрать видео, которое выглядит более реалистично и имеет меньше артефактов сжатия. Каждая пара видео была показана 15 участникам. Каждый участник сравнивал 25 пар видео. На рисунке 3 показан

SR \ кодек	x264	x265	aomenc	vvenc	uavs3e
No SR	1.000	1.000	1.000	<u>1.000</u>	<i>1.000</i>
RealSR	0.196	<u>0.502</u>	<i>1.513</i>	4.470	0.639
ahq-11	<u>0.271</u>	0.724	2.806	<i>1.665</i>	1.750
SwinIR	<i>0.304</i>	0.346	<u>1.505</u>	2.502	<u>0.640</u>
Real-ESRGAN	0.335	<i>0.640</i>	2.411	4.368	2.43
COMISR	0.367	0.741	2.799	0.701	1.603
VRT	1.245	3.175	4.157	4.185	3.663
SOF-VSR	1.273	3.396	4.366	4.349	4.772
BasicVSR++	1.971	3.390	4.199	4.185	4.250
RBPN	1.979	3.434	4.226	4.246	4.470

Таблица 3: Среднее значение BSQ-rate [68], посчитанное на субъективных оценках для каждой пары метода повышения разрешения (SR) и кодека. Лучший результат выделен **жирным**, второй лучший подчеркнут, третий лучший выделен *курсивом*.

SR + кодек	Суб. оценка ↑	ERQA ↑	LPIPS ↓	PSNR ↑
SwinIR + x264	5.855	0.601	0.237	24.961
RealSR + x264	<u>5.838</u>	0.565	0.268	25.449
Real-ESRGAN + x264	<i>5.142</i>	0.560	0.238	25.083
ahq-11 + x264	5.049	0.579	0.217	<u>26.209</u>
COMISR + x264	4.966	0.550	0.256	24.417
SwinIR + x265	4.801	0.585	0.231	25.034
RealSR + x265	4.738	0.584	0.260	25.519
Real-ESRGAN + x265	4.312	0.576	0.232	25.113
SwinIR + uavs3e	4.206	<i>0.597</i>	0.228	24.954
SwinIR + aomenc	3.843	<u>0.598</u>	0.198	25.24
comisr + x265	3.794	0.568	0.242	24.393
ahq-11 + x265	3.785	0.596	<u>0.210</u>	26.256
SwinIR + vvenc	3.732	0.557	<i>0.214</i>	25.152
RealSR + aomenc	3.694	0.562	0.219	<i>25.760</i>

Таблица 4: Сравнение пар методов суперразрешения (SR) и кодеков по субъективной оценке и объективным метрикам на видео "restaurant". Лучший результат выделен **жирным**, второй лучший подчеркнут, третий лучший выделен *курсивом*.

интерфейс сервиса Subjectify.us.

Среди этих 25 вопросов было 3 проверочных вопроса с очевидным предопределенным ответом. Эти вопросы показываются в случайный момент и поз-

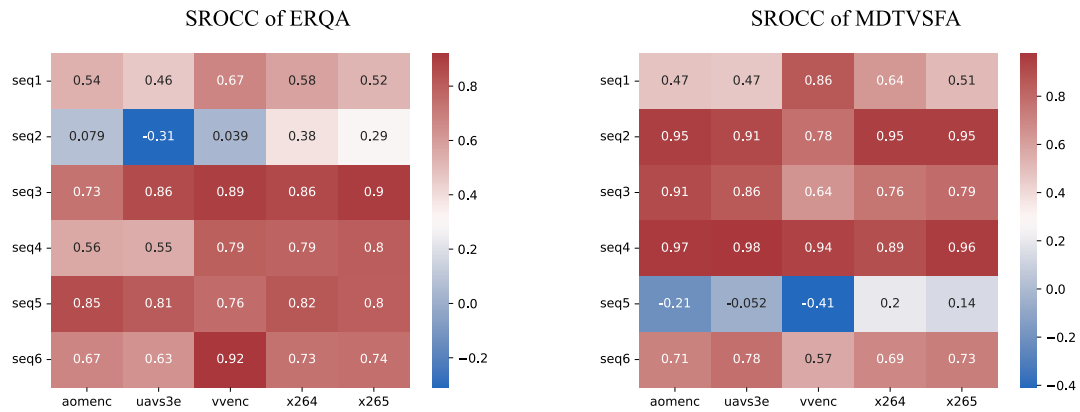


Рисунок 4: Корреляция Спирмена метрик ERQA и MDTVSA с субъективными оценками на каждом тестовом видео. В строках указаны названия исходного видео, а в колонках – кодеки, с помощью которых эти видео были закодированы. Случаи, когда метрика показывала высокую корреляцию, выделены красным, а случаи с низкой корреляцией – синим.

воляют предотвратить атаки ботов и случайные ответы участников. Если пользователь не отвечает на проверочный вопрос, то все его ответы исключаются.

В субъективном сравнении приняли участие 5662 человека. 265 из них не смогли правильно ответить на проверочные вопросы, поэтому их ответы были исключены. Оставшиеся 120316 ответов были использованы для расчета итоговых субъективных оценок с помощью модели Брэдли-Терри [2].

3.5 Результаты бенчмарка

Результаты субъективной оценки каждой пары метода суперразрешения и кодека представлены в таблице 3. Как мы видим, лучшая модель суперразрешения различна для каждого кодека, что доказывает, что не существует универсальной модели суперразрешения, которая одинаково эффективно справляется с искажениями всех стандартов сжатия видео.

No SR – это метод, который был получен применением видеокodeка к исходному видео без понижения разрешения и последующего суперразрешения. Этот метод используется в качестве эталонного при расчете BSQ-rate. Мы видим, что метод *No SR* показывает лучшие результаты для кодека aomenc. Причина этого в том, что кодеки стандарта AV1 имеют режим, который кодирует кадры с низким разрешением и применяет метод суперразрешения после декодирования. Этот режим обычно используется при низком битрейте, что очень похоже на процесс, описанный в методологии нашего бенчмарка.

Общий рейтинг пар методов суперразрешения и кодеков можно увидеть в таблице 4. В ней показано 14 лучших пар, основанных на субъективных оценках. Корреляцию Пирсона (PLCC) и Спирмена (SRCC) объективных метрик с субъективными оценками можно увидеть в таблице 2.

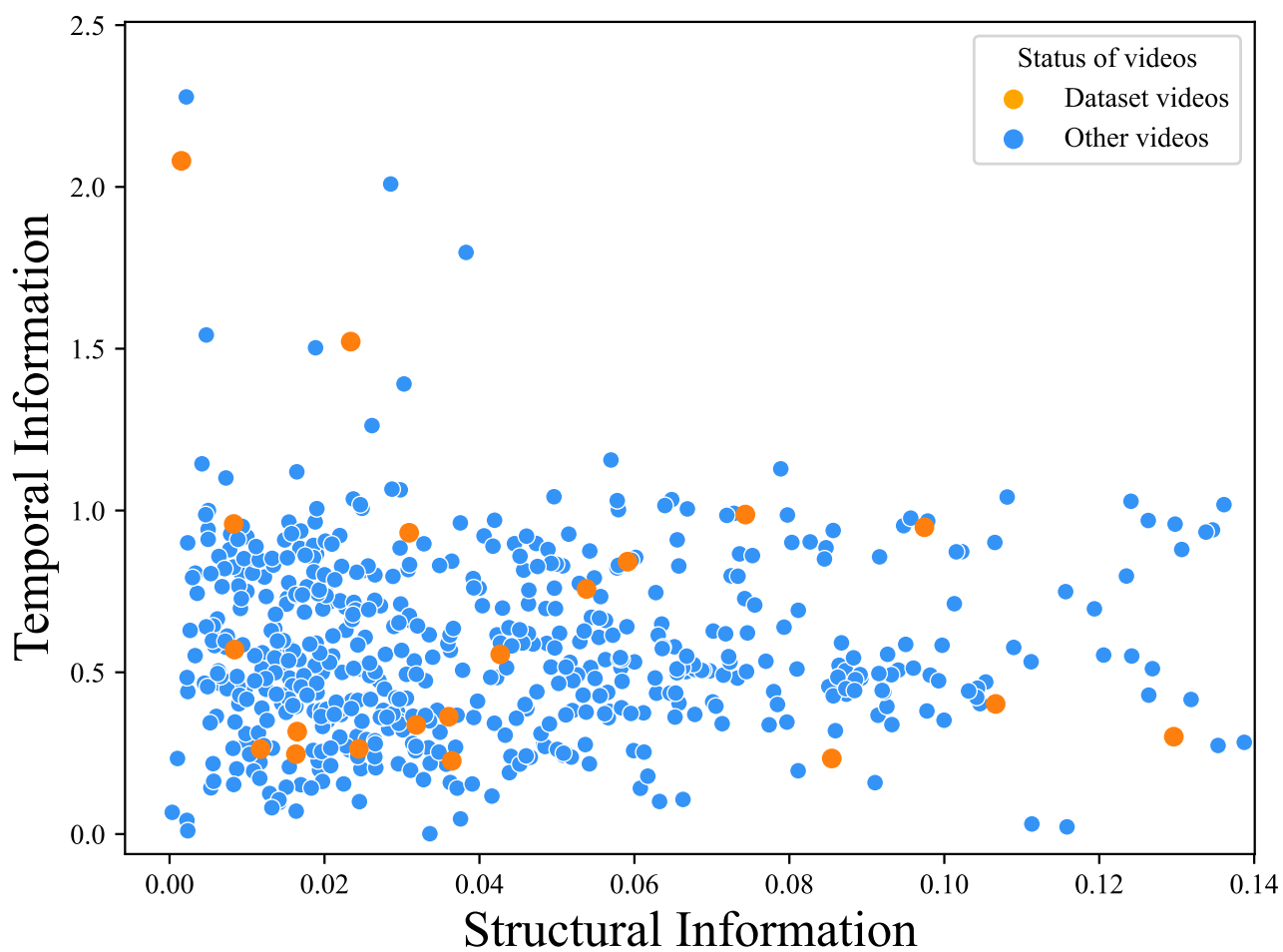


Рисунок 5: Распределение признаков Google SI/TI [54] для видео, которые рассматривались при создании датасета. Выбранные видео показаны оранжевым цветом, а остальные – синим.

4 Метрика качества

Из результатов субъективного сравнения в бенчмарке (2) мы видим, что существующие метрики качества видео плохо коррелируют с субъективными оценками и не подходят для оценки результатов кодирования видео в комбинации с методами суперразрешения.

4.1 Базовый подход

Мы проанализировали, как каждая из метрик работает на каждом отдельном видео и с каждым кодеком. Мы заметили, что ERQA показывает высокую корреляцию на видео, где MDTVSFA имеет низкую корреляцию, и наоборот (рисунок 4). Это означает, что эти две метрики учитывают различные особенности входного видео и могут дополнять друг друга. Поскольку значения ERQA варьируются от 0 до 1, ERQA можно использовать как корректирующее значение для MDTVSFA. Мы попытались объединить метрики ERQA и MDTVSFA, перемножив их результаты. Мы назвали этот метод ERQAхMDTVSFA. Этот подход дал значительный прирост корреляции, как видно из таблицы 2.

Мы расширили наш датасет, чтобы получить больше данных для обучения метрики. Для отбора видео мы использовали ту же методику, что описана в разделе 3.1, однако на этот раз мы собрали 20 новых исходных видео. Распределение SI/TI характеристик видео в наборе данных показано на рисунке 5.

Мы выбрали 4 различных метода суперразрешения (RealSR, Real-ESRGAN, COMISR, BasicVSR++) и 3 видеокodeка (x264, x265, aomenc) для создания искаженных видео. Затем мы уменьшили исходные видео и сжали их с помощью каждого из кодеков с двумя целевыми битрейтами: 0.5 Мбит/с и 2 Мбит/с. После этого мы применили каждый выбранный метод суперразрешения к каждому видео, чтобы создать набор искаженных видео. Мы также сжали исходные видео без понижения разрешения и добавили результаты в набор данных.

Мы провели субъективное сравнение, как в разделе 3.4. Так как субъективные оценки имеют разный масштаб на разных видео, мы изменили их масштаб от 0 до 1 на каждом видео.

4.2 Предлагаемый подход

Хотя отдельные метрики не показывают хороших результатов для поставленной задачи, MDTVSFA, ERQA и LPIPS имеют самую высокую корреляцию среди остальных. Мы решили объединить эти метрики для достижения более высокой корреляции. Мы рассчитали значения MDTVSFA, ERQA и LPIPS на нашем наборе искаженных данных. Чтобы обобщить информацию о видео, мы также рассчитали некоторые другие характеристики каждого видео. Информацию о характеристиках можно посмотреть в таблице 5.

Признак	Описание
MDTVSFA	Значение метрики MDTVSF A
ERQA	Значение метрики ERQA
LPIPS	Значение метрики LPIPS
ERQAxMDTVSFA	$MDTVSF A \times ERQA$
ERQAxLPIPS	$(1 - LPIPS) \times ERQA$
SI/TI	Google SI/TI [54] признаки
Красочность	Красочность [13], усредненная по кадрам видео.
Битрейт	Битрейт входного видео.

Таблица 5: Признаки, на которых обучалась разработанная метрика.

Метрика	Wireless	H.264	MPEG-2	Все данные
PSNR	0.4334	0.4296	0.3588	0.3684
SSIM	0.5233	0.6514	0.5545	0.5257
Speed SSIM [55]	0.5630	0.7086	0.6185	0.5849
VSNR [6]	0.7019	0.6460	0.5915	0.6755
V-VIF [44]	0.5507	0.6807	0.6116	0.5710
Spatial MOVIE [40, 41]	0.7927	0.7066	0.6911	0.7270
Temporal MOVIE [40, 41]	0.8114	<u>0.7797</u>	0.8170	0.8055
MOVIE [40, 41]	<u>0.8109</u>	0.7664	<u>0.7733</u>	0.7890
Разработанная метрика	0.7934	0.8465	0.7554	<u>0.7946</u>

Таблица 6: Сравнение корреляции Спирмена метрик качества видео. Лучший результат выделен **жирным**, второй лучший подчеркнут, третий лучший выделен *курсивом*. Большая часть данных взята из [42].

Для нормализации данных перед обучением мы использовали MinMaxScaler из scikit-learn. Затем мы разделили все видео на искаженные наборы данных с помощью 3-кратной перекрестной валидации. Для обучения метрики мы выбрали модель SVR с линейным ядром из scikit-learn. Мы выбрали эту модель, потому что она проста в обучении и показала лучшие результаты по сравнению с другими моделями регрессии.

Чтобы оптимизировать время работы метрики, мы решили сократить количество признаков. Мы попробовали обучить метрику без каждой возможной пары признаков и посмотреть на значения корреляции. Для каждой пары мы вычислили наихудшее значение SRCC на валидационном наборе. Метрика показала самую высокую корреляцию без признаков, которые являются значениями метрики MDTVSF A и битрейтом входного видео. Удаление других признаков показало значительно более высокие потери корреляции.

В окончательной версии метрики используются следующие признаки:

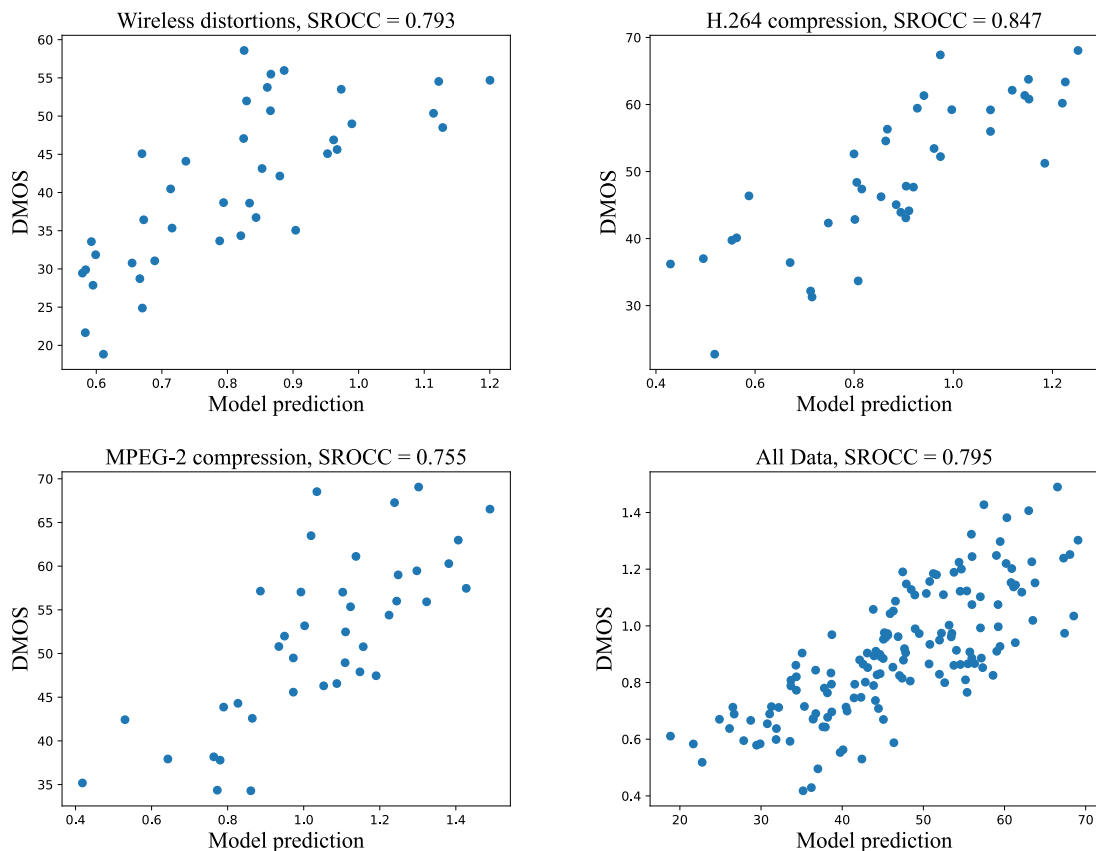


Рисунок 6: Распределения значений метрики и значений DMOS для датасета LIVE VQ Database.

ERQA, LPIPS, ERQAxLPIPS, ERQAxMDTVSFA, SI, TI, красочность.

4.3 Эксперименты

Мы протестировали разработанную метрику на нашем бенчмарке, где она превзошла другие метрики по корреляции Пирсона и Спирмена, как показано в таблице 2.

Мы также оценили предложенную метрику на датасете LIVE Video Quality Database [42, 43]. В таблице 6 показаны значения корреляции Спирмена предложенной метрики со значениями DMOS для каждого из 4 разделов датасета и общую корреляцию Спирмена метрики на всем датасете. На рисунке 6 показаны распределения значений метрики и оценок DMOS в каждом разделе датасета.

Мы видим, что предложенная метрика показывает самую высокую корреляцию в разделе, где видео были сжаты с помощью кодека H.264, и вторую по величине корреляцию в целом. Корреляция в разделе, где видео были сжаты с помощью кодека MPEG-2, не так высока. Скорее всего, это произошло потому, что в обучающем наборе не было видео, закодированных с использованием этого стандарта кодирования.

5 Вывод

В этой статье мы предложили новый бенчмарк для методов суперразрешения для восстановления сжатых видео. Мы оценили 17 моделей суперразрешения, примененных к видео, закодированных с помощью 5 кодеков. Оценка проводилась с помощью объективных метрик качества видео и субъективной оценки. Мы разработали новую метрику качества, которая имеет лучшие корреляции с субъективными оценками, чем другие метрики, на наборе данных нашего бенчмарка и более близка к человеческому восприятию качества сжатых видео.

Наше исследование показывает, что некоторые модели суперразрешения, такие как RealSR [19] и SwinIR [27], могут быть использованы в видеокодеках на стороне декодера для улучшения качества видео при кодировании с низкими битрейтами.

Список литературы

- [1] Andreas Aakerberg, Kamal Nasrollahi and Thomas B Moeslund. “Real-world super-resolution of face-images from surveillance cameras”. In: *IET Image Processing* 16.2 (2022), p. 442—452.
- [2] Ralph Allan Bradley and Milton E Terry. “Rank analysis of incomplete block designs: I. The method of paired comparisons”. In: *Biometrika* 39.3/4 (1952), p. 324—345.
- [3] Yanpeng Cao et al. “Real-time super-resolution system of 4k-video based on deep learning”. In: *2021 IEEE 32nd International Conference on Application-specific Systems, Architectures and Processors (ASAP)*. IEEE. 2021, p. 69—76.
- [4] Kelvin CK Chan et al. “BasicVSR: The search for essential components in video super-resolution and beyond”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, p. 4947—4956.
- [5] Kelvin CK Chan et al. “BasicVSR++: Improving video super-resolution with enhanced propagation and alignment”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, p. 5972—5981.
- [6] Damon M Chandler and Sheila S Hemami. “VSNR: A wavelet-based visual signal-to-noise ratio for natural images”. In: *IEEE transactions on image processing* 16.9 (2007), p. 2284—2298.

- [7] Ying Chen et al. “Higher quality live streaming under lower uplink bandwidth: an approach of super-resolution based video coding”. In: *Proceedings of the 31st ACM Workshop on Network and Operating Systems Support for Digital Audio and Video*. 2021, p. 74—81.
- [8] Keyan Ding et al. “Image Quality Assessment: Unifying Structure and Texture Similarity”. In: *CoRR* abs/2004.07728 (2020). URL: <https://arxiv.org/abs/2004.07728>.
- [9] Lyndon R Duong et al. “Multi-rate adaptive transform coding for video compression”. In: *arXiv preprint arXiv:2210.14308* (2022).
- [10] *Face Recognition by ageitgey*. https://github.com/ageitgey/face_recognition. Online; accessed 2022-11-08.
- [11] Hadi Hadizadeh and Ivan V Bajić. “LCCM-VC: Learned Conditional Coding Modes for Video Coding”. In: *arXiv preprint arXiv:2210.15883* (2022).
- [12] Muhammad Haris, Gregory Shakhnarovich and Norimichi Ukita. “Recurrent back-projection network for video super-resolution”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019, p. 3897—3906.
- [13] David Hasler and Sabine E Suesstrunk. “Measuring colorfulness in natural images”. In: *Human vision and electronic imaging VIII*. T. 5007. SPIE. 2003, p. 87—95.
- [14] Man M Ho, Jinjia Zhou and Gang He. “RR-DnCNN v2. 0: enhanced restoration-reconstruction deep neural network for down-sampling-based video coding”. In: *IEEE Transactions on Image Processing* 30 (2021), p. 1702—1715.
- [15] Yung-Han Ho et al. “Learned video compression for yuv 4: 2: 0 content using flow-based conditional inter-frame coding”. In: *arXiv preprint arXiv:2210.08225* (2022).
- [16] Quan Huynh-Thu and Mohammed Ghanbari. “Scope of validity of PSNR in image/video quality assessment”. In: *Electronics letters* 44.13 (2008), p. 800—801.
- [17] Quan Huynh-Thu and Mohammed Ghanbari. “The accuracy of PSNR in predicting video quality for different video scenes and frame rates”. In: *Telecommunication Systems* 49.1 (2012), p. 35—48.
- [18] Takashi Isobe et al. “Video super-resolution with recurrent structure-detail network”. In: *European conference on computer vision*. Springer. 2020, p. 645—660.
- [19] Xiaozhong Ji et al. “Real-world super-resolution via kernel estimation and noise injection”. In: *proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*. 2020, p. 466—467.

- [20] Mehrdad Khani, Vibhaalakshmi Sivaraman and Mohammad Alizadeh. “Efficient Video Compression via Content-Adaptive Super-Resolution”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, p. 4521—4530.
- [21] Valentin Khrulkov and Artem Babenko. “Neural Side-by-Side: Predicting Human Preferences for No-Reference Super-Resolution Evaluation”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Jun 2021, p. 4988—4997.
- [22] Anastasia Kirillova et al. “ERQA: Edge-restoration quality assessment for video super-resolution”. In: *arXiv preprint arXiv:2110.09992* (2021).
- [23] Alexander Kroner et al. “Contextual encoder–decoder network for visual saliency prediction”. In: *Neural Networks* 129 (2020), p. 261—270. ISSN: 0893-6080. DOI: <https://doi.org/10.1016/j.neunet.2020.05.004>. URL: <https://www.sciencedirect.com/science/article/pii/S0893608020301660>.
- [24] Christian Ledig et al. “Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Jul 2017.
- [25] Dingquan Li, Tingting Jiang and Ming Jiang. “Unified quality assessment of in-the-wild videos with mixed datasets training”. In: *International Journal of Computer Vision* 129.4 (2021), p. 1238—1257.
- [26] Yinxiao Li et al. “Comisr: Compression-informed video super-resolution”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, p. 2543—2552.
- [27] Jingyun Liang et al. “SwinIR: Image restoration using swin transformer”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, p. 1833—1844.
- [28] Jingyun Liang et al. “Vrt: A video restoration transformer”. In: *arXiv preprint arXiv:2201.12288* (2022).
- [29] Zhisheng Lu et al. “Transformer for single image super-resolution”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, p. 457—466.
- [30] Guido Meardi et al. “MPEG-5 part 2: Low complexity enhancement video coding (LCEVC): Overview and performance evaluation”. In: *Applications of Digital Image Processing XLIII* 11510 (2020), p. 238—257.
- [31] Anish Mittal, Anush Krishna Moorthy and Alan Conrad Bovik. “No-Reference Image Quality Assessment in the Spatial Domain”. In: *IEEE Transactions on Image Processing* 21.12 (2012), p. 4695—4708. DOI: 10.1109/TIP.2012.2214050.

- [32] Anish Mittal, Rajiv Soundararajan and Alan C. Bovik. “Making a “Completely Blind” Image Quality Analyzer”. In: *IEEE Signal Processing Letters* 20.3 (2013), p. 209—212. DOI: 10.1109/LSP.2012.2227726.
- [33] *MSU Scene Change Detector*. https://www.compression.ru/video/quality_measure/metric_plugins/scd_en.htm. Online; accessed 2022-11-08.
- [34] *MSU Video Quality Measurement Tool*. https://www.compression.ru/video/quality_measure/video_measurement_tool.html. Online; accessed 2022-11-08.
- [35] *MSU Video Super-Resolution Benchmark: Detail Restoration*. <https://videoprocessing.ai/benchmarks/video-super-resolution.html>. Online; accessed 2022-11-08.
- [36] *MSU Video Upscalers Benchmark 2022: Quality Enhancement*. <https://videoprocessing.ai/benchmarks/video-upscalers.html>. Online; accessed 2022-11-08.
- [37] Jinshan Pan et al. “Deep blind video super-resolution”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, p. 4811—4820.
- [38] Chitwan Saharia et al. “Image super-resolution via iterative refinement”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2022).
- [39] *Screen Resolution Stats Worldwide*. <https://gs.statcounter.com/screen-resolution-stats>. Online; accessed 2022-11-08.
- [40] Kalpana Seshadrinathan and Alan C Bovik. “Motion-based perceptual quality assessment of video”. In: *Human Vision and Electronic Imaging XIV*. T. 7240. SPIE. 2009, p. 283—294.
- [41] Kalpana Seshadrinathan and Alan Conrad Bovik. “Motion tuned spatio-temporal quality assessment of natural videos”. In: *IEEE transactions on image processing* 19.2 (2009), p. 335—350.
- [42] Kalpana Seshadrinathan et al. “A subjective study to evaluate video quality assessment algorithms”. In: *Human Vision and Electronic Imaging XV*. T. 7527. SPIE. 2010, p. 128—137.
- [43] Kalpana Seshadrinathan et al. “Study of subjective and objective quality assessment of video”. In: *IEEE transactions on Image Processing* 19.6 (2010), p. 1427—1441.
- [44] Hamid R Sheikh and Alan C Bovik. “A visual information fidelity approach to video quality assessment”. In: *The first international workshop on video processing and quality metrics for consumer electronics*. T. 7. 2. sn. 2005, p. 2117—2128.

- [45] Wenzhe Shi et al. “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, p. 1874—1883.
- [46] Dewei Su et al. “Local-global fusion network for video super-resolution”. In: *IEEE Access* 8 (2020), p. 172443—172456.
- [47] *Topaz Video Enhance AI*. <https://www.topazlabs.com/topaz-video-ai>. Online; accessed 2022-11-08.
- [48] *uavs3e*. <https://github.com/uavs3/uavs3e>. Online; accessed 2022-11-08.
- [49] *VMAF - Video Multi-Method Assessment Fusion*. <https://github.com/Netflix/vmaf>. Online; accessed 2022-08-06.
- [50] *VNI Complete Forecast Highlights*. https://www.cisco.com/c/dam/m/en_us/solutions/service-provider/vni-forecast-highlights/pdf/Global_Device_Growth_Traffic_Profiles.pdf. Online; accessed 2022-11-08.
- [51] Longguang Wang et al. “Learning for video super-resolution through HR optical flow estimation”. In: *Asian Conference on Computer Vision*. Springer. 2018, p. 514—529.
- [52] Xintao Wang et al. “ESRGAN: Enhanced super-resolution generative adversarial networks”. In: *The European Conference on Computer Vision Workshops (ECCVW)*. Sep 2018.
- [53] Xintao Wang et al. “Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data”. In: *International Conference on Computer Vision Workshops (ICCVW)*. 2021.
- [54] Yilin Wang, Sasi Inguva and Balu Adsumilli. “YouTube UGC dataset for video compression research”. In: *2019 IEEE 21st International Workshop on Multimedia Signal Processing (MMSP)*. IEEE. 2019, p. 1—5.
- [55] Zhou Wang and Qiang Li. “Video quality assessment using a statistical model of human visual speed perception”. In: *JOSA A* 24.12 (2007), B61—B69.
- [56] Zhou Wang, Eero P Simoncelli and Alan C Bovik. “Multiscale structural similarity for image quality assessment”. In: *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*. T. 2. Ieee. 2003, p. 1398—1402.
- [57] Zhou Wang et al. “Image quality assessment: from error visibility to structural similarity”. In: *IEEE transactions on image processing* 13.4 (2004), p. 600—612.

- [58] Adam Wieckowski et al. “VVenC: An Open And Optimized VVC Encoder Implementation”. In: *Proc. IEEE International Conference on Multimedia Expo Workshops (ICMEW)*. 2021, p. 1—2. DOI: 10 . 1109 / ICMEW53276 . 2021 . 9455944.
- [59] Bartłomiej Wronski et al. “Handheld Multi-Frame Super-Resolution”. In: *ACM Trans. Graph.* 38.4 (Jul 2019). ISSN: 0730-0301. DOI: 10 . 1145 / 3306346 . 3323024. URL: <https://doi.org/10.1145/3306346.3323024>.
- [60] Yanze Wu et al. “AnimeSR: Learning Real-World Super-Resolution Models for Animation Videos”. In: *arXiv preprint arXiv:2206.07038* (2022).
- [61] Gang Xu et al. “Temporal modulation network for controllable space-time video super-resolution”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, p. 6388—6397.
- [62] Tianfan Xue et al. “Video enhancement with task-oriented flow”. In: *International Journal of Computer Vision* 127.8 (2019), p. 1106—1125.
- [63] Ren Yang et al. “Aim 2022 challenge on super-resolution of compressed image and video: Dataset, methods and results”. In: *arXiv preprint arXiv:2208.11184* (2022).
- [64] Ren Yang et al. “NTIRE 2022 challenge on super-resolution and quality enhancement of compressed video: Dataset, methods and results”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022, p. 1221—1238.
- [65] Xi Yang et al. “Real-world video super-resolution: A benchmark dataset and a decomposition based learning scheme”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021, p. 4781—4790.
- [66] Richard Zhang et al. “The unreasonable effectiveness of deep features as a perceptual metric”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018, p. 586—595.
- [67] Wei Zhou, Zhou Wang and Zhibo Chen. “Image Super-Resolution Quality Assessment: Structural Fidelity Versus Statistical Naturalness”. In: *arXiv preprint arXiv:2105.07139* (2021).
- [68] Anastasia V Zvezdakova et al. “BSQ-rate: a new approach for video-codec performance comparison and drawbacks of current solutions”. In: *Programming and computer software* 46.3 (2020), p. 183—194.

Оглавление

Введение	3
Обзор области	4
Бенчмарк.....	8
Метрика качества	15
Вывод.....	18
Список литературы.....	18