



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

Препринты ИПМ • Препринт № 86 за 2022 г.

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

М.В. Синюков, Д.Л. Куликов,
Д.С. Ватолин, [В.А. Галактионов](#)

Ограничения применимости
дифференцируемых
эталонных показателей
качества изображений

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



Рекомендуемая форма библиографической ссылки: Ограничения применимости дифференцируемых эталонных показателей качества изображений / М.В. Синюков [и др.] // Препринты ИПМ им. М.В.Келдыша. 2022. № 86. 15 с. <https://doi.org/10.20948/prepr-2022-86>
<https://library.keldysh.ru/preprint.asp?id=2022-86>

**Ордена Ленина
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ
имени М.В.Келдыша
Российской академии наук**

**М.В. Синюков, Д.Л. Куликов,
Д.С. Ватолин, В.А. Галактионов**

**Ограничения применимости
дифференцируемых эталонных
показателей качества изображений**

Москва — 2022

М.В. Синюков, Д.Л. Куликов, Д.С. Ватолин, В.А. Галактионов
Ограничения применимости дифференцируемых эталонных показателей качества изображений

Измерение субъективного качества изображения играет ключевую роль в разработке приложений для обработки изображений. Цель показателя визуального качества — приблизить результаты субъективной оценки. В связи с этим разрабатывается все больше и больше метрик, но их ограничения рассматриваются в небольшом количестве исследований. В этой статье устраняется этот недостаток: мы показываем, как предварительная обработка изображений перед сжатием может искусственно повысить показатели качества, предоставляемые популярными метриками DISTs, LPIPS, HaarPSI и VIF, а также что эти показатели несовместимы с оценками субъективного качества. Мы предлагаем серию нейросетевых моделей предварительной обработки изображений, которые увеличивают DISTs на 34,5%, LPIPS на 36,8%, VIF на 98,0% и HaarPSI на 22,6% в случае изображений, сжатых в формате JPEG. Субъективное сравнение предобработанных изображений показало, что для большинства рассмотренных нами метрик качество изображений падает или остается неизменным, ограничивая применимость этих показателей качества.

Ключевые слова: взлом метрик, видеопроцессинг, сжатие изображений

Siniukov Maksim, Dmitriy Kulikov, Dmitriy Vatolin, Vladimir Galaktionov

Applicability limitations of differentiable full-reference image quality metrics

Subjective image-quality measurement plays a critical role in the development of image-processing applications. The purpose of a visual-quality metric is to approximate the results of subjective assessment. In this regard, more and more metrics are under development, but little research has considered their limitations. This paper addresses that deficiency: we show how image preprocessing before compression can artificially increase the quality scores provided by the popular metrics DISTs, LPIPS, HaarPSI, and VIF as well as how these scores are inconsistent with subjective-quality scores. We propose a series of neural-network preprocessing models that increase DISTs by up to 34.5%, LPIPS by up to 36.8%, VIF by up to 98.0%, and HaarPSI by up to 22.6% in the case of JPEG-compressed images. A subjective comparison of preprocessed images showed that for most of the metrics we examined, visual quality drops or stays unchanged, limiting the applicability of these metrics.

Key words: metrics hacking, video processing, image compression

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

1 Введение

Измерение качества изображения играет ключевую роль в разработке алгоритмов обработки изображений, таких как сжатие, повышение разрешения и адаптивная потоковая передача данных по сети. Несмотря на существование объективных показателей качества, визуальное качество часто является основным средством тестирования новых алгоритмов. Воспринимаемое качество измеряется с помощью субъективной оценки, дающей наиболее точные результаты, однако поскольку проведение субъективной оценки сложно и требует много времени, она не подходит для частого использования; объективные показатели качества более практичны. Однако для обеспечения высокой точности метрики должны хорошо коррелировать с результатами восприятия.

Эталонные метрики — метрики, которые требуют исходное изображение без искажений для оценки качества тестируемого изображения, — делятся на две категории: основанные на машинном обучении и эвристические. Примеры эвристических метрик: NLPD [10], HaarPSI [17] и VIF [19]. Разработчики *normalized Laplacian-pyramid distance measure (NLPD)* [10] основывались на локальном вычитании яркости и учёте усиления сигнала, а также использовали пирамиду Лапласа для имитации человеческого зрения, что обеспечило более высокую корреляцию с MOS, чем хорошо известные RMSE и MS-SSIM при оценке качества сжатого изображения. The Haar wavelet-based perceptual-similarity index (HaarPSI) [17] использует коэффициенты из вейвлет разложения Хаара для оценки локального сходства между двумя изображениями для оценки важности областей изображения. Visual-information fidelity (VIF) [19] оценивает то, сколько информации об исходном изображении можно извлечь из искаженного изображения, тем самым оценивая субъективное качество тестового изображения.

Популярные новые показатели, основанные на машинном обучении, включают PieAPP [15], LPIPS [27] и DISTs [8]. Perceptual image-error assessment through pairwise preference (PieAPP) [15] — это нейросетевой метод, основанный на наблюдении, что нейронная сеть оценки воспринимаемого качества может быть обучена на данных попарного сравнения без явных субъективных оценок изображений. Learned perceptual image-patch similarity (LPIPS) — это метрика на основе сверточной нейронной сети, которая использует признаки, извлеченные предварительно обученной сетью VGG. [27] Deep image structure and texture similarity (DISTs) [8] — это новая метрика субъективного качества, основанная на CNN, которая фокусируется на сходстве текстуры и структуры, поскольку эти характеристики сильно влияют на человеческое восприятие. Независимая исследовательская группа [1] протестировала вышеупомянутые метрики и пришла к выводу, что DISTs демонстрирует самую высокую корреляцию с субъективными оценками.

Корреляция с субъективной оценкой – основной показатель качества метрики. Сравнение методов оценки качества изображения (IQA) [1] показало, что при тестировании на сжатых изображениях без предварительной обработки DISTIS достиг корреляции SROCC 0,847, LPIPS - 0,749. Другая исследовательская группа [14] получила SROCC: 0,84 для PieAPP, 0,87 для HaarPSI, 0,81 для DISTIS и 0,67 для LPIPS. В дополнение к простому субъективному сравнению актуальны более глубокие исследования стабильности метрик на устойчивость к различным искусственным искажениям; некоторые искажения увеличивают значения метрики, уменьшая или оставляя неизменным визуальное качество. В этой статье мы описываем такие искажения и анализируем их влияние как на субъективное качество, так и на показатели метрик. Сегодня разрабатывается все большее и больше показателей воспринимаемого качества, но в большинстве случаев исследователи всё ещё используют старые SSIM и PSNR. Причина такой тенденции — недостаток знаний об ограничениях современных метрик, о которых мы пишем в данной статье.

2 Обзор области

Взлом VMAF, NIQE и PSNR тщательно изучен. Несмотря на то, что настройка параметров алгоритмов под объективные метрики для достижения лучших результатов в сравнении происходит во всей отрасли, это может снизить субъективное качество, точно так же, как настройка под субъективные оценки может понизить показатели объективных метрик. Авторы [5] проанализировали влияние предварительной обработки изображений на субъективное качество изображений и сравнили результаты с другими объективными показателями качества. В их экспериментах субъективное качество повышалось, в то время как PSNR понижался. Их результаты показали, что безэталонная метрика NIQE лучше коррелирует с субъективными оценками для предварительно обработанных изображений, чем другие метрики. Но NIQE также может быть неустойчивой. В другом исследовании [29] авторы продемонстрировали, что NIQE плохо справляется с оценкой качества темных и сильно текстурированных кадров.

Многие алгоритмы обработки изображений включают в себя фильтры настройки под метрики, которые особенно популярны в алгоритмах сжатия изображений. Например, x264 и x265 предоставляют параметры настройки для повышения PSNR и SSIM; разработчики libaom реализовали настройку под VMAF [7]. В их статье показано, что предварительная обработка даёт существенное усиление сжатия (битрейт снизился на 37,91 % для того же VMAF), в то время как rate-distortion оптимизация дала гораздо меньшее увеличение силы сжатия (4,69 %). Авторы [2] заметили, что выравнивание гистограммы и нерезкое маскирование могут искусственно увеличить VMAF на 5-6%; другие

Таблица 1: Зависимость функции потерь на тестовом датасете от размера обучающего датасета.

размер датасета	100	250	500	1000	200	5000	11000	20000	30000
DISTS	0.5496	0.503	0.456	0.441	0.429	0.427	0.423	0.4225	0.4222

исследователи также доказали уязвимость VMAF к предварительной обработке [13]. Было показано [22], что VMAF Neg устойчив к взлому при применении этого подхода, однако VMAF и VMAF Neg оказались уязвимыми при использовании другого метода взлома, в котором изображения сжимались с помощью кодека H.264. Авторы [4] использовали генеративно-состязательный подход для повышения оценок VMAF обучаемого алгоритма сжатия изображений.

Во многих исследованиях, включая [21] и [28], проводились способы поиска уязвимостей безэталонных метрик. В [9] авторы показали, что итеративный алгоритм может понизить LPIPS особым образом, что приводит к отрицательной корреляции LPIPS с MOS. Также появилось несколько методов поиска состязательных примеров для PSNR, включая [18]. В другом исследовании были получены изображения с одинаковыми значениями SSIM (метрика использовалась как показатель субъективного качества изображения), но разными значениями PSNR [6]. Поскольку метрики, которые мы затрагиваем в данной статье, более новые, они не так хорошо изучены. [25]

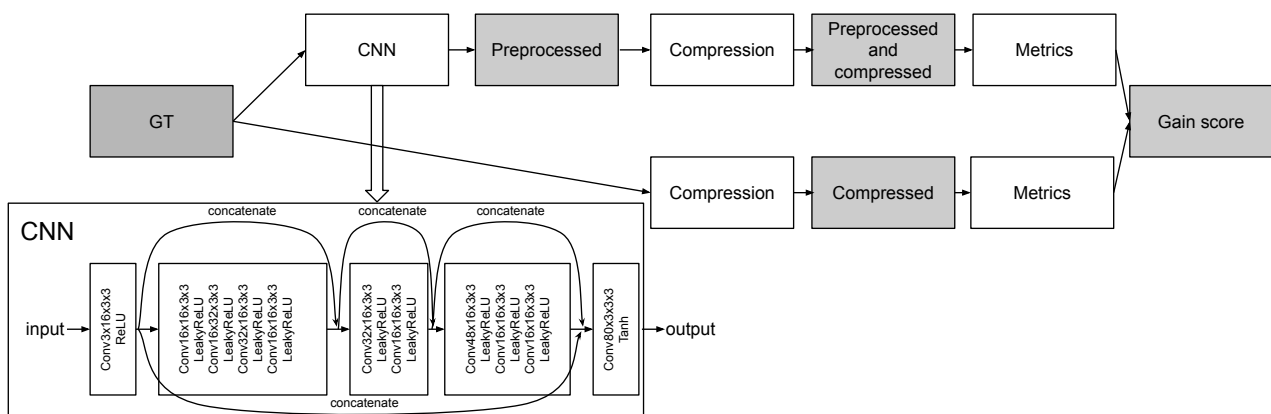


Рисунок 1: Схема взлома.

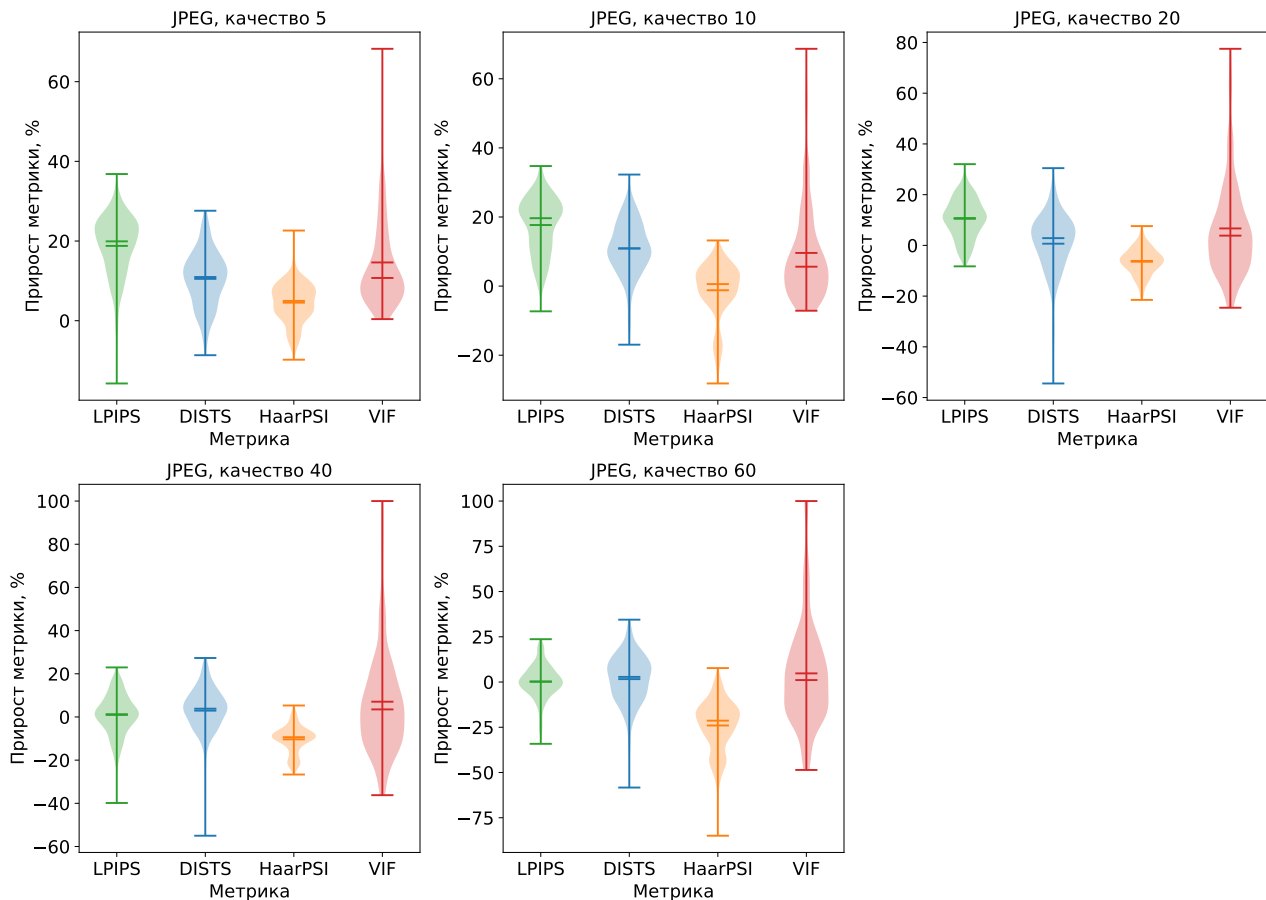


Рисунок 2: Распределение прироста метрик.

3 Метод

Целью исследования было найти уязвимости, которые делают метрику неприменимой к сжатым изображениям, поскольку на практике, как правило, изображения сжимают (с использованием JPEG). Для каждой метрики мы нашли универсальное преобразование таким образом, что при любом входном исходном изображении (GT) мы можем сгенерировать пару изображений: одно сжатое в формате JPEG (C), а другое преобразованное с использованием предварительно обученной нейронной сети, а затем сжатое при помощи JPEG (ТС). Если метрика оценивает ТС выше, чем C, а субъективные оценки и другие метрики идут вразрез, то целевой показатель непригоден для субъективной оценки качества при использовании со сжатием JPEG. Так, разработчики кодеков могут включить подобный этап предварительной обработки в свои алгоритмы, чтобы увеличить оценки уязвимых метрик и занять более высокое место в бенчмарках. Поскольку JPEG является наиболее распространенным алгоритмом сжатия изображений, он имеет первостепенное значение для сравнений алгоритмов в области сжатия изображений.

Мы разработали алгоритм предварительной обработки изображений, осно-

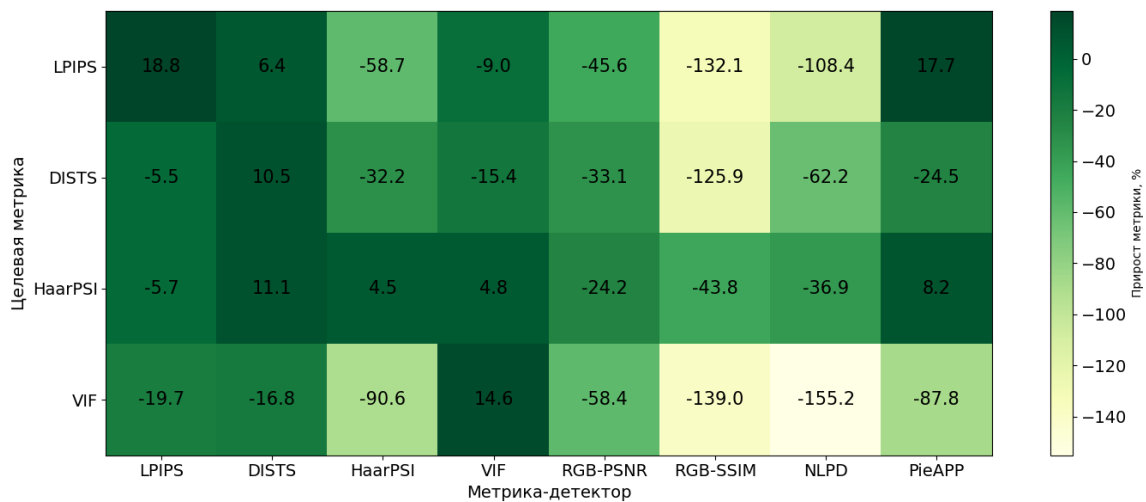


Рисунок 3: Тепловая карта обнаружения взлома.

ванный на легковесной свёрточной нейронной сети, подобной ResNet, она показана на рисунке 1. Общее количество весов составляет 32 707, а размер модели - 133 КБ (136 681 байт), что обеспечивает высокую производительность (до 10,95 кадров в секунду на графическом процессоре Nvidia RTX 2080Ti; 0,17 кадров в секунду на процессоре Intel Core i7-3770; 0,07 кадров в секунду на мобильном процессоре Snapdragon 855 для изображений разрешения FullHD) и делает этот подход применимым для использования в реальном времени. Таким образом, другие методы сжатия могут включать такую предварительную обработку без потери в скорости.

Вторым этапом нашего метода является сжатие. Поскольку обратное пространство ошибки необходимо для обучения препроцессора, мы выбрали дифференцируемые алгоритмы сжатия. Кроме того, поскольку обычные реализации JPEG недифференцируемы, мы использовали дифференцируемый DiffJPEG [20], который имитирует поведение обычного JPEG. Мы выбрали реализацию из [11] для нашей схемы обучения. Последний шаг — расчет эталонной метрики. Мы измеряли каждую метрику на оцениваемом изображении ТС и на изображении GT в качестве эталона. Функция потерь – метрика, как мера различия между двумя вышеупомянутыми изображениями (более низкое значение метрики соответствует более высокому субъективному качеству). Изображения, сгенерированные таким образом, показали аномально высокие значения метрик, которые превышают значения на исходных сжатых изображениях. Этот результат поставил под вопрос корреляцию метрики с субъективными оценками и другими объективными оценками качества.

Все модели предварительной обработки были обучены на 11 000 изображениях из набора данных Vimeo90K [26] и были протестированы на 1500 изображениях из различных последовательностей Vimeo90K; этот набор данных широко используется для обработки изображений и использовался в многочислен-

ных исследованиях [23], [3]. Мы выбрали размер обучающего набора данных в 11 000 изображений, изучив, как размер тренировочного набора данных влияет на прирост метрики на тестовом наборе. Таблица 1 представляет результаты в виде графика, и показывает, что метрика выходит на плато при размере набора данных в 11 000 изображений. Мы обучили нейронные сети с методом ранней остановки: максимальное количество эпох составляло 50, но обучение прекращалось, если функция потерь на тестовом наборе не улучшалась в течение пяти последовательных эпох.

4 Практическое исследование

Мы обучили нейронные сети предварительной обработки при использовании DiffJPEG, но при сжатии обычно применяют недифференцируемый JPEG. Поэтому мы протестировали переносимость результатов эксперимента на различные недифференцируемые реализации JPEG — в частности, OpenCV[12] и PIL[16]. Схема тестирования моделей была следующей: первый этап, обучение моделей предварительной обработки с использованием DiffJPEG с различными значениями параметра качества. Затем вычислялось значение взламываемой метрики, используя обученную модель предварительной обработки для изображения, сжатого в стандартном формате JPEG, и для изображения GT без предварительной обработки в качестве эталона. Таким образом, мы убедились, что показатели метрик имеют аномально высокие значения при сжатии изображений в стандартном формате JPEG, это говорит об ограничении применимости метрик. В таблице 2 представлены результаты; прирост метрик ниже, чем при тестировании с DiffJPEG. Таким образом, модели, обученные с использованием DiffJPEG, могут использоваться на практике при сжатии стандартным JPEG, а также могут быть встроены в специально разработанные кодеры в качестве этапа предварительной обработки, это позволит кодерам получать более высокие показатели в сравнениях кодеров.

Более того, описанный подход работает и в случае сжатия видео, поскольку в видеокодеках ключевые кадры сжимаются алгоритмами сжатия изображений; включение этапа предварительной обработки в эти алгоритмы может привести к неоправданно высоким результатам в сравнениях кодеков. Так, кодек AV1 уже имеет подобный этап предварительной обработки, который включается опцией “-tune vmaf”[7], и повышает оценки уязвимой метрики VMAF. Однако настройка под VMAF в AV1 основана на итеративном алгоритме, который выполняет поиск параметров преобразования полным перебором, значительно замедляя AV1, в то время как наш подход однопроходный и почти не влияет на скорость.

Таблица 2: Переносимость на обычный JPEG при параметре качества 10

	DISTS	LPIPS	HaarPSI	VIF
прирост метрики в среднем	6.31%	18.94%	2.03%	20.40%

Таблица 3: Результаты субъективного сравнения

	прирост метрики	субъективная Брэдли-Терри(ТС/С)	доля обработанных изображений выбранных как лучшие	
			метрикой	людьми
DISTS	5.21%	1.176/0.583	73.8%	22.4%
LPIPS	12.67%	0.714/0.357	90.0%	32.2%
HaarPSI	-1.03%	1.367/0.683	57.5%	17.9%
VIF	12.54%	3.857/1.929	76.3%	1.3%

5 Субъективная оценка

IQA метрики разработаны для имитации субъективного восприятия человека, поэтому если их показатели отрицательно коррелируют с субъективными оценками, то они имеют серьезный недостаток. Чтобы убедиться, что метрики оценивают сгенерированные нами пары изображений вразрез с субъективными оценками человеческого восприятия, мы выбрали несколько пар изображений и предложили участникам сравнения выбрать, какое изображение в каждой паре лучше. В субъективной оценке приняли участие 531 человек. Чтобы убедиться, что участники поняли задание и правильно выполнили свою работу, мы включили еще 2 изображения, для которых был очевиден правильный ответ.

Мы использовали модель Брэдли-Терри для преобразования результатов попарного сравнения в единую меру MOS; Таблица 3 показывает полученные оценки Брэдли-Терри. Они значительно ниже, чем когда метрики тестировались на изображениях, сжатых без какой-либо предварительной обработки. Участники предпочли изображения, которые не были предварительно обработаны, что привело к низкой корреляции между MOS и значениями рассмотренных нами метрик. Следовательно, эти метрики не подходят для субъективной оценки сжатых изображений в IQA из-за их несоответствия оценке человеком.

VIF меньше всего соответствовала MOS, в то время как DISTS больше всего соответствовала MOS. Взломанные метрики перестали соответствовать человеческому восприятию, потому что их значения повысились, в то время как субъективные оценки понизились. Примеры сжатых версий предварительно обработанных изображений приведены на рисунке 4.

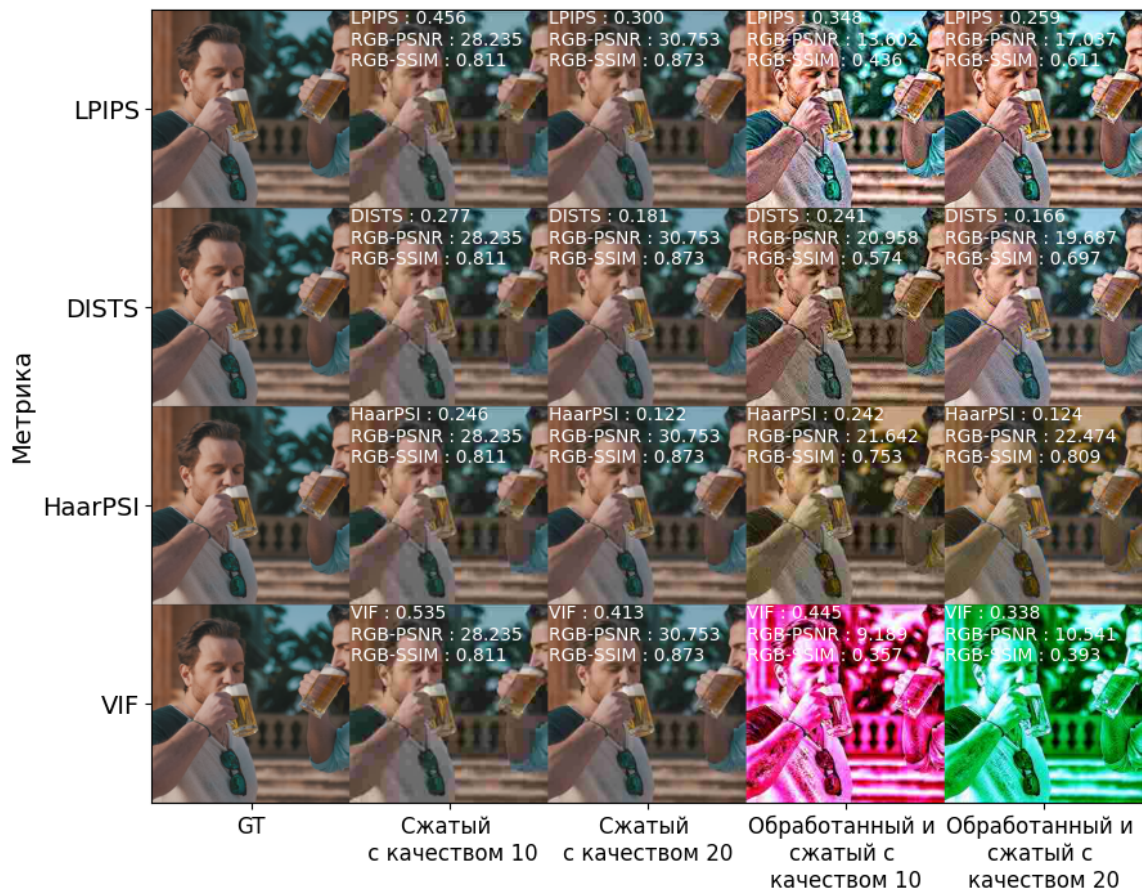


Рисунок 4: Примеры взлома метрик.

6 Обобщение взлома

Мы обнаружили, что прирост значений метрик наблюдается на 93,6% из наших 1500 тестовых изображений, сжатых с использованием DiffJPEG на уровне качества 5 для случая DISTS, 97,8% для LPIPS, 83% для HaarPSI и 100% для VIF. Однако на некоторых изображениях метрики повысились слабо, а на некоторых, наоборот, гораздо сильнее, чем в среднем. DISTS и LPIPS не увеличивались для высококонтрастных изображений и увеличивались больше всего, когда изображения были неконтрастными. HaarPSI не повышался, когда изображения были чрезвычайно яркими. Таким образом, наш подход является универсальным, поскольку он работал почти на всех тестируемых изображениях. Рисунок 2 представляет распределение прироста метрик на тестовом датасете в виде скрипичной диаграммы, показывая, что HaarPSI имеет низкое стандартное отклонение усиления, что означает, что его взлом стабилен, тогда как взлом VIF нестабилен.

7 Обнаружение взлома

Как упоминалось выше, разработчики могут включать в свои алгоритмы методы взлома, поэтому задача обнаружения таких алгоритмов становится актуальной. На практике обнаружение взлома может быть реализовано следующим образом: если результаты субъективной оценки не коррелируют с результатами метрики, как мы описали в предыдущем разделе, метрика была взломана. Но сбор субъективных оценок отнимает много времени и требует большого числа участников. Следовательно, это непрактично для обнаружения взлома. Однако значения других метрик могут показать, была ли взломана целевая метрика. Мы обнаружили, что взлом одной метрики влияет на показатели других. Этот результат показан в виде тепловой карты на рисунке 3. Как правило, другие метрики указывают на предобработанные изображения низкого качества, предоставляя низкие оценки, но, например, DIST5 невосприимчив к взлому LPIPS, возможно, потому, что они имеют схожую внутреннюю структуру. PSNR и SSIM — лучшие детекторы взлома.

8 Будущая работа

Другим типом уязвимостей метрик, представляющим интерес, является незаметный взлом, который лишь незначительно изменяет изображение, не оставляет артефактов и не влияет на значения других метрик. В таких случаях взлом может остаться незамеченным. Первым шагом к поиску таких уязвимостей может стать разработка модели предварительной обработки, которая увеличивает все IQA метрики разом, не оставляя следов взлома. В этом случае субъективные оценки остаются неизменными, в то время как уязвимые метрики дают более высокие оценки.

9 Заключение

Эталонные метрики качества изображений DIST5, LPIPS, HaarPSI и VIF не являются универсальными, поскольку некоторые универсальные преобразования могут увеличить их показатели и уменьшить их корреляцию с MOS. Таким образом, они могут показывать неточную оценку качества изображения и не должны служить основой для принятия решений при оценке качества. Мы показали, что взлом может увеличить эти метрики до 98,0%; самый низкий прирост был в случае HaarPSI. Значения метрик могут вводить в заблуждение при оценке алгоритмов сжатия, поэтому субъективная оценка все еще необходима. Традиционные PSNR и SSIM, однако, невосприимчивы к такому увеличению, что делает их хорошим индикатором для обнаружения взлома, но их корреляция

с MOS слишком мала, чтобы использовать их в качестве субъективных показателей качества, поэтому разработка новых универсальных методов визуальной оценки качества остается актуальной задачей [24]. Разработчики VMAF, например, предложили защищенную от взлома версию своей метрики VMAF Neg. Аналогично, в будущем могут появиться более устойчивые версии метрик, которые мы исследовали в статье.

Список литературы

- [1] Anastasia Antsiferova et al. “Video compression dataset and benchmark of learning-based video-quality metrics”. In: *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*. 2022. URL: <https://openreview.net/forum?id=My5AI9aM49R>.
- [2] Anastasia Antsiferova et al. “Video Distortion Method for VMAF Quality Values Increasing”. In: *CoRR* abs/1907.04807 (2019). arXiv: 1907.04807. URL: <http://arxiv.org/abs/1907.04807>.
- [3] Wenbo Bao et al. “Depth-Aware Video Frame Interpolation”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 2019.
- [4] Li-Heng Chen et al. “ProxIQA: A Proxy Approach to Perceptual Optimization of Learned Image Compression”. In: *IEEE Transactions on Image Processing* 30 (2021), pp. 360–373. DOI: 10.1109/TIP.2020.3036752.
- [5] Jayachandra Chilukamari, Sampath Kannangara, and Grant Maxwell. “Investigation of the effectiveness of video quality metrics in video pre-processing”. In: *2013 IEEE Third International Conference on Consumer Electronics & Berlin (ICCE-Berlin)*. 2013, pp. 1–5. DOI: 10.1109/ICCE-Berlin.2013.6698005.
- [6] Frank M. Ciaramello and Amy R. Reibman. “Supplemental subjective testing to evaluate the performance of image and video quality estimators”. In: *Human Vision and Electronic Imaging XVI*. Ed. by Bernice E. Rogowitz and Thrasyvoulos N. Pappas. Vol. 7865. International Society for Optics and Photonics. SPIE, 2011, 78650Q. DOI: 10.1117/12.876747. URL: <https://doi.org/10.1117/12.876747>.
- [7] Sai Deng, Jingning Han, and Yaowu Xu. “VMAF Based Rate-Distortion Optimization for Video Coding”. In: *2020 IEEE 22nd International Workshop on Multimedia Signal Processing (MMSP)*. 2020, pp. 1–6. DOI: 10.1109/MMSP48831.2020.9287114.

- [8] Keyan Ding et al. “Image Quality Assessment: Unifying Structure and Texture Similarity”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 44.5 (2022), pp. 2567–2581. DOI: 10.1109/TPAMI.2020.3045810.
- [9] Markus Kettunen, Erik Härkönen, and Jaakko Lehtinen. “E-LPIPS: robust perceptual image similarity via random transformation ensembles”. In: *arXiv preprint arXiv:1906.03973* (2019).
- [10] Valero Laparra et al. “Perceptual image quality assessment using a normalized Laplacian pyramid”. In: *Electronic Imaging 2016* (Feb. 2016), pp. 1–6. DOI: 10.2352/ISSN.2470-1173.2016.16.HVEI-103.
- [11] Mlomnitz. *DiffJPEG*. (Visited on 11/02/2022).
- [12] Opencv. *Opencv/opencv: Open source computer vision library*. <https://github.com/opencv/opencv>.
- [13] Jan Ozer. *VMAF is hackable: What now?* Apr. 2020. URL: <https://streaminglearningcenter.com/blogs/vmaf-is-hackable-what-now.html>.
- [14] PIQ. *PyTorch Image Quality*). <https://github.com/photosynthesis-team/piq>.
- [15] Ekta Prashnani et al. “PieAPP: Perceptual Image-Error Assessment Through Pairwise Preference”. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2018, pp. 1808–1817. DOI: 10.1109/CVPR.2018.00194.
- [16] Python-Pillow. *Python-pillow/pillow: The friendly pil fork (Python Imaging Library)*. <https://github.com/python-pillow/Pillow>.
- [17] Rafael Reisenhofer et al. “A Haar wavelet-based perceptual similarity index for image quality assessment”. In: *Signal Processing: Image Communication* 61 (2018), pp. 33–43. ISSN: 0923-5965. DOI: <https://doi.org/10.1016/j.image.2017.11.001>. URL: <https://www.sciencedirect.com/science/article/pii/S0923596517302187>.
- [18] Mahmood Sharif, Lujo Bauer, and Michael Reiter. “On the Suitability of Lp-Norms for Creating and Preventing Adversarial Examples”. In: June 2018, pp. 1686–16868. DOI: 10.1109/CVPRW.2018.00211.
- [19] H.R. Sheikh and A.C. Bovik. “Image information and visual quality”. In: *IEEE Transactions on Image Processing* 15.2 (2006), pp. 430–444. DOI: 10.1109/TIP.2005.859378.
- [20] Richard Shin and Dawn Song. “Jpeg-resistant adversarial images”. In: *NIPS 2017 Workshop on Machine Learning and Computer Security*. Vol. 1. 2017, p. 8.

- [21] Ekaterina Shumitskaya, Anastasia Antsiferova, and Dmitriy Vatolin. “Universal Perturbation Attack on Differentiable No-Reference Image- and Video-Quality Metrics”. In: *The British Machine Vision Conference* (2022).
- [22] Maksim Siniukov et al. “Hacking VMAF and VMAF NEG: Vulnerability to Different Preprocessing Methods”. In: *2021 4th Artificial Intelligence and Cloud Computing Conference. AICCC '21*. Kyoto, Japan: Association for Computing Machinery, 2022, pp. 89–96. ISBN: 9781450384162. DOI: 10.1145/3508259.3508272. URL: <https://doi.org/10.1145/3508259.3508272>.
- [23] Xintao Wang et al. “EDVR: Video Restoration With Enhanced Deformable Convolutional Networks”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*. June 2019.
- [24] Zhou Wang and Eero P Simoncelli. “Maximum differentiation (MAD) competition: A methodology for comparing computational models of perceptual quantities”. In: *Journal of Vision* 8.12 (2008), pp. 8–8.
- [25] Zhou Wang et al. “Image quality assessment: from error visibility to structural similarity”. In: *IEEE Transactions on Image Processing* 13.4 (2004), pp. 600–612. DOI: 10.1109/TIP.2003.819861.
- [26] Tianfan Xue et al. “Video Enhancement with Task-Oriented Flow”. In: *International Journal of Computer Vision* 127.8 (Aug. 2019), pp. 1106–1125. ISSN: 1573-1405. DOI: 10.1007/s11263-018-01144-2. URL: <https://doi.org/10.1007/s11263-018-01144-2>.
- [27] Richard Zhang et al. “The Unreasonable Effectiveness of Deep Features as a Perceptual Metric”. In: *CoRR* abs/1801.03924 (2018). arXiv: 1801.03924. URL: <http://arxiv.org/abs/1801.03924>.
- [28] Weixia Zhang et al. *Perceptual Attacks of No-Reference Image Quality Models with Human-in-the-Loop*. 2022. DOI: 10.48550/ARXIV.2210.00933. URL: <https://arxiv.org/abs/2210.00933>.
- [29] A. Zvezdakova et al. “Barriers towards no-reference metrics application to compressed video quality analysis: On the example of no-reference metric NIQE”. In: *CEUR Workshop Proceedings* 2485 (2019), pp. 22–27.

Оглавление

Введение	3
Обзор области	4
Метод	6
Практическое исследование	8
Субъективная оценка	9
Обобщение взлома	10
Обнаружение взлома	11
Будущая работа	11
Заключение	11
Список литературы	12