Записки научных
семинаров ПОМИ
Том 552, 2026 г.

И. А. Молодецких, К. В. Малышев, М. Р. Миргалеев, Е. Н. Богатырёв, Д. С. Ватолин

МЕТОДОЛОГИЯ СБОРА И АНАЛИЗА СУБЪЕКТИВНОЙ ЗАМЕТНОСТИ АРТЕФАКТОВ НЕЙРОСЕТЕВЫХ МОДЕЛЕЙ СУПЕРРАЗРЕШЕНИЯ

§1. Введение

Задача суперразрешения одиночного изображения (SISR) состоит в восстановлении изображения высокого разрешения (HR) по его версии низкого разрешения (LR). Недавние достижения в области глубокого обучения и генеративных состязательных сетей (GAN) значительно повысили субъективное качество изображений, но создали новую серьёзную проблему — появление визуально неприятных артефактов. Эти артефакты — такие как неестественные структуры, смазанные лица и искажения текстур — снижают визуальное качество результатов SISR и ограничивают их практическую применимость. Даже самые современные методы, основанные на трансформерах и диффузионных моделях [12, 24], по-прежнему склонны к генерации артефактов.

Для решения этой проблемы в последнее время начинают появляться методы детектирования артефактов, такие как LDL [11], DeSRA [21] и PAL4VST [27]. Как правило, эти методы полагаются на вручную размеченные наборы данных с двоичными масками артефактов. В данной работе рассматривается гипотеза, что артефакты различаются по своей заметности для зрителя. Например, искажения на регулярных структурах вроде зданий, или узнаваемых объектах вроде человеческих лиц, мгновенно привлекают внимание и могут вызывать дискомфорт (рисунок 1). С другой стороны, артефакты на воде, траве и других органических поверхностях зачастую почти незаметны (рисунок 2). Если рассматривать эти случаи одинаково, есть риск, что метод обнаружения переобучится на несущественные артефакты, пропуская

Ключевые слова: суперразрешение, артефакты, компьютерное зрение, глубокое обучение.

Работа выполнялась с использованием суперкомпьютера “МГУ-270” МГУ имени М. В. Ломоносова.

Рис. 1. Примеры заметных артефактов. Слева — из предлагаемого набора данных: модель RealSR размазала отверстия на панели радиоприёмника. Справа — из набора данных DeSRA: метод LDL сгенерировал текстурный артефакт (линии) на сумке. Заметность указывает долю асессоров, подтвердивших наличие артефакта в выделенной области.

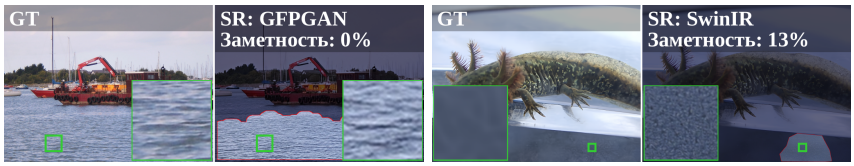

Рис. 2. Пример незаметных артефактов. Слева — из предлагаемого набора данных: поверхность воды была неверно восстановлена GFPGAN, создав артефакт, но поскольку это естественная поверхность, он не является заметным для человека. Справа — из набора данных DeSRA: метод SwinIR сгенерировал текстурный артефакт (точки) на стенке, но эта область несалиентна, поэтому данный артефакт также не заметен для наблюдателя.

действительно раздражающие — что в итоге ухудшит визуальное восприятие. Таким образом, двоичная разметка существующих наборов данных является их серьёзным недостатком.

Чтобы устранить этот недостаток, в рамках данной работы создан обширный набор данных из 1302 примеров артефактов суперразрешения, полученных 11 современными методами SISR из 500 исходных

изображений, где каждый артефакт снабжён оценкой заметности, размеченной с помощью краудсорсинга. Также собраны оценки заметности для всех 593 артефактов из набора данных DeSRA (изначально размеченного в лабораторных условиях Xie и др.). Эти оценки показали, что почти половина артефактов в этом наборе не является заметной для большинства зрителей.

Основной вклад работы заключается в следующем:

(1) Предложена методология сбора данных и краудсорсинговой разметки заметности артефактов, включающая постобработку масок для упрощения восприятия и процедуры контроля качества. Проведён анализ вариативности ответов, обосновывающий выбор количества асессоров.
(2) Представлен первый набор артефактов SISR с аннотациями заметности. Набор включает 1302 примера артефактов и охватывает широкий спектр искажений, сгенерированных 11 методами SISR. Кроме того, собраны оценки заметности для 593 артефактов из набора данных DeSRA [21].
(3) Проанализированы 11 методов суперразрешения на предмет их склонности к генерации артефактов. Показано, что даже новейшие модели высокого качества, такие как SUPIR [24], остаются сильно подверженными этой проблеме.

Наборы данных с аннотациями заметности опубликованы по ссылке `https://drive.google.com/drive/folders/1Rue6vPgXnxZTFw0aJvWWcM2mZlRlGtg8?usp=sharing`.

§2. Обзор литературы

Задача **суперразрешения одиночного изображения (SISR)** активно исследуется в течение последнего десятилетия. Ранние методы оптимизировали поканальные функции потерь, такие как L_1 и L_2, что улучшало показатели PSNR/SSIM, но плохо восстанавливало реалистичные детали. Подходы на основе GAN и реалистичные деградационные модели для обучения повысили качество генерируемых деталей [10, 17, 26].

Недавние достижения в области SISR включают архитектуры на основе трансформеров и диффузионных моделей. Подходы с трансформерами, такие как HAT [2] и DRCT [5], используют многоуровневый механизм внимания, что повышает качество при сохранении эффективности. Диффузионные методы, например ResShift [25] и SinSR [19],

ускоряют сэмплирование с помощью сдвига остатков и детерминированной дистилляции, в то время как StableSR [15] использует априоры Stable Diffusion [14] для увеличения изображений в реальных условиях, а SUPIR [24] масштабирует модели и данные для фотореалистичного восстановления, управляемого текстом. Однако эти достижения сопровождаются новыми проблемами, в частности — появлением визуальных артефактов.

Детектирование артефактов SR. Задача выявления и подавления артефактов суперразрешения начинает развиваться в последние годы. Существуют подходы, интегрированные в обучение SR, такие как LDL [11]. Здесь в процессе обучения SR-модели формируются карты искажённых областей, которые затем используются для регуляризации, что уменьшает интенсивность артефактов без заметной потери деталей. Другие методы, в частности DeSRA [21, 22], направлены на локализацию и ослабление артефактов уже обученных моделей SR в условиях отсутствия эталонных изображений высокого разрешения. В этой работе также был предложен набор данных с двоичными масками артефактов, размеченными в лабораторных условиях, что сделало возможным обучение и объективную оценку детекторов артефактов.

Схожие по постановке задачи встречаются и в других направлениях генеративного восстановления, например в PAL4Inpainting [28] и PAL4VST [27], где модели обучаются на задачу сегментации артефактов по наборам данных с двоичными масками.

Несмотря на значительный прогресс в области детектирования артефактов, нехватка специализированных наборов данных артефактов остаётся существенным ограничением. Существующие наборы, такие как DeSRA, содержат только двоичные маски артефактов. Такие маски отражают лишь факт наличия искажения, но не передают степень его выраженности или субъективную заметность, что снижает способность обученных на таких наборах детекторов к обобщению и ограничивает их применимость в реальных условиях.

Предлагаемая в данной работе методология направлена на устранение этого ограничения путём создания нового набора данных с аннотированными областями артефактов и непрерывными оценками их заметности.

§3. Методология сбора данных

Существующие наборы данных, такие как DeSRA [21], содержат только двоичные маски артефактов и не включают информацию о том, насколько заметны эти артефакты для наблюдателей. Чтобы сделать возможным исследование *заметности* артефактов, нужен набор данных, в котором каждый артефакт аннотирован как двоичной маской, так и оценкой заметности.

Предлагаемый набор данных основан на Open Images [9] — коллекции из 300 000 разнообразных естественных изображений (лицензия CC BY 2.0). Из коллекции случайным образом отбираются исходные фотографии разрешением 768×1024 пикселей. Эти изображения уменьшаются в 4 раза бикубической интерполяцией, а затем увеличиваются обратно несколькими современными нейросетевыми методами SR, формируя множество изображений-кандидатов для поиска артефактов.

Далее, двоичные маски кандидатов артефактов получаются посредством ручной разметки, а также путём запуска существующих метрик визуального качества (SSIM, DISTS, LPIPS) и алгоритмов детекции артефактов (LDL, DeSRA). Для каждого алгоритма кандидаты артефактов ранжировались по силе искажений, затем ложноположительные срабатывания удалялись посредством ручного отсмотра. Оставшиеся кандидаты артефактов проходят краудсорсинговую аннотацию заметности.

3.1. Краудсорсинговая аннотация. Для сбора данных использовалась платформа Яндекс.Задания. Участникам показывались пары изображений с подписями “Исходное” и “Увеличенное”, при этом область с артефактом визуально выделялась. Участникам предлагалось ответить, содержит ли выделенная область искажённый объект или текстуру. Пример вопроса показан на рисунке 3.

Каждое изображение оценивалось 30 разными участниками. Заметность вычислялась как доля участников, проголосовавших за наличие артефакта.

Перед основным заданием участники проходили четыре обучающих вопроса с объяснением правильных ответов, а затем — четыре экзаменационных вопроса с заранее известным, но скрытым ответом. Далее, для контроля качества, каждая группа из 20 вопросов содержала 4 контрольных вопроса. Все ответы участников, ошибшихся в двух и

Рис. 3. Интерфейс, представленный участникам при сборе субъективных данных.

более контрольных вопросах, исключались. Всего в аннотации участвовало 758 участников, успешно завершили её 596 участников.

В разд. 3.4 анализируется влияние числа участников на вариативность ответов.

3.2. Выбор “исходного” изображения. В практическом применении суперразрешения эталонное изображение полного разрешения не доступно: имеется лишь вход низкого разрешения (LR) и результат SR. Следовательно, оценивание не может опираться на точность совпадения с неизвестным GT высокого разрешения, а должно фокусироваться на правдоподобии восстановленных деталей и отсутствии артефактов. Для одного и того же входного изображения низкого разрешения существует множество подходящих выходных изображений высокого разрешения.

Эмпирически это подтверждается в несоответствии рейтингов в специализированных бенчмарках: бенчмарке SR по точности восстановления исходного изображения [8] и бенчмарке SR, направленном на перцептивное качество результата [7]. Таблицы лидеров в этих бенчмарках существенно различаются.

В связи с этим в интерфейсе аннотации в качестве “исходного” изображения целесообразно показывать именно вход низкого разрешения,

а не эталонное высокое разрешение. В предлагаемой методологии входное изображение низкого разрешения перед показом участникам масштабируется до целевого размера методом ближайшего соседа и отображается как “Исходное”. Использование именно метода ближайшего соседа делает очевидным для асессоров, что им показано входное изображение низкого разрешения.

3.3. Постобработка масок. Метод детекции артефактов должен формировать максимально “плотные” маски вокруг артефактов, так как такие маски полезнее для последующего анализа и автоматической коррекции. Однако, плотные маски затрудняют субъективную визуальную оценку присутствия артефакта в выделенной области.

Чтобы устранить эту проблему, перед показом участникам к маскам применяются морфологические операции:

(1) Размыкание с ядром 25×25 пикселей для удаления мелких точек.
(2) Наращивание с круговым ядром 64×64 пикселя, чтобы включить контекст вокруг артефакта.
(3) Замыкание с ядром 25×25 пикселей, чтобы устранить дыры и отступить от границ изображения.

Пример на рисунке 4 демонстрирует, что плотная маска затрудняет восприятие артефакта по сравнению с результатом после постобработки.

Для проверки корректности этой процедуры, была проведена двойная краудсорсинговая аннотация заметности артефактов в наборе DeSRA — с постобработкой и без неё. Для сравнения использовались раздельные группы участников, при этом порядок вопросов был сохранён. В результате средняя заметность артефактов по всему набору DeSRA составила 49,4% с постобработкой и 47,7% с оригинальными масками, что показывает, что постобработка помогает участникам лучше оценивать наличие артефакта.

3.4. Разброс краудсорсинговых аннотаций. Предложенный процесс краудсорсинговой аннотации заметности использует 30 участников для оценки каждого изображения. Для выбора этого числа был проведён анализ вариативности ответов.

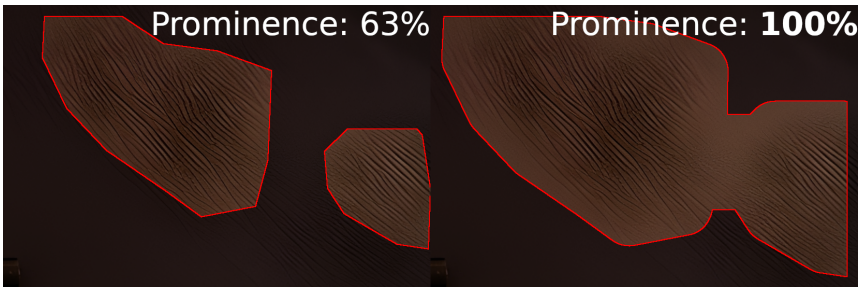

Рис. 4. Пример постобработки масок, повышающей видимость артефактов. Слева — оригинальная маска артефакта из набора DeSRA; справа — та же маска после постобработки.

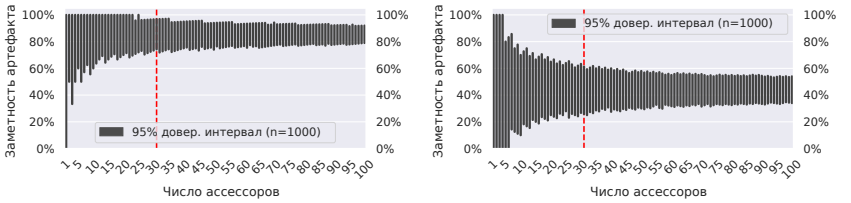

Рис. 5. Результаты бутстрэп-анализа для изображения с сильно заметным артефактом (слева) и слабо заметным артефактом (справа). Красная линия обозначает выбранное нами количество асессоров — 30.

Для этого анализа были отобраны 11 изображений, восстановленных методами SR и содержащих артефакты различной степени выраженности, после чего проведена краудсорсинговая аннотация заметности с увеличенным числом участников: каждое изображение было оценено 200 участниками вместо 30. Далее голоса анализировались бутстрэпом. Для каждого количества асессоров k от 1 до 100 случайным образом выбирались k голосов с возвращением, и по этим голосам вычислялась оценка заметности. Процедура повторялась n=1000 раз; затем для каждого k вычислялись 95%-е доверительные интервалы.

Рисунок 5 показывает доверительные интервалы для двух примеров изображений: с сильно- и слабо-заметным артефактом. При малом

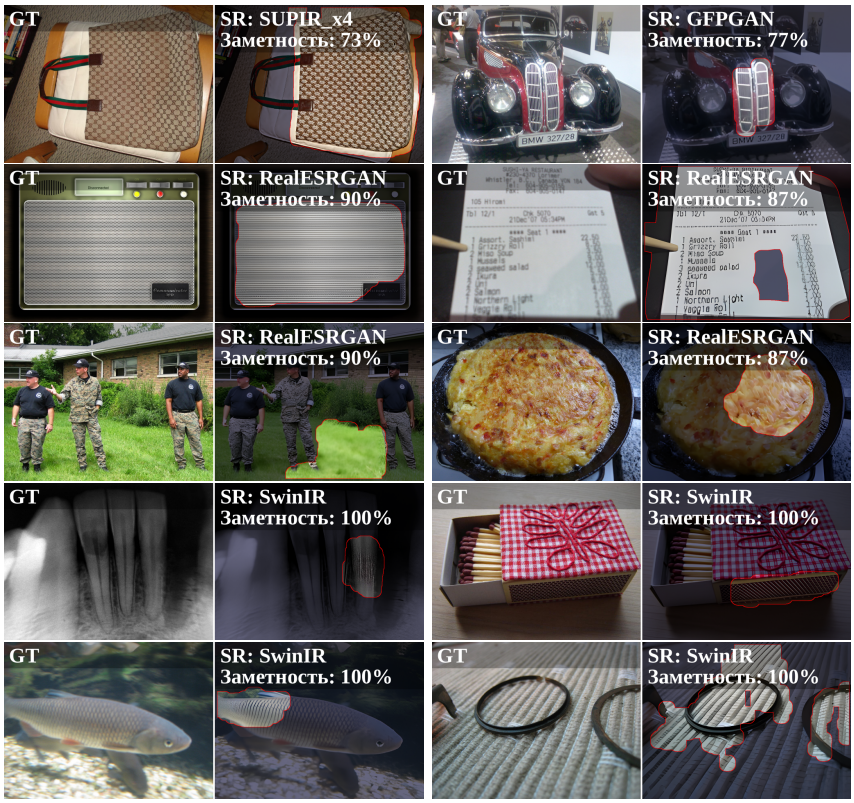

Рис. 6. Примеры артефактов из предложенного набора.

числе асессоров (1–5) доверительный интервал часто охватывает весь диапазон заметности от 0% до 100%, что означает, что пять случайных участников могут как все указать наличие артефакта, так и все его не заметить. Это особенно характерно для неочевидных случаев около 50% заметности. С увеличением числа асессоров доверительный интервал сужается и достигает примерно $\pm 10\%$ при 100 асессорах.

Для основной части процесса аннотации было выбрано 30 асессоров — компромисс между надёжностью результата ($\pm 20\%$) и временем/стоимостью привлечения большого числа участников.

	DeSRA	Предложенный
Методов SR	3	**11**
Артефактов	593	**1302**
Источник масок	лабораторный	автоматический
Тип разметки	двоичный	**заметность**

Таблица 4.1. Сравнение набора DeSRA и предложенного набора данных артефактов.

§4. Предложенный набор данных

Итоговый предложенный набор состоит из 1302 примеров артефактов с двоичными масками и оценками заметности на основе 500 исходных фотографий. На рисунке 6 представлены примеры.

Для поиска кандидатов артефактов было взято 2101 исходных фотографий из набора Open Images и обработано одиннадцатью популярными методами SR [2, 5, 6, 12, 15–17, 19, 20, 24, 25] по процедуре, описанной в разд. 3. Таким образом, поиск кандидатов артефактов выполнялся на 23 111 изображениях — результатах супреразрешения.

Дополнительно в рамках данной работы были собраны аннотации заметности для всех 593 изображений из набора артефактов DeSRA.

4.1. Сравнение с набором DeSRA. В таблице 4.1 и на рисунке 7 предложенный набор данных сравнивается с DeSRA и показывается распределение оценок заметности.

Предложенный набор данных смещён в сторону нулевой заметности (отсутствие или почти незаметные артефакты); причина в том, что исходные маски формировались по результатам существующих метрик качества изображений, плохо приспособленных для поиска артефактов, и методов детектирования, не различающих едва заметные и высокозаметные артефакты. Это смещение следует учитывать при проведении объективных оценок и обучении моделей на предложенном наборе, например, с помощью создания равномерно распределённых подвыборок, или перевзвешивания.

Набор DeSRA имеет более равномерное распределение с тенденцией к средним значениям заметности. Примечательно, что даже при лабораторной аннотации почти половина артефактов DeSRA имеет заметность ниже 50% — то есть большинство наблюдателей их не замечают.

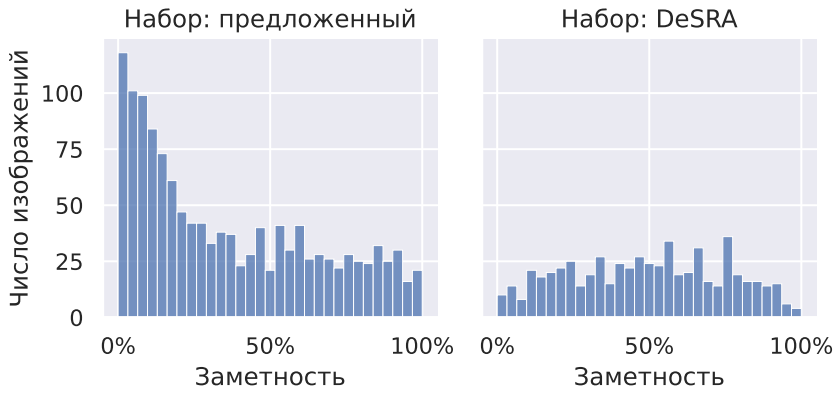

Рис. 7. Распределение заметности для предложенного набора данных (на основе Open Images) и для набора DeSRA (для которого в рамках данной работы были собраны аннотации заметности).

Этот результат подтверждает, что двоичные маски недостаточны для точной оценки артефактов суперразрешения.

4.2. Оценка устойчивости моделей суперразрешения к артефактам. В рамках данной работы была проведена оценка устойчивости 11 популярных моделей SR к генерации артефактов. Для этого использовалась процедура подготовки из разд. 3, но без какой-либо ручной фильтрации масок. Для каждого метода детектирования артефактов выбирались 10 наиболее сильных кандидатов артефактов на метод SR, всего 653 кандидата. Затем посредством краудсорсинга были собраны оценки заметности для этих кандидатов. В результатах приводятся общее число масок артефактов, найденных методом детектирования, их средняя заметность, и количество “уверенных” масок, соответствующих сильно видимым артефактам (50% заметности или выше).

Таблица 4.2 содержит результаты, сгруппированные по моделям SR. DRCT [5] и HAT-L [2], обе на основе трансформеров, показывают отличные результаты почти без заметных артефактов. Далее следуют три диффузионные модели: SinSR [19], ResShift [25] и StableSR [15].

Метод SR	Тип	Масок найдено	Средняя заметность$^{\downarrow}$	Заметных масок найдено$^{\downarrow}$
DRCT [5]	Трансформер	43	**11.85%**	**1**
HAT-L [2]	Трансформер	53	13.53%	**1**
SinSR [19]	Диффузионный	60	19.39%	3
ResShift [25]	Диффузионный	60	20.22%	7
StableSR [15]	Диффузионный	60	28.55%	14
RealSR [6]	Свёрточный	51	28.70%	10
SeeSR [20]	Диффузионный	61	32.34%	14
GFPGAN [16]	Свёрточный	45	32.74%	11
SwinIR [12]	Трансформер	59	41.09%	17
SUPIR [24]	Диффузионный	70	45.29%	20
RealESRGAN [17]	Свёрточный	61	48.42%	19

Таблица 4.2. Результаты краудсорсинга заметности по моделям суперразрешения.

Метод	Масок найдено	Средняя заметность$^{\downarrow}$	Заметных масок найдено$^{\downarrow}$	Объединённая оценка
LDL (t=0.005)	12	77.11%	11	8.48%
bd_jup (t=0.1)	110	17.03%	13	2.21%
LDL (t=0.0005)	51	34.84%	15	5.23%
LDL (t=0.001)	40	43.00%	16	6.88%
ssm_jup (t=0.15)	110	23.09%	20	4.62%
SSIM (t=0.55)	74	36.62%	26	9.52%
DeSRA	110	32.03%	31	9.93%
DISTS (t=0.25)	108	38.80%	**38**	**14.75%**

Таблица 4.3. Результаты краудсорсинга заметности по методам детектирования артефактов. Порог для каждого метода указан как t=0.xx.

Далее, таблица 4.3 содержит результаты, сгруппированные по методам детектирования артефактов. Метод оценки качества изображений DISTS [3] показал наилучшую способность находить заметные артефакты; на втором месте — метод детектирования артефактов DeSRA [21].

Исходное изображение	Корр. Пирсона	Корр. Спирмана
image1	1.00	1.00
image2	1.00	0.95
image3	0.93	0.74
image4	0.84	0.95
image5	0.76	0.40
image6	0.47	0.00
image7	0.31	0.20
image8	0.23	0.40
image9	0.22	0.32
image10	0.21	0.40
image11	-0.63	-0.80
image12	-0.76	-0.95

Таблица 4.4. Корреляция между субъективными оценками и $(1 - p)$, где p — оценка заметности. Субъективные оценки вычислялись независимо для каждой группы изображений, имеющих одно и то же входное изображение низкого разрешения.

Следует отметить, что LDL с порогом 0.005 находит области с сильно видимыми артефактами, но значительно уступает другим методам по общему числу заметных масок. Для учёта обоих показателей, числа найденных масок и средней заметности, они были перемножены аналогично произведению precision $\times$ recall из [21]; результаты приведены в последнем столбце. LDL была также протестирована с двумя более низкими порогами, что увеличило общее число найденных масок. Однако, эти маски в основном соответствовали малозаметным артефактам, что привело к худшему объединённому показателю.

4.3. Субъективное качество изображений с артефактами. Когда изображение после суперразрешения содержит артефакт, его субъективное качество обычно снижается, что отражается в оценках ассессоров. Это наблюдение подчёркивает важность детектирования артефактов, поскольку изображения без артефактов, как правило, воспринимаются визуально более приятными.

Метод SR	Тип	Масок найдено	Средняя заметность$^{\downarrow}$	Заметных масок найдено$^{\downarrow}$
SeeSR	Диффузионный	61	**7.15%**	0
ResShift	Диффузионный	60	<u>7.20%</u>	0
SinSR	Диффузионный	60	9.19%	2
HAT-L	Трансформер	50	9.60%	1
DRCT	Трансформер	50	13.33%	0
SwinIR	Трансформер	60	17.12%	4
RealESRGAN	Свёрточный	60	17.46%	5
SUPIR	Диффузионный	62	17.56%	6

Таблица 4.5. Краудсорсинговая оценка заметности артефактов для моделей суперразрешения на 6 наборах данных.

Для проверки этого утверждения было проведено парное субъективное сравнение изображений из предложенного набора данных. Участникам показывались пары изображений, полученных методом SR, и они выбирали то, которое им больше нравилось. В исследовании приняли участие 842 человека. Итоговые оценки рассчитывались с использованием модели Брэдли–Терри и включали 16 720 голосов.

Затем была проанализирована корреляция между этими субъективными оценками и $(1 - p)$, где p — оценка заметности артефакта на изображении. Как показано в таблице 4.4, корреляция была положительной в большинстве случаев, что подтверждает гипотезу о том, что изображения без артефактов получают более высокие субъективные оценки.

4.4. Субъективная оценка на дополнительных наборах данных суперразрешения. В рамках данной работы была проведена дополнительная субъективная оценка на 6 широко используемых наборах изображений [1,4,13,18,23], следуя методике, описанной в разд. 4.2. Всего в этой оценке использовалось 420 исходных изображений, каждое из которых было обработано 8 моделями SR.

В таблицах 4.5 и 4.6 показаны результаты, сгруппированные соответственно по моделям SR и по методам детектирования артефактов. Интересно, что модели SR демонстрируют гораздо лучшую устойчивость к артефактам, чем в основном сравнении в разд. 4.2, вероятно,

Метод	Масок найдено	Средняя заметность$^{\downarrow}$	Заметных масок найдено$^{\downarrow}$	Объединённая оценка
ssm_jup (t=0.15)	80	7.42%	1	0.07
bd_jup (t=0.1)	80	7.34%	2	0.15
LDL (t=0.005)	80	9.46%	2	0.19
DeSRA	74	12.21%	3	0.37
DISTS (t=0.25)	76	**18.03%**	**9**	**1.62**

Таблица 4.6. Краудсорсинговая оценка заметности артефактов для методов детектирования артефактов на 6 наборах данных. Порог для каждого метода указан как t=0.xx.

потому что эти наборы данных часто используются при обучении и оценке SR.

§5. Заключение

В данной работе рассматривается проблема визуальных артефактов в задаче суперразрешения одиночного изображения — проблема, которая остаётся актуальной даже для новейших и наиболее мощных моделей, таких как [24]. Артефакты следует характеризовать через их *заметность* для человеческих наблюдателей, а не через двоичные маски, поскольку восприятие различных артефактов сильно варьируется. Это подтверждается экспериментальным результатом: большинство участников не замечают почти половину артефактов в наборе данных DeSRA, размеченном в лабораторных условиях.

Основной вклад работы заключается в новом наборе данных, содержащем 1302 примера артефактов с аннотациями заметности для 11 современных методов SISR. Также представлены аннотации заметности для всех 593 артефактов из набора данных DeSRA.

Результаты данного исследования выходят за рамки области суперразрешения. Концепция заметности артефактов, вероятно, применима и к другим задачам обработки и восстановления изображений, таким как нейросетевое сжатие. Методы оценки качества, учитывающие заметность, могут направить разработку методов обработки изображений на структурированные области, где артефакты наиболее заметны.

Следует отметить ограничения данного исследования. Маски артефактов в представленном наборе данных являются достаточно приближёнными, поскольку даже для человеческих ассессоров точное определение границ артефактов является нетривиальной задачей из-за их локальной согласованности. Инициализация кандидатов масок выполнялась с использованием существующих методов детектирования, что вносит определённую неточность.

Существует несколько направлений для будущих исследований. Естественным расширением является применение подхода к задаче суперразрешения видео. В данной работе изображения рассматриваются по отдельности, и межкадровые артефакты, такие как мерцание, не затрагиваются. Наконец, более мощные модели, такие как [24], начинают переходить от простых искажений текстур к более семантическим артефактам, таким как замена объектов, что требует дальнейшего изучения.

Наборы данных с аннотациями заметности доступны по ссылке `https://drive.google.com/drive/folders/1Rue6vPgXnxZTFw0aJvWWcM2mZlRlGtg8?usp=sharing`.

Список литературы

1. M. Bevilacqua, A. Roumy, and C. Guillemot, *Set5 and Set14 Datasets*, 2012. Available: `https://figshare.com/articles/dataset/BSD100_Set5_Set14_Urban100/21586188`.
2. X. Chen, X. Wang, W. Zhang, X. Kong, Y. Qiao, J. Zhou, and C. Dong, *HAT: Hybrid Attention Transformer for Image Restoration*, ArXiv preprint arXiv:2309.05239 (2023).
3. K. Ding, K. Ma, S. Wang, and E. P. Simoncelli, *Image Quality Assessment: Unifying Structure and Texture Similarity.* — IEEE Transactions on Pattern Analysis and Machine Intelligence **44** (2022), no. 5, 2567–2581.
4. C. Dong and C. C. Loy, *General-100 Dataset*, 2016. Available: `http://mmlab.ie.cuhk.edu.hk/projects/FSRCNN.html`.
5. C.-C. Hsu, C.-M. Lee, and Y.-S. Chou, *DRCT: Saving Image Super-Resolution Away from Information Bottleneck*, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2024, pp. 6133–6142.
6. X. Ji, Y. Cao, Y. Tai, C. Wang, J. Li, and F. Huang, *Real-World Super-Resolution via Kernel Estimation and Noise Injection*, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2020.
7. N. Karetin and I. Molodetskikh, *Video Upscalers Benchmark*, 2022. Available: `https://videoprocessing.ai/benchmarks/video-upscalers.html`.

8. A. Kirillova, E. Lyapustin, A. Antsiferova, and D. Vatolin, *ERQA: Edge-Restoration Quality Assessment for Video Super-Resolution*, in: Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications — Volume 4: VISAPP, 2022, pp. 315–322.
9. A. Kuznetsova, H. Rom, N. Alldrin, J. Uijlings, I. Krasin, J. Pont-Tuset, S. Kamali, S. Popov, M. Malloci, A. Kolesnikov, *et al.*, *The Open Images Dataset V4: Unified Image Classification, Object Detection, and Visual Relationship Detection at Scale.* — International Journal of Computer Vision **128** (2020), no. 7, 1956–1981.
10. C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, *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, 2017, pp. 4681–4690.
11. J. Liang, H. Zeng, and L. Zhang, *Details or Artifacts: A Locally Discriminative Learning Approach to Realistic Image Super-Resolution*, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 5657–5666.
12. J. Liang, J. Cao, G. Sun, K. Zhang, L. Van Gool, and R. Timofte, *SwinIR: Image Restoration Using Swin Transformer*, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 1833–1844.
13. D. Martin, C. Fowlkes, and J. Malik, *BSDS200 Dataset*, 2001. Available: `https://huggingface.co/datasets/goodfellowliu/BSDS200`.
14. R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, *High-Resolution Image Synthesis with Latent Diffusion Models*, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022, pp. 10684–10695.
15. J. Wang, Z. Yue, S. Zhou, K. C. K. Chan, and C. C. Loy, *Exploiting Diffusion Prior for Real-World Image Super-Resolution.* — International Journal of Computer Vision (2024).
16. X. Wang, Y. Li, H. Zhang, and Y. Shan, *Towards Real-World Blind Face Restoration with Generative Facial Prior*, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021, pp. 9168–9178.
17. X. Wang, L. Xie, C. Dong, and Y. Shan, *Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data*, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 1905–1914.
18. X. Wang, K. Yu, *et al.*, *Historical Dataset*, 2018. Available: `https://github.com/xinntao/BasicSR`.
19. Y. Wang, W. Yang, X. Chen, Y. Wang, L. Guo, L.-P. Chau, Z. Liu, Y. Qiao, A. C. Kot, and B. Wen, *SinSR: Diffusion-Based Image Super-Resolution in a Single Step*, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 25796–25805.
20. R. Wu, T. Yang, L. Sun, Z. Zhang, S. Li, and L. Zhang, *SeeSR: Towards Semantics-Aware Real-World Image Super-Resolution*, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 25456–25467.
21. L. Xie, X. Wang, X. Chen, G. Li, Y. Shan, J. Zhou, and C. Dong, *DeSRA: Detect and Delete the Artifacts of GAN-Based Real-World Super-Resolution Models*, ArXiv preprint arXiv:2307.02457 (2023).

22. L. Xie, X. Wang, S. Shi, J. Gu, C. Dong, and Y. Shan, *Mitigating Artifacts in Real-World Video Super-Resolution Models*, in: Proceedings of the AAAI Conference on Artificial Intelligence **37**, 2023, pp. 2956–2964.
23. J. Yang, J. Wright, and T. Huang, *T91 Super-Resolution Dataset*, 2010. Available: `https://github.com/open-mmlab/mmsr`.
24. F. Yu, J. Gu, Z. Li, J. Hu, X. Kong, X. Wang, J. He, Y. Qiao, and C. Dong, *Scaling Up to Excellence: Practicing Model Scaling for Photo-Realistic Image Restoration in the Wild*, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024, pp. 25669–25680.
25. Z. Yue, J. Wang, and C. C. Loy, *ResShift: Efficient Diffusion Model for Image Super-Resolution by Residual Shifting*, in: Advances in Neural Information Processing Systems **36**, 2023, pp. 13294–13307.
26. K. Zhang, J. Liang, L. Van Gool, and R. Timofte, *Designing a Practical Degradation Model for Deep Blind Image Super-Resolution*, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 4791–4800.
27. L. Zhang, Z. Xu, C. Barnes, Y. Zhou, Q. Liu, H. Zhang, S. Amirghodsi, Z. Lin, E. Shechtman, and J. Shi, *Perceptual Artifacts Localization for Image Synthesis Tasks*, in: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023, pp. 7579–7590.
28. L. Zhang, Y. Zhou, C. Barnes, S. Amirghodsi, Z. Lin, E. Shechtman, and J. Shi, *Perceptual Artifacts Localization for Inpainting*, ArXiv preprint arXiv:2208.03357 (2022).

Molodetskikh I. A., Malyshev K. V., Mirgaleev M. R., Bogatyrev E.N., Vatolin D. S. Methodology for collecting and analyzing subjective prominence of artifacts in learning-based super-resolution.

Modern image super-resolution methods based on deep neural networks can generate highly detailed and visually appealing results, but they often introduce artifacts—incorrect details that degrade perceptual image quality. Importantly, the human perception of such artifacts varies: some are barely noticeable, while others severely affect visual quality. Existing datasets and artifact detection methods typically use binary labeling ("artifact present/absent"), which does not reflect subjective visibility and makes it impossible to quantitatively assess its impact on perceived quality.

This work proposes a methodology for subjective annotation of super-resolution artifacts with respect to their perceptual visibility, using a crowdsourcing approach with quality control. We describe the data preparation pipeline, including mask post-processing to improve assessors' perception, as well as the approach to aggregating subjective evaluations. The resulting dataset contains over 1,300 artifacts produced by eleven modern super-resolution models. We also collected visibility annotations for 593 images from the existing DeSRA artifact dataset and found that

most of its artifacts are barely noticeable to human observers despite being labeled in laboratory conditions.

The proposed dataset and methodology enable the study of artifacts from the perspective of human perception and can be used to train artifact detectors focused on perceptually significant distortions.

Поступило 20 марта 2026 г.

Московский государственный университет
имени М.В. Ломоносова,
Россия, 119991, Москва, Ленинские горы, д. 1.
Центр искусственного интеллекта МГУ.
E-mail: ivan.molodetskikh@graphics.cs.msu.ru

Московский государственный университет
имени М.В. Ломоносова,
Россия, 119991, Москва, Ленинские горы, д. 1.
E-mail: kirill.malyshev@graphics.cs.msu.ru

Московский государственный университет
имени М.В. Ломоносова,
Россия, 119991, Москва, Ленинские горы, д. 1.
E-mail: mark.mirgaleev@graphics.cs.msu.ru

Московский государственный университет
имени М.В. Ломоносова,
Россия, 119991, Москва, Ленинские горы, д. 1.
E-mail: evgeney.zimin@graphics.cs.msu.ru

Московский государственный университет
имени М.В. Ломоносова,
Россия, 119991, Москва, Ленинские горы, д. 1.
Центр искусственного интеллекта МГУ.
Институт искусственного интеллекта МГУ.
E-mail: dmitriy@graphics.cs.msu.ru