



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

Препринты ИПМ • Препринт № 18 за 2025 г.

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

**А.К. Денисов, А.Д. Жданов,
Е.Д. Бирюков, С.В. Быковский**

**Метод повышения качества
изображения на базе
диффузионной модели с
использованием опорных
изображений**

Статья доступна по лицензии
[Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/)



Рекомендуемая форма библиографической ссылки: Метод повышения качества изображения на базе диффузионной модели с использованием опорных изображений / А.К. Денисов [и др.] // Препринты ИПМ им. М.В.Келдыша. 2025. № 18. 16 с. EDN: [USPIGY](https://uspi.gy)
<https://library.keldysh.ru/preprint.asp?id=2025-18>

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

**А.К. Денисов, А.Д. Жданов, Е.Д. Бирюков,
С.В. Быковский**

**Метод повышения качества
изображения на базе диффузионной
модели с использованием опорных
изображений**

Москва – 2025

Денисов А.К., Жданов А.Д., Бирюков Е.Д., Быковский С.В.

Метод повышения качества изображения на базе диффузионной модели с использованием опорных изображений

В статье представлен метод восстановления изображений, основанный на применении диффузионной модели с использованием опорных изображений, содержащих дополнительную информацию (референсов). Предложенный подход направлен на улучшение качества восстановления деталей изображения за счет интеграции информации из референсов. Диффузионные модели, известные своей способностью генерировать изображения высокой четкости, применяются для устранения ограничений, присущих сверточным нейронным сетям, таких как недостаточное качество изображений, восстановленных после сильных деградаций. При этом качество изображения, полученного с помощью диффузионной модели, можно дополнительно увеличить с помощью внедрения информации с опорных изображений. В данной работе описывается предлагаемая архитектура модели, описан способ ее обучения и показаны результаты сравнения с моделью, не использующей опорные изображения. Результаты экспериментов подтверждают эффективность предложенного подхода, в частности, улучшение качества результирующего изображения более чем на 9% по метрике PSNR и на 14% по метрике LPIPS.

Ключевые слова: восстановление изображений, нейронные сети, диффузионные модели, super-resolution, глубокое обучение

*Aleksei Konstantinovich Denisov, Andrei Dmitrievich Zhdanov,
Elissey Dmitrievich Birukov, Sergei Vyacheslavovich Bykovskii*

A method for enhancing image quality based on a reference diffusion model

The paper presents an image restoration method based on the application of a diffusion model utilizing reference images. The proposed approach aims to enhance the quality of restored image details by integrating information from references. Diffusion models, known for their ability to generate high-resolution images, are employed to address the limitations of convolutional neural networks, such as the insufficient quality of images restored after severe degradation. Furthermore, the quality of images generated by the diffusion model can be further improved by incorporating information from reference images. This study describes the proposed model architecture, training procedure, and comparison results with the no-reference model. The experimental results confirm the effectiveness of the proposed approach, specifically demonstrating an improvement in the quality of the resulting image by more than 9% in terms of the PSNR metric and by 14% in terms of the LPIPS metric.

Key words: image restoration, neural networks, diffusion models, super-resolution, deep learning

Введение

Задача увеличения разрешения изображений (super-resolution, далее – задача SR) является одной из ключевых проблем в области компьютерного зрения и обработки изображений. Она находит применение в самых разных сферах: от медицины и спутниковой съемки до систем видеонаблюдения и развлечений. Цель задачи SR заключается в восстановлении высококачественного изображения с высоким разрешением из его версии с низким разрешением, при этом сохраняя детали и текстуры оригинала. Сложные деградации в изображениях могут возникать по различным причинам, связанным с физическими ограничениями оборудования, воздействием окружающей среды или цифровой обработкой. Вот основные источники таких деградаций:

- аппаратные ограничения:
 - низкое разрешение сенсоров: камеры с низким разрешением фиксируют меньше деталей, что приводит к потере мелких текстур и размытию изображения;
 - оптические искажения: абберации, виньетирование, искажения перспективы или размытость краев;
 - шум сенсоров: цифровые камеры, особенно в условиях слабого освещения, могут вносить значительный шум, связанный с электроникой;
- воздействие окружающей среды:
 - погодные условия: дождь, туман, снег, пыль или дым могут значительно ухудшить видимость и снизить резкость изображения;
 - движение: смазывание изображения возникает из-за движения объекта или камеры, особенно при длительной выдержке;
 - освещение: недостаточное освещение или резкие перепады яркости могут приводить к недостаткам изображения, таким как переэкспонирование или затемнение;
- процессы обработки и передачи данных:
 - компрессия: использование методов сжатия изображений (например, JPEG) может привести к потере качества, появлению артефактов и снижению детализации;
 - интерполяция: при масштабировании изображений низкого разрешения часто используются методы интерполяции, которые могут вызывать артефакты.

Эти сложные деградации существенно осложняют задачи обработки изображений и требуют применения мощных алгоритмов восстановления, таких как методы на основе глубокого обучения, в том числе с использованием диффузионных моделей.

В последние годы область поиска решений задачи SR активно развивается благодаря успехам методов глубокого обучения. Современные подходы, такие как сверточные нейронные сети (CNN), генеративно-состязательные сети (GAN, generative adversarial networks), а также трансформеры, значительно улучшили качество восстановления изображений. Однако большинство этих методов сталкивается с рядом ограничений, включая сложности в восстановлении мелких деталей, особенно при значительных искажениях изображения, а также недостаток информации в исходных данных.

Одним из перспективных направлений в решении задачи SR является применение диффузионных моделей, которые зарекомендовали себя как мощный инструмент генерации изображений с высокой четкостью и детализацией. Эти модели используют процесс обратной диффузии для восстановления сложных структур и текстур, что делает их особенно полезными для задач восстановления изображений с сильными потерями информации. Однако диффузионные модели в своей базовой форме могут страдать от несоответствия сгенерированных деталей реальной структуре изображения, что ограничивает их применение в задачах SR.

В этой работе предлагается подход, сочетающий возможности диффузионных моделей с использованием информации из опорных изображений. Такой подход позволяет дополнить недостающие данные, полученные с помощью диффузионной модели, информацией из дополнительных источников. Это позволяет не только восстановить мельчайшие детали, но и повысить общую точность и реалистичность результата.

В работе решается задача разработки и исследования метода увеличения качества изображения с помощью диффузионной модели с использованием опорных изображений. Описаны метод получения тренировочных данных для обучения модели с референсами, архитектура модуля диффузионной модели, позволяющего интегрировать информацию с опорных изображений в признаки внутри диффузионной модели, а также приведены результаты сравнения двух моделей, обученных на одних и тех же данных с использованием и без использования референсов.

Обзор литературы

Задача увеличения разрешения изображений (задача SR) за последние годы существенно изменилась благодаря внедрению глубокого обучения, которое трансформировало подходы к обработке изображений.

Сверточные нейронные сети (CNN). Одной из первых работ, заложивших основу применения сверточных нейронных сетей (CNN) для SR, стала SRCNN, предложенная Донг (Dong) и соавторами (2016) [1]. Эта работа показала превосходство CNN над классическими подходами к увеличению разрешения. В дальнейшем Ким (Kim) и соавторы (2016) разработали архитектуру VDSR [2], основанную на очень глубокой нейронной сети, которая

использовала метод восстановления остаточной информации (residual learning) для повышения качества изображений. Позже Лим (Lim) и соавторы (2017) представили EDSR [3] – улучшенную архитектуру, оптимизированную за счет удаления избыточных компонентов, что позволило повысить эффективность обучения и качество получаемого изображения.

Существенный вклад в развитие методов решения задачи SR внесли генеративно-состязательные сети (GAN). Например, в SRGAN Ледига (Ledig) и соавторов (2017) [4] использовалось состязательное обучение для создания визуально привлекательных изображений, при этом авторы ориентировались на качество восприятия изображений человеком. Дальнейшим развитием стали подходы BSRGAN (Blind Super-Resolution) [5] и Real-ESRGAN [6] – подходы в области решений задачи SR, которые фокусируются на моделировании синтетических деградаций изображений для улучшения работы моделей в реальных условиях. Они решают ключевую проблему области: стандартные методы часто обучаются на данных с искусственными и упрощенными деградациями (например, простое понижение разрешения через бикубическую интерполяцию), которые не соответствуют более сложным и разнообразным искажениям, встречающимся в реальной жизни. В то же время сложный многостадийный конвейер деградаций, применяющийся в Real-ESRGAN, включает в себя различные искажения, такие как аддитивный шум, JPEG-компрессию, свертку с различными ядрами, имитацию кольцевых искажений на изображении и т.д.

Использование опорных изображений. Наряду с методами увеличения разрешения на основе одного изображения (single image super-resolution, SISR) важным направлением исследований стало использование опорных изображений для повышения разрешения (reference-based super-resolution, RefSR). Ключевой особенностью RefSR является применение дополнительных изображений, которые содержат фрагменты той же сцены в высоком разрешении. Метод SRNTT [7] предлагает многоуровневую архитектуру, позволяющую эффективно переносить текстуры между фрагментами целевого и опорного изображений. Ванг (Wang) и соавторы (2020) [8] разработали фреймворк C2-Matching, который включает три этапа обучения: первые два используют контрастивное обучение для оптимизации сети извлечения признаков из целевого и опорного изображений, а на заключительном этапе обучается сеть, восстанавливающая изображение на основе сопоставленных признаков. Одним из значительных достижений является метод DATSR, предложенный Цао (Cao) и соавторами (2022) [9], который применяет механизм адаптивного внимания (deformable attention). Это позволяет модели динамически концентрироваться на релевантных признаках в опорных изображениях, что особенно полезно при наличии пространственных вариаций, и способствует более качественному восстановлению изображения с высоким разрешением. Одной из ключевых сложностей RefSR методов остается ограниченный объем данных для обучения. В работе Жанг (Zhang) и соавторов

(2021) [10] представлен большой и разнообразный набор данных (LMR), предназначенный для обучения и оценки различных RefSR моделей. Также в этой работе предложен подход, который позволяет использовать сразу несколько опорных изображений, что улучшает качество итогового изображения. Еще один подход был реализован в работе [11], где используется предобученная диффузионная модель для генерации опорного изображения. Детали, перенесенные с генерированного референса, помогают улучшить качество целевого изображения.

Эти исследования демонстрируют значительный прогресс в развитии методов RefSR и раскрывают их потенциал, однако их существенным ограничением является работа с изображениями, полученными с помощью синтетических деградаций (чаще всего бикубической интерполяции), в то время как для работы в реальном сценарии необходимо справляться со сложными, неизвестными заранее искажениями.

Диффузионные модели. В то время как большинство методов обработки изображений основываются на использовании сверточных нейронных сетей и трансформеров, диффузионные модели стали активно набирать популярность в последние годы. Эти модели зарекомендовали себя как мощный инструмент в задачах генеративного моделирования благодаря их способности усваивать структуру сложных распределений данных. Принцип работы диффузионных моделей заключается в итеративном преобразовании зашумленных данных в образцы, соответствующие целевому распределению.

Значительный вклад в развитие диффузионных моделей внесла работа Хо (Ho) и соавторов, в которой представлен метод DDPM [12]. Этот подход использует прямой процесс диффузии для добавления шума к данным, а затем обучаемый обратный процесс для последовательного его удаления, что позволяет генерировать новые данные. В работе также акцентируется внимание на важности правильного управления интенсивностью шума на каждом этапе диффузии и параметризации модели, что стало основой для дальнейших исследований.

На базе DDPM был предложен метод DDIM, разработанный Сонг (Song) и соавторами [13], который сосредоточен на повышении эффективности генерации. В отличие от DDPM, DDIM использует неявный процесс семплинга, что позволяет ускорить генерацию, сохраняя при этом высокое качество выходных данных. Этот подход демонстрирует универсальность диффузионных моделей и их пригодность для задач с ограничениями по вычислительным ресурсам.

Следующим шагом в развитии диффузионных моделей стало появление скрытых диффузионных моделей (Latent Diffusion Models, далее LDM) [14]. В основе LDM лежит идея выполнения диффузионных процессов в скрытом пространстве низкой размерности, например, в пространстве латентных переменных, полученных вариационным автоэнкодером, вместо работы в исходном пространстве данных. Это существенно снижает вычислительные

затраты, сохраняя высокое качество генерации, и делает подход пригодным для таких задач, как создание изображений и видео с высоким разрешением.

Кроме задач генерации, диффузионные модели успешно применяются в задачах преобразования изображения в изображение (image-to-image), включая восстановление. В этой области используются два подхода. В первом случае диффузионная модель обучается с нуля под конкретную задачу, где изображение для восстановления подается вместе с шумом в качестве входных данных. Так работают методы SRDiff [15] и LDM [14]. Второй подход предполагает использование предварительно обученных генеративных моделей с добавлением специальных адаптеров для решения конкретных задач. Примером является метод SUPIR [16], в котором в качестве базовой модели используется Stable Diffusion XL [17].

Таким образом, методы super-resolution на основе диффузионных моделей являются одним из вариантов продолжения развития методов восстановления изображений на основе глубокого обучения. Одновременно с этим можно заметить, что использование опорных изображений, несмотря на развитие этого направления в классических моделях, в данный момент редко используется в диффузионных моделях. Одним из примеров может служить метод COSER [18], но в данной работе используется не реальное опорное изображение, а сгенерированное с помощью другой генеративной модели. Мы исследуем влияние использования реальных опорных изображений на качество восстановления изображения диффузионной моделью.

Предложенный метод

В данной работе используется модель, принимающая на вход изображение низкого качества (LR), а также одно или несколько опорных изображений. За основу взята диффузионная модель, работающая в латентном пространстве путем выполнения процессов диффузии и восстановления не в исходном пространстве данных высокой размерности (например, пикселей изображения), а в компактном латентном пространстве. Такой подход позволяет значительно снизить вычислительные затраты при сохранении высокого качества результатов.

Чтобы обеспечить эффективную обработку, LDM сначала преобразует изображение высокой размерности x в латентное пространство меньшей размерности z с помощью предобученного энкодера \mathcal{E} . Это представление содержит ключевую структурную и семантическую информацию об изображении в сжатой форме: $z = \mathcal{E}(x)$, где \mathcal{E} представляет собой сжимающую часть вариационного автоэнкодера с квантованием векторов (VQ-VAE, vector-quantized variational autoencoder), оптимизированную для качественного кодирования.

После завершения процессов диффузии и восстановления в латентном пространстве полученное представление z преобразуется обратно в исходное

изображение с использованием VQ-VAE декодера \mathcal{D} : $\hat{x} = \mathcal{D}(z)$. Декодер отвечает за восстановление деталей, цветовой информации и текстур, опираясь на обученную структуру латентного пространства.

Прямой процесс диффузии постепенно добавляет гауссов шум к латентному представлению z на протяжении T дискретных шагов, превращая чистую латентную переменную z_0 в зашумленную переменную z_T . Обратный процесс параметризуется нейронной сетью ϵ_θ , которая предсказывает шум, добавленный на каждом шаге. Начиная с z_T , модель шаг за шагом удаляет шум для восстановления исходного латентного представления z_0 .

Целью обучения является предсказание шума ϵ , добавленного в процессе прямой диффузии, в процессе обучения минимизируется следующая функция потерь:

$$L_{LDM} := \mathbb{E}_{\mathcal{E}(x), \epsilon \sim \mathcal{N}(0,1), t} [\|\epsilon - \epsilon_\theta(z_t, t)\|_2^2].$$

В фазе тестирования (инференса) модель генерирует данные путем сэмплинга латентной переменной z_T из стандартного гауссова распределения: $z_T \sim \mathcal{N}(0, I)$. При этом для обуславливания модели на изображение низкого качества оно склеивается с z_T , и результат этой операции подается в модель. Затем применяется обратный процесс диффузии, чтобы поэтапно преобразовать z_T в чистое латентное представление z_0 . После этого декодер \mathcal{D} преобразует z_0 в изображение высокого разрешения: $\hat{x} = \mathcal{D}(z_0)$.

В качестве основы диффузионной модели используется сеть на основе архитектуры U-Net, состоящая из чередующихся ResNet-блоков и attention модулей. На рис. 1 показана общая схема используемой модели. \mathcal{E} и \mathcal{D} – VQ-VAE энкодер и декодер, E и D – энкодер и декодер диффузионной модели, E_{LR} и E_{Ref} – энкодеры для изображения низкого разрешения и опорных изображений соответственно.

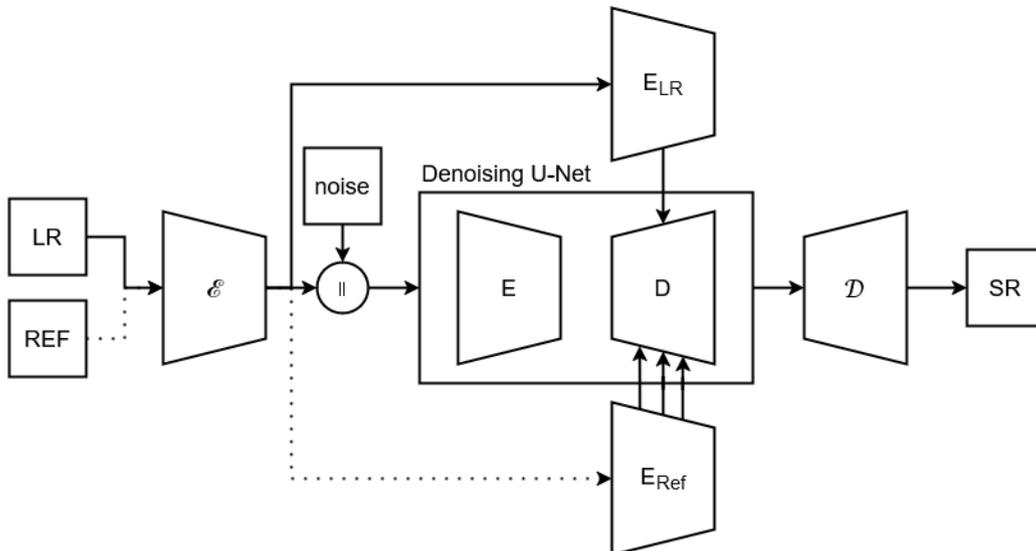


Рис. 1. Общая схема используемой модели

Для интеграции информации с опорных изображений используется следующий механизм. Опорные изображения и LR пропускаются через VQ-VAE энкодер \mathcal{E} , а затем через копию U-Net энкодера диффузионной модели (один для LR и один для референсов). Карты признаков с каждого уровня энкодера затем подаются на каждый уровень основного декодера, где происходит операция перекрестного внимания (cross-attention, CA) между признаками, извлеченными из LR и из референсов. Эта операция выполняется для каждого референса в отдельности, а затем результаты от каждого референса агрегируются и складываются с выходом соответствующего слоя декодера U-Net. В операции CA в качестве запроса (query) подаются признаки из LR, а в качестве ключа (key) и значения (value) используются признаки, извлеченные из опорных изображений. На рис. 2 приведена схема интеграции опорной информации. Символом «||» обозначена операция склеивания (конкатенации) по измерению каналов.

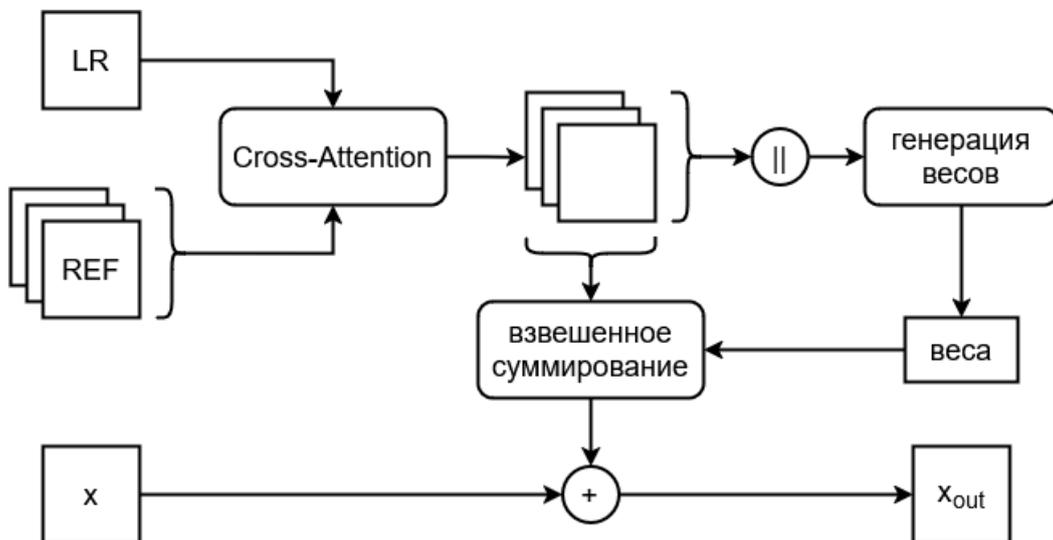


Рис. 2. Схема интеграции опорной информации

Результаты

Для проверки предлагаемого метода интеграции опорных изображений было обучено две модели: с и без предложенного выше механизма (LDM и LDM+CA соответственно). Архитектура модели LDM используется такая же, как в работе [14].

Так как обучение диффузионной модели требует больших временных затрат при условии ограниченных вычислительных мощностей, а также для облегчения задачи поиска опорных изображений при проверке предлагаемого метода было принято решение сузить область задачи до изображений птиц. В качестве набора данных использовались 80000 изображений птиц разных видов. Для получения универсальной модели, работающей с изображениями любых объектов, необходимо произвести обучение на большем количестве данных с разнообразными объектами на изображениях.

Модели были обучены на 900 тысячах фрагментов изображений размером 512 на 512 пикселей, в качестве модели деградации используются Real-ESRGAN деградации. Для получения опорных изображений на этапе обучения использовалась двухстадийная схема: на первом этапе с помощью CLIP [19] выбираются наиболее похожие на целевое изображения, на втором этапе с помощью специальной сети на основе ResNet-34 [20] выбираются 10 фрагментов 512 на 512 пикселей, наиболее похожих на целевой фрагмент. Модели обучались 2000000 итераций на восьми NVIDIA V100. Для валидации использовался набор из 20 изображений большого разрешения (>1 мегапикселя), к каждому из которых с помощью CLIP было найдено 10 опорных.

Для оценки качества восстановления изображений использовались следующие метрики: PSNR, SSIM, LPIPS [21], CLIP-IQA [22] и FID [23].

PSNR (Peak Signal-to-Noise Ratio) – это метрика, используемая для оценки качества восстановленных или сжатых изображений по сравнению с их эталонными версиями. Она измеряет разницу между двумя изображениями, выраженную в виде отношения максимальной мощности сигнала (интенсивности пикселей) к мощности шума (ошибки между пикселями эталонного и оцениваемого изображения). PSNR вычисляется следующим образом:

$$PSNR = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right),$$

где MAX_I – максимальное значение интенсивности пикселя (например, 255 для изображений с 8-битной глубиной цвета), а MSE (Mean Squared Error) – среднеквадратичная ошибка между пикселями двух изображений.

SSIM (Structural Similarity Index Measure) – это метрика, которая оценивает качество изображений, основываясь на человеческом восприятии структуры, яркости и контрастности. В отличие от PSNR, SSIM не просто сравнивает пиксели, а анализирует, насколько похожи структурные и визуальные элементы двух изображений. SSIM вычисляется следующим образом:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)},$$

где μ_x, μ_y – средние значения яркости двух изображений, σ_x^2, σ_y^2 – дисперсии яркости для каждого изображения, σ_{xy} – ковариация между яркостями двух изображений, C_1, C_2 – малые константы для предотвращения деления на ноль. Вышеперечисленные значения вычисляются в окрестности рассматриваемого пикселя, для оценки всего изображения применяется метод скользящего окна.

Метрика LPIPS измеряет расстояние между признаками, которые извлекаются из эталонного и оцениваемого изображений с использованием предобученной нейронной сети, например, AlexNet. Данная метрика

демонстрирует высокую корреляцию с субъективными оценками качества изображений человеком (перцептивное качество), что подтверждено авторами методики.

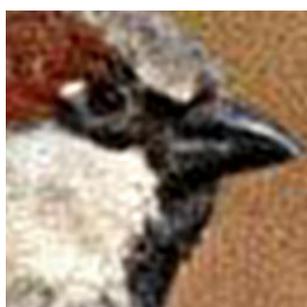
Метрика CLIP-IQA основана на мультимодальной модели CLIP, которая устанавливает соответствие между изображениями и их текстовыми описаниями. Для её расчёта используется косинусное расстояние между описанием анализируемого изображения, сгенерированного с помощью модели CLIP, и заранее заданным набором описаний для изображений высокого и низкого качества. Это позволяет оценивать качество изображений без необходимости наличия эталона.

Метрика FID (Fréchet Inception Distance) применяется для оценки качества генерации изображений. Она вычисляет расстояние между распределениями целевых данных и сгенерированных сэмплов. Для этого используется предобученная модель Inception v3, которая извлекает признаки с последнего слоя. Затем рассчитываются средние значения и ковариации этих признаков, что позволяет определить степень сходства между распределениями.

Таблица 1

Метод	PSNR↑	SSIM↑	LPIPS↓	CLIP-IQA↑	FID↓
LDM	21.252	0.6435	0.4186	0.6454	12.449
LDM + CA	23.312	0.6632	0.3582	0.6787	8.626

В таблице 1 приведены количественные сравнения между двумя обученными моделями на основе описанных метрик. Стрелками около названия метрики указано направление улучшения качества изображения (минимизация или максимизация метрики), жирным шрифтом выделены лучшие значения в столбцах. Видно, что модель, использующая опорные изображения, лучше базовой модели по всем метрикам, в частности, более чем на 2 dB (9%) по PSNR, на 0.02 (3%) по SSIM, на 0.06 (14%) по LPIPS, на 0.03 (5%) по CLIP-IQA и более чем 3.8 (30%) по FID. Это показывает преимущество использования дополнительной референсной информации при восстановлении изображения.



LR



LDM



LDM + CA



HR

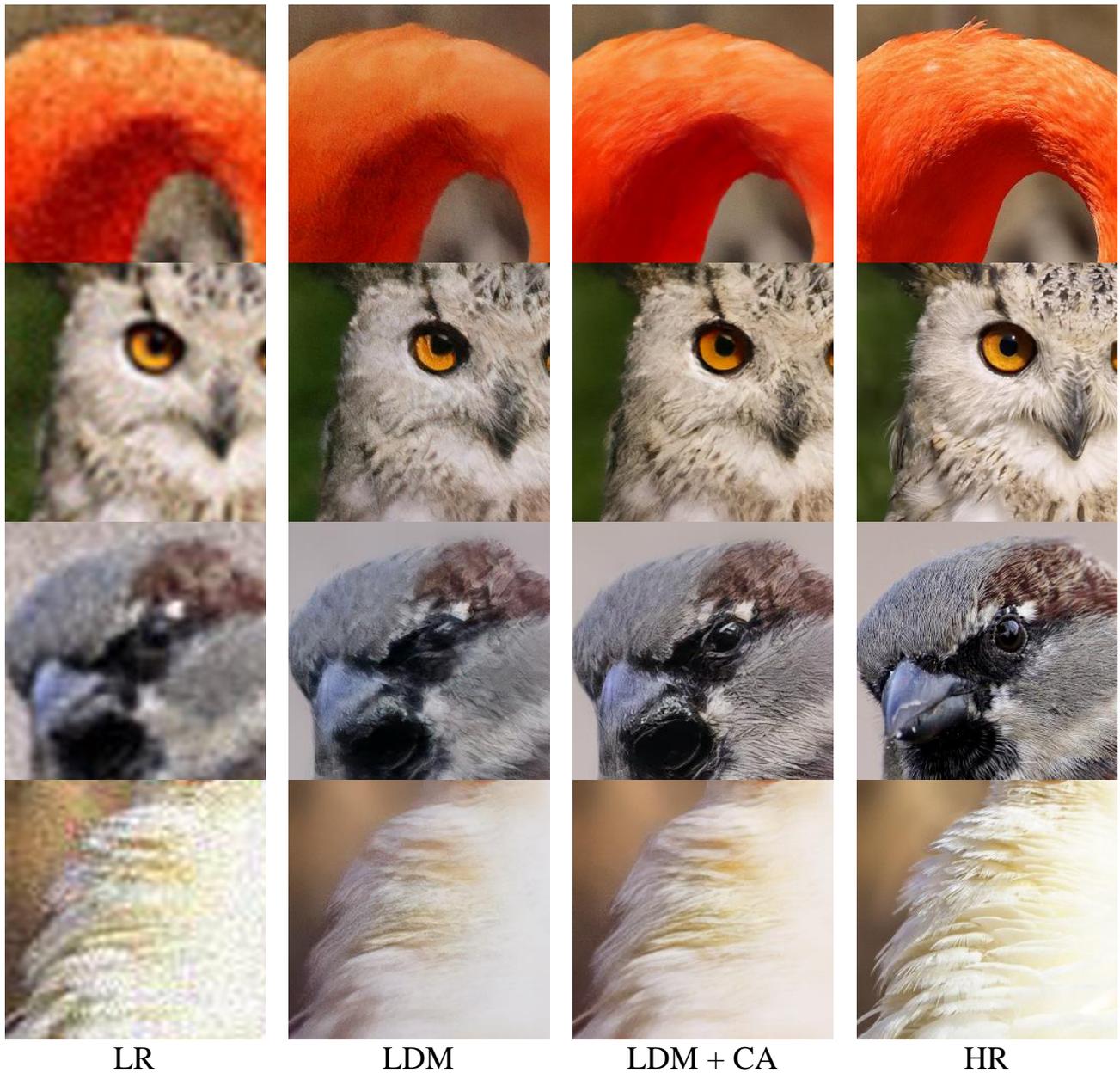


Рис. 3. Визуальные результаты. Слева направо приведены: входное изображение (LR), результат обработки моделью без опорных изображений (LDM), результат обработки моделью с опорными изображениями (LDM+CA), эталонное изображение (HR)

На рис. 3 приведены изображения для визуального сравнения результатов работы обеих моделей. В качестве опорных изображений в модель подавались изображения птицы того же вида, что и целевое изображение. Видно, что при использовании референсов на восстановленном изображении присутствует больше деталей и текстур, меньше шума и артефактов.

Заключение

Методы решения задачи увеличения качества изображений с использованием нейронных сетей активно развиваются в последние годы. Методы на основе опорных изображений стали развитием сверточных сетей, генерирующих результат за один проход. Параллельно им для решения задачи повышения качества изображений стали применяться диффузионные модели. В работе предложен метод использования информации с опорных изображений в диффузионной модели. Были обучены две диффузионных модели, одна из которых имеет внутри модуль для внедрения информации с опорных изображений. Данные модели были обучены для восстановления изображений после Real-ESRGAN деградаций. Модель, использующая опорные изображения, демонстрирует более высокое качество восстановленного изображения, что подтверждается визуальным анализом, а также с помощью количественной оценки с использованием различных метрик, в частности, более чем на 2 dB (9%) по PSNR и на 0.06 (14%) по LPIPS. Данный результат показывает положительный эффект от использования дополнительной информации с опорных изображений в диффузионной SR модели.

Библиографический список

1. Dong C., Loy C.C., He K., Tang X. Image Super-Resolution Using Deep Convolutional Networks // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2016. V. 38. P. 295–307.
2. Kim J., Lee K.M., Lee J.G. Accurate Image Super-Resolution Using Very Deep Convolutional Networks // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 1646–1654.
3. Lim B., Son S., Kim H., Nah S., Lee K.M. Enhanced Deep Residual Networks for Single Image Super-Resolution // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2017. P. 136–144.
4. Ledig C., Theis L., Huszár F., Caballero J., Cunningham A., Acosta A.U., Bishop M. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. P. 4681–4690.
5. Zhang K., Liang J., van Gool L., Timofte R. Designing a Practical Degradation Model for Deep Blind Image Super-Resolution. 2021.
6. Wang X., Xie L., Dong C., Shan Y. Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data // Proceedings of ICCVW. 2021.
7. Zhang Z., Wang Z., Lin Z.L., Qi H. Image Super-Resolution by Neural Texture Transfer // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019. P. 7974–7983.

8. Wang X., Zhang S., Li Y. C2-Matching: Cross-Scale Correspondence for Reference-Based Image Super-Resolution // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020. P. 10167–10176.
9. Cao J., Liang J., Zhang K., Li Y., Zhang Y., Wang W., Van Gool L. Reference-based Image Super-Resolution with Deformable Attention Transformer // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022. P. 1865–1874.
10. Zhang Z., Yu J., Zhang C., Xu Y. LMR: A Large-Scale Multi-Reference Dataset for Reference-based Super-Resolution // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021. P. 13692–13701.
11. Li G., Wei X., Zhao L., Liu Z., Sun J., Zheng Z., Zhang Q., Li H., Liu Z. Self-Reference Image Super-Resolution via Pre-trained Diffusion Large Model and Window Adjustable Transformer // Proceedings of the 31st ACM International Conference on Multimedia. 2023.
12. Ho J., Chen X., Srinivas A., Duan Y., Abbeel P. [Denoising Diffusion Probabilistic Models](#).
13. Song J., Sohl-Dickstein J., Kingma D.P., Kumar M., Ermon S., Poole B. [Denoising Diffusion Implicit Models](#).
14. Rombach R., Blattmann A., Lorenz D., Esser P., Ommer B. [High-Resolution Image Synthesis with Latent Diffusion Models](#).
15. Li H., Yang Y., Chang M., Feng H., Xu Z., Li Q., Chen Y. SRDiff: Single Image Super-Resolution with Diffusion Probabilistic Models // Neurocomputing. 2021. V. 479. P. 47–59.
16. Yu F., Gu J., Li Z., Hu J., Kong X., Wang X., He J., Qiao Y., Dong C. Scaling Up to Excellence: Practicing Model Scaling for Photo-Realistic Image Restoration in the Wild // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2024. P. 25669–25680.
17. Podell D., English Z., Lacey K., Blattmann A., Dockhorn T., Muller J., Penna J., Rombach R. [SDXL: Improving Latent Diffusion Models for High-Resolution Image Synthesis](#).
18. Sun H., Li W., Liu J., Chen H., Pei R., Zou X., Yan Y., Yang Y. CoSeR: Bridging Image and Language for Cognitive Super-Resolution // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2024.
19. Radford A., Kim J.W., Hallacy C., Ramesh A., Goh G., Agarwal S., Sastry G., Askell A., Mishkin P., Clark J., Krueger G., Sutskever I. Learning Transferable Visual Models From Natural Language Supervision // Proceedings of the International Conference on Machine Learning (ICML). 2021.
20. He K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 770–778.

21. Zhang R., Isola P., Efros A. A., Shechtman E., Wang O. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2018.
22. Wang J., Chan K., Loy C. C. Exploring CLIP for Assessing the Look and Feel of Images // Proceedings of the AAAI Conference on Artificial Intelligence. 2023.
23. Heusel M., Ramsauer H., Unterthiner T., Nessler B., Hochreiter S. GANs Trained by a Two Time-Scale Update Rule Converge to a Local Nash Equilibrium // Proceedings of NIPS. 2017.

Оглавление

Введение	3
Обзор литературы.....	4
Предложенный метод.....	7
Результаты.....	9
Заключение.....	13
Библиографический список.....	13