НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
март–апрель 2025 Том 25 № 2 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
March–April 2025 Vol. 25 No 2 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

doi: 10.17586/2226-1494-2025-25-2-321-327
УДК 004.032.26

Метод увеличения разрешения изображения с использованием референсных изображений на основе диффузионной модели

Алексей Константинович Денисов[1]✉, Сергей Вячеславович Быковский[2], Павел Валерьевич Кустарев[3]

[1,2,3] Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

[1] denisov@itmo.ru✉, https://orcid.org/0000-0001-8135-1723
[2] sergei_bykovskii@itmo.ru, https://orcid.org/0000-0003-4163-9743
[3] kustarev@itmo.ru, https://orcid.org/0000-0001-9326-0837

Аннотация

Введение. В настоящий момент активно развиваются различные методы восстановления изображений на основе методов глубокого машинного обучения. С помощью таких методов решаются задачи восстановления утраченных областей, подавления шумов и увеличения разрешения изображений. В задаче увеличения разрешения важную роль играют методы, основанные на применении референсных изображений, позволяющих восстановить недостающую информацию на основном изображении. Такие методы реализуются с использованием сверточных нейронных сетей, широко востребованных в задачах компьютерного зрения. В применяемых в настоящее время методах область изображения, не представленная на референсном изображении, часто отличается худшим качеством по сравнению с остальным изображением, что заметно визуально. Наряду со сверточными нейронными сетями в задачах восстановления изображений активно применяются диффузионные модели, позволяющие генерировать изображения с высоким качеством и четкостью, однако их недостатком часто бывает несоответствие сгенерированных деталей реальным. В работе обсуждается проблема улучшения качества восстановления изображений на основе применения референсных изображений с использованием диффузионной модели. **Метод.** Для получения хорошего конечного результата предложена гибридная архитектура нейронной сети диффузионной модели, состоящая из трех основных блоков: базового модуля диффузионной модели, модуля использования референсной информации и модуля слияния. Обучение предложенной гибридной модели, а также сравниваемой с ней сверточной нейронной сети, использующей референсные изображения, и диффузионной моделью выполнено с использованием набора данных Large-Scale Multi-Reference Dataset (LMR). **Основные результаты.** По результатам тестирования обученных моделей на тестовой выборке набора данных LMR проведено качественное (визуальное) и количественное сравнение работы трех моделей. Гибридная модель продемонстрировала более высокое качество, четкость и однородность изображения в сравнении со сверточной нейронной сетью с использованием референсных изображений и лучшее восстановление реальных деталей по сравнению с диффузионной моделью. Количественные оценки подтвердили, что гибридная модель также показала более высокие результаты по сравнению с остальными моделями. **Обсуждение.** Результаты работы могут быть использованы для увеличения разрешения любых изображений с использованием референсной информации.

Ключевые слова

обработка изображений, диффузионные модели, super-resolution, глубокое обучение, восстановление изображений

Ссылка для цитирования: Денисов А.К., Быковский С.В., Кустарев П.В. Метод увеличения разрешения изображения с использованием референсных изображений на основе диффузионной модели // Научно-технический вестник информационных технологий, механики и оптики. 2025. Т. 25, № 2. С. 321–327. doi: 10.17586/2226-1494-2025-25-2-321-327

Reference-based diffusion model for super-resolution

Aleksei K. Denisov[1]✉, Sergei V. Bykovskii[2], Pavel V. Kustarev[3]

[1,2,3] ITMO University, Saint Petersburg, 197101, Russian Federation

[1] denisov@itmo.ru✉, https://orcid.org/0000-0001-8135-1723

[2] sergei_bykovskii@itmo.ru, https://orcid.org/0000-0003-4163-9743

[3] kustarev@itmo.ru, https://orcid.org/0000-0001-9326-0837

Abstract

This article is devoted to digital image processing algorithms, namely, super-resolution task. Currently, various methods of image restoration based on deep learning are actively developing. These methods are used to solve image restoration problems, such as inpainting, denoising, and super-resolution. One important class of super-resolution methods is reference-based super-resolution that allows restoring the missing information in the main image using reference images. Methods of this class are mainly represented by convolutional neural networks which are widely used in computer vision problems. Despite the significant achievements of existing methods, they have one significant drawback: the image area not represented in the reference image often has worse quality compared to the rest of the image, which is clearly visible to the observer. In addition to convolutional neural networks, diffusion models are actively used in image restoration problems. They are capable of generating images with high quality and diverse fine details but suffer from a lack of fidelity between the generated details and the real ones. The aim of this work is to improve the quality of the reference-based image restoration method using the diffusion model. A hybrid architecture of the diffusion model denoising neural network is proposed consisting of three main blocks: the basic denoising module, the reference-based module, and the fusion module for the final result generation. Three models were trained: a diffusion model, a reference-based convolutional neural network, and a proposed hybrid model. All three models were trained and evaluated on the Large-Scale Multi-Reference Dataset dataset. Based on the results of the trained models testing, a qualitative (visual) and quantitative comparison of the three models was done. The hybrid model demonstrated higher image quality, clarity, and consistency compared to the convolutional neural network using references and better restoration of real details compared to the diffusion model. According to the quantitative evaluation, the hybrid model also showed higher results compared to pure models. The results of this work can be used to increase the resolution of any images using reference information.

Keywords

image processing, diffusion models, super-resolution, deep learning, image restoration

For citation: Denisov A.K., Bykovskii S.V., Kustarev P.V. Reference-based diffusion model for super-resolution. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2025, vol. 25, no. 2, pp. 321–327 (in Russian). doi: 10.17586/2226-1494-2025-25-2-321-327

Введение

В эпоху быстрого развития цифровых технологий постоянно растет спрос на высококачественные изображения в различных приложениях, от медицинской диагностики до сферы развлечений. Несовершенство процесса получения цифровых изображений, начиная от конструктивных особенностей оптических систем и специфики работы сенсоров до сложных условий съемки, связанных с погодными условиями, низкой освещенностью, движением объектов сцены и т. д., делает необходимыми алгоритмы восстановления и улучшения качества изображений. Одной из важных подзадач улучшения изображений является увеличение разрешения (Super-Resolution, SR). Выделяются методы, работающие с одним изображением (Single Image Super-Resolution, SISR), и использующие референсные изображения (Reference-Based Super-Resolution, RefSR). RefSR выступает в качестве перспективного решения, которое использует дополнительные референсные изображения высокого разрешения для повышения качества целевого изображения с низким разрешением. В отличие от традиционных SISR-методов, RefSR использует большое количество деталей, доступных в референсных изображениях для более реалистичной реконструкции. Глубокое обучение, в частности сверточные нейронные сети (CNN), произвело революцию в области обработки изображений, предоставив мощные инструменты для моделирования сложных отображений между входными и выходными изображениями. Одной из особенностей RefSR-методов на основе CNN является существенное различие в качестве деталей восстановленного изображения в областях, присутствующих и отсутствующих на референсных изображениях.

Одновременно с этим развиваются SISR-методы, использующие в основе диффузионные модели. Такие модели способны за счет итеративного процесса генерации выходного изображения восстанавливать большое количество реалистичных деталей. Проблемой этих методов является нежелательная генерация неверных деталей в тех случаях, когда заведомо известно, что должно быть изображено в данном месте (например, известного архитектурного сооружения). Иными словами, классические RefSR-методы страдают от наличия на результирующем изображении областей с низкой четкостью и качеством деталей, а мощные SISR-методы на основе диффузионных моделей генерируют ложные детали, которые хорошо заметны в некоторых случаях. В настоящей работе делается попытка исправить недостатки двух SR-методов и использовать их сильные стороны для решения задачи SR-изображений с использованием референсной информации.

Обзор научных публикаций

Методы SR-изображений за последние годы претерпели значительные изменения, особенно с внедрением методов глубокого обучения, которые преобразовали область обработки изображений. В работе [1] были заложены основы использования CNN в задачах SR

и продемонстрированы значительные улучшения по сравнению с традиционными методами интерполяции. Основываясь на этом успехе, авторы работы [2] предложили архитектуру VDSR, которая использовала очень глубокую сеть для увеличения качества с помощью обучения восстановлению остаточной информации (residual learning). В [3] была представлена улучшенная глубокая остаточная сеть (EDSR). В настоящей работе архитектура сети была оптимизирована для удаления ненужных модулей с целью повышения эффективности обучения и точности SR. Генеративно-состязательные сети (GAN) также оказали значительное влияние на SR с появлением SRGAN [4]. Целью работы [4] было создание визуально привлекательных результатов SR с помощью применения состязательного обучения для улучшения перцептивного качества изображений. В [5] значительное внимание уделено синтезированию тренировочных данных для имитации различных деградаций изображений для улучшения качества работы в реальных сценариях. Эти исследования иллюстрируют быстрый прогресс в методах SR-изображений, подчеркивая переход от классических технологий к более сложным моделям глубокого обучения, которые используют возможности CNN для повышения качества изображения.

Параллельно с методами SISR важной областью исследований стали методы RefSR, особенностью которого является использование дополнительных изображений, содержащих часть той же сцены в высоком разрешении. Метод SRNTT [6] использует многоуровневую структуру для переноса текстур между фрагментами входного и референсного изображений. Фреймворк C2-Matching, предложенный в [7], использует три стадии обучения, в первых двух из которых с помощью контрастивного обучения тренируется сеть извлечения признаков из целевого и референсного изображений, а на третьей стадии обучается сеть, восстанавливающая результирующее изображение на основе извлеченных и сопоставленных признаков. Еще одним заметным вкладом является метод DATSR [8], в котором применяется механизм адаптивного внимания (deformable attention) для динамического фокуса на соответствующих признаках в референсных изображениях, позволяя модели генерировать изображения с высоким разрешением, лучше справляясь с пространственными вариациями в данных. Важной проблемой RefSR-методов является дефицит данных для обучения. В работе [9] представлен LMR: большой многообразный набор данных для обучения и бенчмаркинга различных RefSR-моделей, кроме того, в работе был предложен метод, позволяющий использовать сразу несколько референсных изображений и улучшить таким образом качество выходного изображения. В [10] используется предобученная диффузионная модель для генерации референсного изображения, детали с которого затем переносятся на целевое изображение. Эти исследования подчеркивают значительный прогресс в методах RefSR и показывают перспективы их использования. При этом одной из существенных проблем остается восстановление областей изображения, отсутствующих на референсе и из-за этого страдающих от недостатка реалистичных текстур.

В то время как большинство методов обработки изображений использовали в своей основе CNN и трансформеры, в последнее время активно стали применяться диффузионные модели. Такие модели стали мощной основой в генеративном моделировании. Диффузионные модели направлены на усваивание структуры сложных распределений данных с помощью процесса, который итеративно преобразует зашумленные изображения в образцы данных из целевого распределения. Основополагающим вкладом в диффузионные модели является работа Denoising Diffusion Probabilistic Models (DDPM) [11], где представлен фреймворк, в котором процесс прямой диффузии постепенно добавляет шум к данным, в то время как обучаемый обратный процесс удаляет этот шум для генерации новых образцов данных. В работе DDPM подчеркивается важность зависимости интенсивности шума от шага диффузии и типов параметризации, устанавливаются базовые уровни качества, на которых основаны последующие исследования. Основываясь на DDPM, в [12] представлен метод Denoising Diffusion Implicit Models (DDIM), который направлен на повышение эффективности генерации путем предложения неявного процесса сэмплинга. Использование DDIM приводит к более быстрому процессу генерации при сохранении высокого качества. Работа DDIM подчеркивает универсальность подходов, основанных на диффузии, предлагая практические идеи относительно вычислительной эффективности.

Дальнейшие достижения видны в разработке моделей скрытой диффузии (Latent Diffusion Models, LDM) [13]. Ключевая идея LDM состоит в выполнении процессов диффузии в сжатом скрытом пространстве, например, вариационного автоэнкодера, а не напрямую в пространстве данных высокой размерности. Это значительно снижает вычислительную сложность при сохранении высокого качества генерации, демонстрируя применимость модели к таким задачам, как синтез изображений и видео высокого разрешения. Помимо использования в задачах генерации изображений, диффузионные модели нашли свое применение в различных image-to-image задачах, например восстановлении изображений. Используются два основных подхода. В первом — обучение диффузионной модели с нуля целевой задаче, при этом изображение для восстановления обычно подается внутрь денойзера диффузионной модели вместе с шумом. Так решаются image-to-image задачи в таких методах как SRDiff [14], LDM [13]. Другим подходом является использование больших предобученных генеративных моделей и дальнейшее обучение целевой задаче специальных адаптеров к ним. Примером такого метода является работа SUPIR [15], использующая Stable Diffusion XL в качестве базовой модели.

Предложенный метод

В данной работе предлагается гибридная диффузионная модель, использующая помимо целевого изображения низкого разрешения одно или несколько референсных изображений.

В качестве основы используется LDM, представляющая собой вероятностную модель, которая выучи-

вает распределение данных $p(x)$ путем постепенного удаления шума из сэмпла, полученного из нормального распределения, что соответствует выучиванию обратного процесса фиксированной марковской цепи длины T. Эти модели можно интерпретировать как равновзвешенную последовательность шумоподавляющих автоэнкодеров $\epsilon_\theta(x_t, t)$; $t = 1, \ldots, T$, которые обучены предсказывать чистую версию входа x_t, где x_t — зашумленная версия входа x. Для эффективного сжатия входных данных и возможности работать с более большими изображениями используется VQ-VAE, состоящий из энкодера $\mathcal{E}$ и декодера $\mathcal{D}$ с коэффициентом сжатия $f = 4$. Таким образом, диффузионная модель работает с латентной версией входа $z = \mathcal{E}(x)$, а для получения результирующего изображения в пространстве пикселов необходимо пропустить выход модели через декодер. При этом обучение происходит в латентном пространстве, минимизируется следующая функция потерь: $L_{LDM} := \mathbb{E}_{z,\epsilon,t}[\|\epsilon - \epsilon_\theta(z_t, t)\|_2^2]$, где $\epsilon \sim \mathcal{N}(0, 1)$ — гауссов шум, добавляемый на каждом шаге; θ — параметры модели; t — шаг времени, случайно выбираемый во время обучения; $\mathbb{E}[\cdot]$ — математическое ожидание.

Для переноса информации с референсных изображений используется механизм, предложенный в работе [9]. Для каждого референсного изображения отдельно проводится следующая процедура: из изображения низкого разрешения (LR) и референсного изображения извлекаются признаки на разных уровнях; между фрагментами этих карт признаков производится сопоставление; для каждого фрагмента LR выбирается наилучший фрагмент референсного изображения (в пространстве признаков); с помощью адаптивной свертки (deformable convolution) производится уточнение и улучшение агрегированного референсного изображения. После того как полностью выполнена процедура, применяется специальный модуль слияния, комбинирующий несколько агрегированных референсных изображений в одно, после чего производится восстановление LR с использованием полученного комбинированного референсного изображения. В настоящей работе используется претренированная модель данного модуля.

Архитектура предложенной модели представлена на рис. 1. Модель состоит из трех модулей: U-Net сети денойзера (Diffusion Unet), извлечения информации из референсных изображений (модуль обозначен как RefSR, референсные изображения — как Refs) и слияния (Fusion Module). На вход поступают LR и референсные изображения. В результате работы модели получается восстановленное изображение SR. На рисунке не показаны предобученные энкодер и декодер, используемые для работы в латентном пространстве.

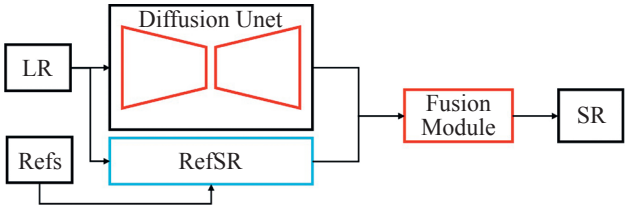

Рис. 1. Предложенная гибридная модель восстановления изображения.

Красным цветом выделены обучаемые модули, синим — претренированные

Fig. 1. Proposed hybrid restoration model architecture.

Blue denotes frozen modules, red denotes trained modules

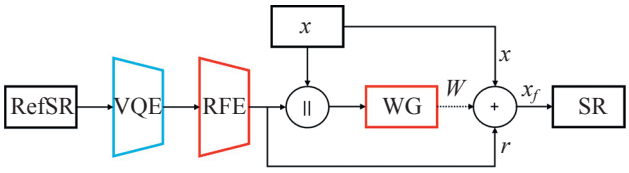

Рис. 2. Схема предложенного модуля слияния (Fusion Module).

VQE — VQ Encoder, RFE — Reference Feature Extractor, WG — Weights Generator. Красным цветом выделены обучаемые модули, синим — претренированные

Fig. 2. Proposed fusion module.

VQE — VQ Encoder, RFE — Reference Feature Extractor, WG — Weights Generator. Blue denotes frozen modules, red denotes trained modules

На рис. 2 представлена схема разработанного модуля слияния. Изображение с перенесенной референсной информацией проходит через VQ энкодер, затем через блок извлечения признаков (RFE), состоящий из пяти остаточных блоков, после чего конкатенируется с признаками из U-Net и идет в модуль генерации весовых коэффициентов (WG), состоящий из трех сверточных слоев. Финальный результат (SR) получается путем взвешенного суммирования по следующей формуле (обозначения соответствуют рис. 2): $x_f = Wx + (1 - W)r$, где x — исходные признаки из U-Net; x_f — измененные признаки U-Net; W — весовые коэффициенты; r — признаки, извлеченные из изображения с перенесенной референсной информацией.

Результаты

Для сравнения предлагаемого метода со стандартной диффузионной моделью (LDM) и сверточной сетью с использованием референсной информации (LMR) был использован набор данных из работы [9]. Он содержит 112 142 набора [целевое изображение, референсное изображение 1 … референсное изображение 5] для обучения, а также 142 набора с переменным (2–6) количеством референсных изображений для тестирования. В качестве модели деградации использовался 4x bicubic downsampling.

Так как используемый набор данных был предложен в той же работе, что метод LMR, была использована предобученная авторами модель. LDM и предлагаемая гибридная модель были обучены по следующей схеме: целевое изображение низкого разрешения подавалось в диффузионную модель, его каналы склеивались (конкатенировались) с каналами зашумленного целевого изображения высокого разрешения. При этом в модуль слияния гибридной модели одновременно подавалось изображение, восстановленное с помощью референсной информации. Обучение проводилось в течение 400 000 итераций с размером батча 4 на 8 NVIDIA V100 на изображениях размером 512 × 512 пикселов.

В качестве метрик для оценки качества классификации использовались следующие метрики: Learned Perceptual Image Patch Similarity (LPIPS) [16], Contrastive Language-Image Pre-training Image Quality Analysis (CLIP-IQA) [17], Fréchet Inception Distance (FID) [18]. Метрика LPIPS представляет собой расстояние между признаками, вычисленными от эталонного и оцениваемого изображений. Признаки вычисляются с помощью предобученной сети, например, AlexNet. Авторы данной метрики показывают значительную корреляцию с перцептивной оценкой качества изображения. Метрика CLIP-IQA построена на основе мультимодальной модели CLIP, связывающей изображения с их описаниями. Данная метрика основана на вычислении косинусного расстояния между описанием оцениваемого изображения и некоторым набором описаний заведомо «хорошего» и «плохого» изображений. Таким образом, для оценки качества изображения не требуется эталон. Метрика FID используется для оценки качества генерации и показывает расстояние между целевым распределением и распределением сгенерированных сэмплов после обработки их с помощью предобученной Inception v3 и вычисления средних и ковариаций последнего слоя.

В таблице приведены результаты обучения трех рассматриваемых методов: сверточной сети с использованием референсных изображений (LMR), обычной диффузионной модели (LDM) и предлагаемой гибрид-

Таблица. Результаты сравнения качества восстановленных изображений

Table. Quantitative evaluation of the proposed method

Метод	LPIPS-Alexnet ↓	CLIP-IQA ↑	FID ↓
LMR	0,1488	0,4678	15,225
LDM	0,1480	0,5857	11,400
LDM-Ref	0,1156	0,6063	10,169

Примечание. Стрелки у названия метрик указывают на уменьшение (↓) или увеличение (↑) величины метрики с увеличением качества изображения.

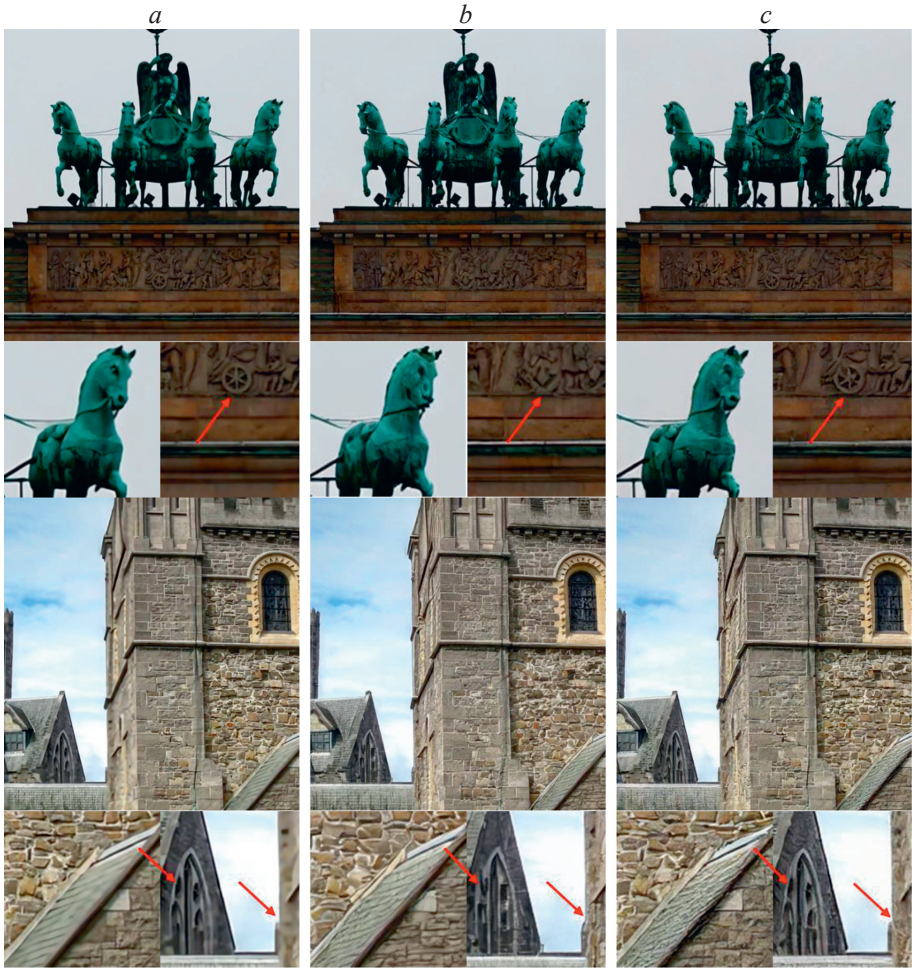

Рис. 3. Изображения, полученные в результате работы трех сравниваемых методов: LMR (*a*); LDM (*b*); гибридная модель (LDM-Ref) (*c*)

Fig. 3. Visual results of trained models: LMR (*a*); LDM (*b*); hybrid model (LDM-Ref) (*c*)

ной диффузионной модели с использованием референсных изображений (LDM-Ref).

Из таблицы видно, что предлагаемый метод улучшает все рассматриваемые метрики, используя сильные стороны каждой из составных частей гибридной модели. По сравнению с LDM и LMR улучшается попиксельное сходство текстур выходного изображения и эталонного (в LDM эта метрика страдает от ненастоящих деталей, в LMR — от плохо восстановленных областей, где отсутствует референс), что показывает уменьшение LPIPS более чем на 20 %. CLIP-IQA и FID особенно улучшаются по сравнению с LMR (увеличение почти на 30 % CLIP-IQA и уменьшение FID на 33 %) благодаря использованию диффузионной модели и генерации более четких текстур.

На рис. 3 представлены результаты работы трех сравниваемых методов. Использованы изображения из набора данных LMR [9]. Красными стрелками показаны области интереса. Видно, что LDM генерирует ненастоящие детали — в области колеса, стрельчатой арки, в то время как на результате работы LMR присутствуют слаботекстурированные области. Предлагаемый метод стремится исправить оба этих недостатка, балансируя между количеством деталей и их реалистичностью, увеличивая четкость и резкость по сравнению с LMR, и однородность, и согласованность деталей и текстур по сравнению с LDM.

Заключение

В работе были рассмотрены достоинства и недостатки двух существующих классов методов увеличения разрешения изображений: сверточных нейронных сетей с использованием референсных изображений и диффузионных моделей. В то время как первые страдают от низкого качества областей, где отсутствует референсная информация, вторые генерируют ложные детали на изображении.

Была предложена гибридная модель, построенная на базе LDM, включающая в себя специальный модуль слияния, добавляющий референсную информацию. По результатам визуальной оценки, а также количественного сравнения предложенного метода с чистой диффузионной моделью и чистой сверточной сетью с использованием референсных изображений, было показано преимущество в перцептивном качестве предложенного метода, а также уменьшение метрики LPIPS на минимум 20 % и значительное улучшение по CLIP-IQA и FID по сравнению с обоими рассматриваемыми методами.

Предложенный метод может применяться для увеличения разрешения любых изображений с использованием референсной информации.

Литература

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. N 2. P. 295–307. https://doi.org/10.1109/TPAMI.2015.2439281
2. Kim J., Lee J.K., Lee K.M. Accurate image Super-Resolution using very deep convolutional networks // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 1646–1654. https://doi.org/10.1109/CVPR.2016.182
3. Lim B., Son S., Kim H., Nah S., Lee K.M. Enhanced deep residual networks for single image Super-Resolution // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). 2017. P. 1132–1140. https://doi.org/10.1109/CVPRW.2017.151
4. Ledig C., Theis L., Huszár F., Caballero J., Cunningham A., Acosta A., Aitken A., Tejani A., Totz J., Wang Z., Shi W. Photo-realistic single image Super-Resolution using a generative adversarial network // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. P. 105–114. https://doi.org/10.1109/10.1109/CVPR.2017.19
5. Wang X., Xie L., Dong C., Shan Y. Real-ESRGAN: training real-world blind Super-Resolution with pure synthetic data // Proc. of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). 2021. P. 1905–1914. https://doi.org/10.1109/ICCVW54120.2021.00217
6. Zhang Z., Wang Z., Lin Z., Qi H. Image Super-Resolution by neural texture transfer // Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019. P. 7974–7983. https://doi.org/10.1109/CVPR.2019.00817
7. Jiang Y., Chan K.C.K., Wang X., Loy C.C., Liu Z. Robust Reference-based Super-Resolution via C2-Matching // Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2021. P. 2103–2112. https://doi.org/10.1109/CVPR46437.2021.00214
8. Cao J., Liang J., Zhang K., Li Y., Zhang Y., Wang W., Van Gool L. Reference-based image Super-Resolution with deformable attention transformer // Lecture Notes in Computer Science. 2022. V. 13678. P. 325–342. https://doi.org/10.1007/978-3-031-19797-0_19

References

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, vol. 38, no. 2, pp. 295–307. https://doi.org/10.1109/TPAMI.2015.2439281
2. Kim J., Lee J.K., Lee K.M. Accurate image Super-Resolution using very deep convolutional networks. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 1646–1654. https://doi.org/10.1109/CVPR.2016.182
3. Lim B., Son S., Kim H., Nah S., Lee K.M. Enhanced deep residual networks for single image Super-Resolution. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017, pp. 1132–1140. https://doi.org/10.1109/CVPRW.2017.151
4. Ledig C., Theis L., Huszár F., Caballero J., Cunningham A., Acosta A., Aitken A., Tejani A., Totz J., Wang Z., Shi W. Photo-realistic single image Super-Resolution using a generative adversarial network. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 105–114. https://doi.org/10.1109/10.1109/CVPR.2017.19
5. Wang X., Xie L., Dong C., Shan Y. Real-ESRGAN: training real-world blind Super-Resolution with pure synthetic data. *Proc. of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, 2021, pp. 1905–1914. https://doi.org/10.1109/ICCVW54120.2021.00217
6. Zhang Z., Wang Z., Lin Z., Qi H. Image Super-Resolution by neural texture transfer. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 7974–7983. https://doi.org/10.1109/CVPR.2019.00817
7. Jiang Y., Chan K.C.K., Wang X., Loy C.C., Liu Z. Robust Reference-based Super-Resolution via C2-Matching. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 2103–2112. https://doi.org/10.1109/CVPR46437.2021.00214
8. Cao J., Liang J., Zhang K., Li Y., Zhang Y., Wang W., Van Gool L. Reference-based image Super-Resolution with deformable attention transformer. *Lecture Notes in Computer Science*, 2022, vol. 13678, pp. 325–342. https://doi.org/10.1007/978-3-031-19797-0_19

9. Zhang L., Li X., He D., Li F., Ding E., Zhang Z. LMR: a large-scale multi-reference dataset for Reference-based Super-Resolution // Proc. of the IEEE/CVF International Conference on Computer Vision (ICCV). 2023. P. 13072–13081. https://doi.org/10.1109/ICCV51070.2023.01206
10. Li G., Xing W., Zhao L., Lan Z., Sun J., Zhang Z., Zhang Q., Lin H., Lin Z. Self-Reference image Super-Resolution via pre-trained diffusion large model and window adjustable transformer // Proc. of the 31st ACM International Conference on Multimedia. 2023. P. 7981–7992. https://doi.org/10.1145/3581783.3611866
11. Ho J., Jain A., Abbeel P. Denoising diffusion probabilistic models // arXiv. 2020. arXiv:2006.11239. https://doi.org/10.48550/arXiv.2006.11239
12. Song J., Meng C., Ermon S. Denoising diffusion implicit models // arXiv. 2020. arXiv:2010.02502. https://doi.org/10.48550/arXiv.2010.02502
13. Rombach R., Blattmann A., Lorenz D., Esser P., Ommer B. High-Resolution image synthesis with latent diffusion models // Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022. P. 10674–10685. https://doi.org/10.1109/CVPR52688.2022.01042
14. Li H., Yang Y., Chang M., Chen S., Feng H., Xu Z., Li Q., Chen Y. SRDiff: Single Image Super-Resolution with diffusion probabilistic models // Neurocomputing. 2022. V. 479. P. 47–59. https://doi.org/10.1016/j.neucom.2022.01.029
15. Yu F., Gu J., Li Z., Liu 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 // Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2024. P. 25669–25680. https://doi.org/10.1109/CVPR52733.2024.02425
16. Zhang R., Isola P., Efros A.A., Shechtman E., Wang O. The unreasonable effectiveness of deep features as a perceptual metric // Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2018. P. 586–595. https://doi.org/10.1109/CVPR.2018.00068
17. Wang J., Chan K.C.K., Loy C.C. Exploring CLIP for assessing the look and feel of images // Proc. of the 37th AAAI Conference on Artificial Intelligence. 2023. V. 37. N 2. P. 2555–2563. https://doi.org/10.1609/aaai.v37i2.25353
18. 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 // Proc. of the 31st International Conference on Neural Information Processing Systems (NIPS '17). 2017. P. 6629–6640.

9. Zhang L., Li X., He D., Li F., Ding E., Zhang Z. LMR: a large-scale multi-reference dataset for Reference-based Super-Resolution. *Proc. of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023, pp. 13072–13081. https://doi.org/10.1109/ICCV51070.2023.01206
10. Li G., Xing W., Zhao L., Lan Z., Sun J., Zhang Z., Zhang Q., Lin H., Lin Z. Self-Reference image Super-Resolution via pre-trained diffusion large model and window adjustable transformer. *Proc. of the 31st ACM International Conference on Multimedia*, 2023, pp. 7981–7992. https://doi.org/10.1145/3581783.3611866
11. Ho J., Jain A., Abbeel P. Denoising diffusion probabilistic models. *arXiv*, 2020, arXiv:2006.11239. https://doi.org/10.48550/arXiv.2006.11239
12. Song J., Meng C., Ermon S. Denoising diffusion implicit models. *arXiv*, 2020, arXiv:2010.02502. https://doi.org/10.48550/arXiv.2010.02502
13. Rombach R., Blattmann A., Lorenz D., Esser P., Ommer B. High-Resolution image synthesis with latent diffusion models. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pp. 10674–10685. https://doi.org/10.1109/CVPR52688.2022.01042
14. Li H., Yang Y., Chang M., Chen S., Feng H., Xu Z., Li Q., Chen Y. SRDiff: Single Image Super-Resolution with diffusion probabilistic models. *Neurocomputing*, 2022, vol. 479, pp. 47–59. https://doi.org/10.1016/j.neucom.2022.01.029
15. Yu F., Gu J., Li Z., Liu 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. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024, pp. 25669–25680. https://doi.org/10.1109/CVPR52733.2024.02425
16. Zhang R., Isola P., Efros A.A., Shechtman E., Wang O. The unreasonable effectiveness of deep features as a perceptual metric. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018, pp. 586–595. https://doi.org/10.1109/CVPR.2018.00068
17. Wang J., Chan K.C.K., Loy C.C. Exploring CLIP for assessing the look and feel of images. *Proc. of the 37th AAAI Conference on Artificial Intelligence*, 2023, vol. 37, no. 2. pp. 2555–2563. https://doi.org/10.1609/aaai.v37i2.25353
18. 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. *Proc. of the 31st International Conference on Neural Information Processing Systems (NIPS '17)*, 2017, pp. 6629–6640.

Авторы

Денисов Алексей Константинович — ассистент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, SC 57210698353, https://orcid.org/0000-0001-8135-1723, denisov@itmo.ru

Быковский Сергей Вячеславович — кандидат технических наук, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, SC 57216469537, https://orcid.org/0000-0003-4163-9743, sergei_bykovskii@itmo.ru

Кустарев Павел Валерьевич — кандидат технических наук, декан, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, SC 35317916600, https://orcid.org/0000-0001-9326-0837, kustarev@itmo.ru

Authors

Aleksei K. Denisov — Assistant, ITMO University, Saint Petersburg, 197101, Russian Federation, SC 57210698353, https://orcid.org/0000-0001-8135-1723, denisov@itmo.ru

Sergei V. Bykovskii — PhD, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, SC 57216469537, https://orcid.org/0000-0003-4163-9743, sergei_bykovskii@itmo.ru

Pavel V. Kustarev — PhD, Dean, ITMO University, Saint Petersburg, 197101, Russian Federation, SC 35317916600, https://orcid.org/0000-0001-9326-0837, kustarev@itmo.ru

Статья поступила в редакцию 26.11.2024
Одобрена после рецензирования 05.02.2025
Принята к печати 19.03.2025

Received 26.11.2024
Approved after reviewing 05.02.2025
Accepted 19.03.2025