

УДК 004.89
Код ГРНТИ 28.23.15

doi 10.54708/19926502_2026_30311348

Интегрированная система восстановления и аугментации изображений с использованием глубоких нейронных сетей трансформерного типа

Н.И. Бережнов*, А.А. Сирота

ФГБОУ ВО «Воронежский государственный университет», г. Воронеж, Россия

Аннотация. Представлена система для автоматического восстановления и аугментации изображений сцен, искаженных различными помехами, атмосферными осадками, на основе моделей глубоких нейронных сетей – трансформеров. В отличие от традиционных подходов, в которых задачи восстановления изображений и генерации и синтеза искусственных помех решаются отдельно, предлагается интегрированная система, рассматривающая их как взаимно обратимые процессы. Архитектура комплекса включает в себя модуль аугментации, синтезирующий погодные и другие искажения на изображениях, и модуль восстановления, устраняющий эти искажения. Оба модуля построены на моделях-трансформерах с модифицированным механизмом внимания. Обучение системы осуществляется по циклической схеме: в модуле аугментаций формируются искаженные изображения, которые затем улучшаются в модуле восстановления с обратным преобразованием. При этом оптимизируются функции потерь и чувствительные ошибки, обеспечивая реалистичность как сгенерированных, так и восстановленных изображений. Приведены результаты экспериментов, показывающие, что предложенный подход превосходит многие современные модели по метрикам качества (PSNR, SSIM, FID и др.).

Ключевые слова: нейронные сети, трансформеры, механизм внимания, аугментация данных, восстановление изображений, аугментация изображений.

*beregnochnikita@yandex.ru

Введение

Современные системы компьютерного зрения предъявляют высокие требования к качеству визуальной информации, но в реальных условиях входные изображения часто искажены из-за шумов, атмосферных осадков, ограничений сенсоров и помех передачи данных. Задачи восстановления и улучшения изображений, а также эффективной аугментации данных, становятся критически важными на фоне растущей сложности прикладных задач и разнообразия условий съемки. Методы глубокого обучения, в частности архитектуры на основе моделей-трансформеров, продемонстрировали высокую устойчивость к сложным искажениям изображений. Этого удалось достичь главным образом из-за механизма внимания (attention), который позволяет эффективно учитывать как локальные, так и глобальные взаимосвязи между элементами изображения, что помогает устранять артефакты, шумы, тени и другие искажения [1].

Однако исследователи при применении трансформеров для восстановления изображений сталкиваются с рядом ограничений. Во-первых, механизм внимания может избыточно фокусироваться на отдельных деталях изображения, снижая эффективность удаления шума. Во-вторых, глубокие модели склонны к переобучению при ограниченном объеме данных. В-третьих, многие трансформерные модели громоздки и требуют больших вычислительных ресурсов, особенно при обучении. В-четвертых, сбор достаточно разнообразного набора обучающих изображений затруднен, что усложняет формирование качественного датасета. Поэтому проблема нехватки данных остается актуальной, требуются как классические приемы аугментации, так и генеративные модели для синтеза новых условно-реалистичных изображений. Генерация синтетических данных (например, с имитацией атмосферных осадков, цифровых дефектов, нестандартных шумов) расширяет

возможности моделирования входных данных и улучшает обобщающую способность нейросетевых моделей [2].

В связи с этим целью настоящей работы является разработка программного комплекса восстановления и аугментации (ПК ВИА) изображений на базе нейронных сетей трансформерного типа, способных моделировать и устранять широкий класс искажений изображений. В комплексе реализованы: трансформер для восстановления изображений с модифицированным механизмом внимания на основе внесения стохастической регуляризации для повышения устойчивости моделей к переобучению, а также трансформерный модуль с перекрестным вниманием для адаптивной генерации погодных искажений.

Обзор существующих методов

В современной литературе прослеживается тенденция к созданию систем, где процессы генерации искажений и их устранения рассматриваются во взаимосвязи. В работе Real-ESRGAN [3] задача «слепого» супер-разрешения решается с использованием генеративно-состязательной архитектуры, где генератор восстанавливает изображение высокого качества, а дискриминатор, построенный по архитектуре U-Net и снабженный спектральной нормализацией, оценивает достоверность результата по сравнению с эталонными изображениями. Однако в данном подходе генератор и дискриминатор взаимодействуют только в прямом направлении (от восстановления к оценке качества), что ограничивает потенциал совместного обучения обеих задач.

Метод DASR (Degradation-Aware SR) [4] предполагает кодирование искажений в скрытое представление, обучаемое без явной оценки параметров деградации. Используя обучение на контрасте (contrastive learning), энкодер DASR учится различать пары изображений с одинаковыми искажениями от пар с различными, формируя векторное пространство, отражающее признаки шумов и искажений. Затем это представление используется, чтобы адаптировать нейронную сеть для супер-разрешения под конкретный класс искажений. Такой подход демонстрирует системное понимание связи «деградация – восстановление», но не включает в себя интегрированную генерацию искаженного входа в процессе обучения: модель не генерирует зашумленные примеры как часть обучения.

На сегодняшний день универсальные архитектуры восстановления изображений и видео развиваются в направлении учета множества видов искажений. Так, DPMambaIR [5] использует алгоритм Degradation-Aware Prompt State Space Model для кодирования признаков деградации и блок High-Frequency Enhancement для компенсации потерь деталей, что обеспечивает высокое качество при меньших вычислительных затратах, но остается зависимой от предобученного энкодера и не включает генерацию искажений в контуре обучения. В свою очередь, UniWRV [6] ориентирована на видеоданные с гибридными атмосферными эффектами, сочетая модуль априорной информации о погодных осадках (WPGM) для пространственной адаптации и динамическую маршрутизацию (DRA) для объединения признаков; модель успешно справляется с дождем, снегом и туманом, однако полагается на синтетический датасет и не предусматривает генерацию самих погодных осадков.

Также стоит отметить модель TransWeather [7] – трансформер с одним энкодером и декодером, способными одновременно удалять дождь, туман и снег из изображений. В энкодере используются intra-patch transformer блоки, которые улучшают внимание внутри патчей и более эффективно обрабатывают мелкие искажения, а в декодере вводятся обучаемые weather type embeddings, направляющие сеть на конкретный тип погодного искажения. Модель показывает значительный прирост по сравнению с методами, ориентированными на один тип искажения, и работает быстрее и экономичнее, чем предыдущие All-in-One [8] подходы с множеством энкодеров.

Отдельное направление связано с двунаправленной обработкой погодных эффектов, где изучаются как синтез осадков, так и их удаление. При этом генерация в большинстве случаев выступает лишь в роли внешнего источника данных и не взаимодействует с моделью

восстановления через механизмы адаптивной обратной связи. Рассматривая обучение без учителя и GAN подход, систему, подобную ПК ВИА, можно реализовать через пару генераторов и дискриминаторов, но это делает архитектуру тяжеловесной. В предлагаемом подходе в работе достаточно использовать две нейронные сети, не обязательно одного типа: одну для восстановления, другую для генерации.

Механизм внимания в моделях трансформерах

Механизм внимания (attention) лежит в основе трансформеров и отвечает за адаптивное выделение важных фрагментов входных данных. Важным его подвидом является самовнимание, когда для расчета карт внимания используется только один входной тензор изображения.

В классическом блоке самовнимания входное изображение представляется последовательностью признаков (embeddings), которые через обучаемые линейные преобразования отображаются в три тензора: запросы Q , ключи K и значения V . Затем вычисляется матрица внимания путем произведения тензоров Q, K , нормирования Softmax-функцией. После этого результирующая матрица внимания используется как набор весов для линейной комбинации строк тензора значений V , что позволяет нейронной сети обращать внимание на важные части изображения в соответствии со степенью сходства между запросами и ключами. Формально результатом самовнимания является:

$$\text{Attention}(Q, K, V) = \text{Softmax}((QK^T) / \sqrt{d})V,$$

где d – размерность скрытого пространства. Таким образом, выход в каждой позиции получается как взвешенная комбинация всех позиций входа, что и позволяет моделям-трансформерам улавливать как локальные, так и глобальные зависимости между элементами изображения. В отличие от сверток, имеющих локальное рецептивное поле, самовнимание учитывает контекст всего изображения целиком. Это дает трансформерам преимущество в учете глобальных структур (например, границ объектов, распределения освещенности и т.д.), позволяя достигать высоких результатов в задачах обработки изображений [1]. Именно такая форма внимания (самовнимание) в данной работе реализуется в архитектуре трансформера для восстановления изображений, архитектура которого детально описана в работе авторов [9].

В задаче синтеза искажений и аугментации данных в данной работе применяется другая модель WeatherTransformer, представленная в работе авторов [10] – трансформер с перекрестным вниманием. Принципы, описанные выше, сохраняются; отличие заключается в том, что тензор K формируется не из исходных изображений, а из выбранных шаблонов, что позволяет целенаправленно моделировать искажения, сохраняя структуру сцены. Главная идея реализованного в этой модели подхода состоит в переносе погодных искажений из одного изображения-шаблона на другое с включением в него атмосферных осадков без наличия существенных артефактов. Такой подход позволяет быстро и надежно расширять обучающие данные, содержание изображения одной и той же сцены в различных погодных условиях.

При переходе к задачам восстановления изображений трансформеры уже зарекомендовали себя как один из наиболее перспективных классов моделей. Однако их прямое применение к изображениям высокого разрешения сопровождается значительными вычислительными затратами. Использование механизма самовнимания по всем пикселям изображения приводит к квадратичному росту сложности относительно их числа, что ограничивает практическое применение классических архитектур трансформеров.

В связи с этим в известных работах были предложены различные модификации механизма внимания, направленные на снижение вычислительной сложности при сохранении качества восстановления. Один из подходов – локальное внимание: изображение разбивается на непересекающиеся блоки и внимание вычисляется внутри них. Такой принцип реализован в архитектуре Swin Transformer и нашел отражение в модели для восстановления SwinIR [11]. Последняя состоит из трех основных компонентов: блока поверхностного извлечения

признаков, глубокой части с блоками Swin Transformer и остаточными связями, а также модуля реконструкции изображения. Помимо этого, за счет механизма перекрытия окон SwinIR демонстрирует высокие результаты в задачах сверхразрешения, подавления шума и устранения артефактов сжатия, при этом сохраняя небольшую вычислительную сложность и относительно небольшое число параметров.

Другим направлением стало использование методов сжатия канального пространства, когда матрица внимания аппроксимируется более компактным представлением. Так, например, для модели Restormer можно применить механизм Channel Bottleneck Self-Attention (CBSA) [12], при котором внутренняя размерность признаков перед вычислением весов внимания уменьшается в L раз. Эксперименты показали, что при сжатии каналов в 4–8 раз качество восстановления снижается незначительно, оставаясь сопоставимым с лучшими современными методами, тогда как число параметров нейронной сети заметно сокращается.

Дополнительный вклад в устойчивость моделей вносят методы регуляризации внимания. Наиболее распространенным является зануление случайной части значений тензора QK^T (dropout attention), что препятствует переобучению и способствует увеличению обобщающей способности нейронной сети, что было теоретически и экспериментально показано в работе авторов [9]. Совмещение регуляризации с модификациями внимания позволяет строить устойчивые и вычислительно эффективные трансформерные модели, применимые для широкого спектра задач восстановления изображений.

Таким образом, SwinIR и Restormer демонстрируют не только высокую точность восстановления при сложных искажениях, но и эффективные решения проблемы вычислительной сложности, характерной для классических трансформеров. В настоящей работе эти архитектуры с модификацией механизма внимания служат основой для реализации программного комплекса, объединяющего аугментацию и восстановление изображений.

Описание модулей ПК ВИА

Структурная схема предлагаемого программного комплекса ПК ВИА показана на Рис. 1.

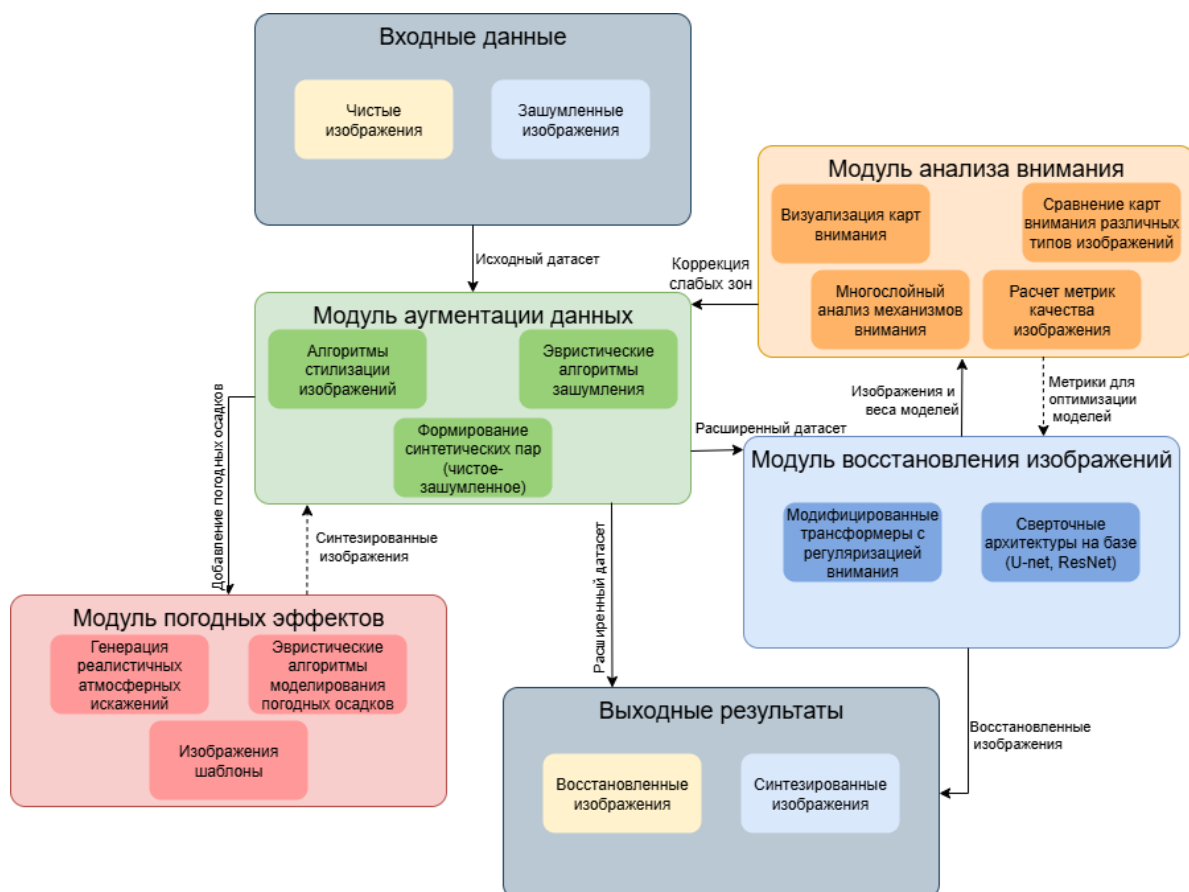


Рисунок 1. Структурная схема предлагаемого программного комплекса ПК ВИА.

Как общий результат выполненных исследований и разработок авторами предложен программный комплекс для восстановления и аугментации изображений. ПК ВИА представляет собой платформу для обработки изображений, объединяющую два функционально связанных направления обработки информации: восстановление изображений и генерацию шумов, искажений и эффектов погодных осадков для аугментации данных. Ключевая особенность ПК заключается в том, что эти задачи являются взаимнообратимыми процессами.

В рамках предложенной системы циклического обучения обосновано использование модульной архитектуры, обеспечивающей гибкость и адаптивность при работе с изображениями. Каждый модуль играет ключевую роль в соответствующих этапах цикла.

Модуль восстановления изображений включает как базовые сверточные архитектуры, так и модифицированные трансформеры с регуляризацией механизмов внимания, формируя основу первичного и повторного восстановления на стадиях обучения.

Модуль аугментации данных позволяет контролируемо расширять обучающую выборку за счет генерации различных искажений (включая зашумление, стилизованный перенос и синтез изображений), что критически важно при формировании синтетических пар изображений (чистого и зашумленного).

Модуль погодных эффектов реализует генерацию реалистичных атмосферных искажений с адаптивной интенсивностью, способствуя улучшению генеративной модели шума в обратной задаче.

Модуль анализа внимания используется на этапе проверки и отладки: визуализация карт внимания, многослойный анализ и сравнение различных типов внимания позволяют оценить интерпретируемость модели и уточнить ее слабые зоны для последующего дообучения.

Методика развертывания и применения программного комплекса (ПК ВИА) охватывает несколько этапов, соответствующих ключевым задачам в жизненном цикле обучения и применения нейросетевых моделей восстановления изображений.

Этап 1. Подготовка обучающей выборки. На первом этапе осуществляется формирование обучающего набора изображений, включающего как реальные, так и синтетически искаженные данные. Для генерации последних используется модуль аугментации данных, реализующий как стандартные эвристические алгоритмы искажения (зашумление, размытие, обесцвечивание), так и продвинутые методы переноса стиля и синтеза атмосферных условий. Особое внимание уделяется моделированию погодных явлений (дождь, снег, туман), реализованному через модуль погодных эффектов и WeatherTransformer, в котором интенсивность и характер искажений адаптируются к контексту исходного изображения. Формируются пары изображений вида «искаженное – исходное», используемые для обучения моделей восстановления.

Этап 2. Обучение и настройка моделей восстановления. На данном этапе задействуется модуль восстановления изображений, включающий глубокие сверточные и трансформерные архитектуры (класса SwinIR, Restormer и др.), модифицированные с целью повышения устойчивости к шумам и улучшения обобщающей способности. В процессе итеративного обучения используются циклы генерации и восстановления: изображения из обучающей выборки проходят через алгоритмы аугментации и затем подаются в нейросетевую модель восстановления, которая учится восстанавливать исходный их вид. Результаты восстанавливаемых изображений сравниваются с эталонами с помощью стандартных метрик (PSNR, SSIM) и введенных автором визуально-интерпретируемых показателей.

Этап 3. Проверка и интерпретация моделей. На этом этапе подключается модуль анализа внимания, предоставляющий визуализацию карт внимания и их динамику развития в процессе обучения. Это позволяет проводить исследование нейросетевой модели: выявлять области, на которые она фокусируется при восстановлении, и оценивать интерпретируемость механизма внимания. Использование этого модуля особенно важно при адаптации моделей под новые типы искажений, а также при выявлении «слепых зон» – участков изображений, где качество восстановления систематически оказывается неудовлетворительным.

Этап 4. Применение моделей и синтез новых искажений. После завершения обучения, модели восстановления могут быть использованы не только на реальных искаженных изображениях, но и в обратном режиме для генерации пар восстановленное (чистое) изображение и искомое зашумленное с целью подачи их на вход моделей синтеза и генерации. Такая взаимно-обратная архитектура реализует основной принцип ПК ВИА: если модель способна эффективно устранять определенный класс искажений, то она может быть использована для их воспроизведения с контролируемой интенсивностью и реалистичностью. Это позволяет динамически расширять обучающую выборку без необходимости ручной разметки данных.

Этап 5. Циклическое обучение и адаптация. ПК ВИА предполагает использование циклического обучения, при котором результаты анализа внимания и эффективности восстановления на тестовых данных используются для дообучения или адаптации моделей синтеза и генерации. Например, при обнаружении ухудшения качества восстановления в специфических погодных условиях может быть инициирована повторная генерация обучающих данных с усилением соответствующего погодного эффекта. Таким образом, обеспечивается гибкость и адаптивность модели к изменениям условий эксплуатации.

Результаты

Экспериментальная оценка разработанного комплекса проводилась на датасете All-Weather, предложенного в работе All-in-One Network [8], включающего в себя изображения с дождем (Raindrop), снегом (Snow100K) и туманом (Outdoor-Rain) с соответствующими эталонами; всего примерно 20 тысяч обучающих изображений и тысяча тестовых.

Для количественной оценки использовались распространенные метрики: пиковое отношение сигнал/шум (PSNR, dB) – характеризует точность восстановления по отношению к эталону (выше – лучше), индекс структурного подобия (SSIM) – измеряет сходство структур и контрастов (от 0 до 1, ближе к 1 с лучше), а также метрика восприятия FID (Frechet Inception Distance) – оценивает расстояние между распределениями признаков реальных и восстановленных изображений (ниже – лучше). PSNR и SSIM были рассчитаны для каждой тестовой пары, и затем усреднены; FID рассчитан сразу для выборки из 1000 восстановленных изображений и соответствующих им эталонов.

Было проведено сравнение работы системы с рядом современных подходов. Для задач восстановления от погодных эффектов использовалась модель TransWeather [7]. Также рассматривалась модель All-in-One [8], представляющая сверточную нейронную сеть для комплексного удаления погодных помех. Наконец, чтобы оценить пользу интегрированного подхода, использовались алгоритмы без обучения генератора, то есть на основе фиксированных искусственных аугментаций (Albumentations [13]) и ограниченного набора реальных данных. Такой вариант отражает традиционную схему обучения (когда для генерации тренировочных искажений используются эвристические алгоритмы и существующие данные). Результаты сравнения приведены в Табл. 1.

Таблица 1. Сравнение различных моделей при восстановлении изображений в условиях погодных искажений.

Метод	PSNR (↑)	SSIM (↑)	FID (↓)
All-in-One	24,71	0,898	–
TransWeather	27,96	0,95	–
SwinIR	25,0	0,85	76,2
SwinIR с Albumentations	26,18	0,86	75
Restormer	26,6	0,93	74,42
Restormer в системе ПК ВИА	29,1	0,94	72,42
SwinIR в системе ПК ВИА	30,0	0,94	71,60

Прочерки означают, что в оригинальных статьях данные метрики не замерялись. Для All-in-One и TransWeather метрики взяты для датасета с туманом и дождем. Наилучший результат показали модели, обученные в парадигме циклического обучения в интегрированной системе ПК ВИА. На Рис. 2 показан пример исходного и восстановленного изображений.



Рисунок 2. Пример зашумленного и восстановленного изображений.

Для сравнения эффективности синтеза погодных осадков была выбрана модель CycleGAN [14], основанная на принципе обучения выполнению преобразований изображений из одного домена в другой и обратно, при этом минимизируя расхождения с исходными данными. Такой подход позволяет выполнять перенос стиля без использования парных данных, что делает его удобным для задач синтеза атмосферных осадков. Однако в условиях сложных погодных искажений модель часто генерирует визуальные артефакты, разрывы текстур и нарушения геометрической структуры сцены, что ограничивает ее применимость для высококачественного синтеза изображений.

Также был выбран более современный подход HRIGNet [15], основанный на диффузионной модели с учетом высокочастотных деталей, который снижает показатель FID и улучшает структурные метрики, но по PSNR уступает классическим GAN-подходам. Для ПК ВИА был использован подход с перекрестным вниманием на основе модели WeatherTransformer, предложенный авторами в статье [10], дающий наилучший результат из-за реалистичности переносимых искажений и сохранения структуры сцены. Результаты показаны в Табл. 2.

Таблица 2. Сравнение методов при синтезе изображений в плохих погодных условиях.

Метод	PSNR (↑)	SSIM (↑)	FID (↓)
CycleGAN	21,604	0,639	47,073
HRIGNet	18,595	0,747	32,111
WeatherTransformer	21,88	0,78	30,47
WeatherTransformer в системе ПК ВИА	22,30	0,79	29,80

Из результатов видно, что при синтезе искажений модель WeatherTransformer в интегрированной системе ПК ВИА также улучшила свои метрики. На Рис. 3 представлены примеры синтезированных изображений WeatherTransformer в системе ПК ВИА.



Рисунок 3. Примеры обработки изображений:

(а) исходные изображения, (б) изображения-шаблоны, (в) синтезированные изображения.

Заключение

В работе предложена и исследована интегрированная система и реализующий ее программный комплекс, объединяющие алгоритмы восстановления и аугментации изображений на основе нейронных сетей трансформерного типа.

В рамках программного комплекса реализована циклическая схема обучения, в которой модули восстановления и генерации итеративно улучшают друг друга на основе синтетических и реальных данных. Экспериментально подтверждена эффективность такого подхода: использование синтетических и искаженных изображений в качестве обучающих данных улучшило результаты в задачах восстановления изображений.

Литература:

1. Khan S., Naseer M., Hayat M., et al. Transformers in vision: A survey // ACM Computing Surveys. 2022. Vol. 54(10). P. 1–41. DOI: 10.1145/3505244.
2. Бережнов Н.И., Сирота А.А. Исследование обобщающей способности методов глубокого обучения для улучшения качества изображений. В сб.: XXIII Международная конференция «Информатика: проблемы, методология, технологии». Воронеж: ИПЦ ВГУ, 2023. С. 510–518. [Berezhnov N.I., Sirota A.A. Investigation of the generalization ability of deep learning methods for image quality improvement. In: Proceedings of the XXIII International Conference “Informatics: Problems, Methodology, Technologies”. Voronezh: Publishing Center of Voronezh State University, 2023. P. 510–518 (in Russian)].
3. Xintao Wang, Liangbin Xie, Chao Dong, Ying Shan. Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). 2021. IEEE, 2021. Art. 00217. DOI: 10.1109/ICCVW54120.2021.00217.
4. Longguang Wang, Yingqian Wang, Xiaoyu Dong, et al. Unsupervised degradation representation learning for blind super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021. IEEE, 2021. P. 10581–10590. DOI: 10.1109/CVPR46437.2021.01052.
5. Liu Z., Zhou S., Dai Y., et al. DPMambaIR: All-in-One image restoration via degradation-aware prompt state space model [arXiv preprint]. arXiv:2504.17732. arXiv, 2025. DOI: 10.48550/arXiv.2504.17732.
6. Wan Y., Shao M., Cheng Y., et al. Removing multiple hybrid adverse weather in video via a unified model [arXiv preprint]. arXiv:2503.06200. arXiv, 2025. DOI: 10.48550/arXiv.2503.06200.
7. Valanarasu J.M.J., Yasarla R., Patel V.M. TransWeather: Transformer-based restoration of images degraded by adverse weather conditions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022. IEEE, 2022. P. 2353–2363. DOI: 10.1109/CVPR52688.2022.00239.

8. Li Q., Tan R. T., Cheong L.-F. All in One Bad Weather Removal Using Architectural Search // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2020. P. 3175–3185. DOI: 10.1109/CVPR42600.2020.00325.
9. Бережнов Н.И., Сирота А.А. Регуляризация механизма самовнимания в блоках трансформеров и ее применение в задачах классификации и восстановления изображений // Искусственный интеллект и принятие решений. 2025. № 2. С. 114–129. [Berezhnov N.I., Sirota A.A. Regularization of the self-attention mechanism in transformer blocks and its application to image classification and restoration tasks // Artificial Intelligence and Decision Making. 2025. No. 2. P. 114–129 (in Russian)].
10. Бережнов Н.И., Сирота А.А. Модели глубокого обучения для синтеза изображений с включением атмосферных осадков с целью решения задач компьютерного зрения в различных погодных условиях // Вестник Воронежского гос. ун-та. Серия: Системный анализ и информационные технологии. 2025. № 2. С. 89–104. [Berezhnov N.I., Sirota A.A. Deep learning models for synthesizing images with precipitation effects for computer vision tasks under various weather conditions // Vestnik Voronezhskogo Gosudarstvennogo Universiteta. Series: Systems Analysis and Information Technologies. 2025. No. 2. P. 89–104 (in Russian)].
11. Liang J., Cao J., Sun G., et al. SwinIR: Image restoration using Swin transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). 2021. IEEE, 2021. P. 1833–1844. DOI: 10.1109/ICCVW54120.2021.00058.
12. Бережнов Н.И., Сирота А.А. Совершенствование механизмов внимания для архитектуры трансформер в задачах повышения качества изображений // Компьютерная оптика. 2024. Т. 48. № 5. С. 726–733. [Berezhnov N.I., Sirota A.A. Improving attention mechanisms for the transformer architecture in image quality enhancement tasks // Computer Optics. 2024. Vol. 48. No. 5. P. 726–733 (in Russian)].
13. Buslaev A., Iglovikov V.I., Khvedchenya E., et al. Albumentations: Fast and flexible image augmentations // Information. 2020. Vol. 11. P. 125. DOI: <https://doi.org/10.3390/info11020125>.
14. Zhu J.-Y., Park T., Isola P., Efros A.A. Unpaired Image-to-Image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2017. IEEE, 2017. P. 2223–2232. DOI: 10.1109/ICCV.2017.244.
15. Zhou K., Zhao S., Deng H., et al. High-Resolution rainy image synthesis: Learning from rendering [arXiv preprint]. arXiv:2502.16421. arXiv, 2025. DOI: 10.48550/arXiv.2502.16421.

Об авторах:

БЕРЕЖНОВ Никита Игоревич, ассистент кафедры технологий обработки и защиты информации, факультет компьютерных наук, ФГБОУ ВО Воронежский государственный университет, berezhnovnikita@yandex.ru, ORCID: 0000-0002-3532-1002.

СИРОТА Александр Анатольевич, д-р техн. наук, проф., заведующий кафедрой технологий обработки и защиты информации, факультет компьютерных наук, ФГБОУ ВО Воронежский государственный университет, sir@cs.vsu.ru, ORCID: 0000-0002-5785-8513.

Metadata:

Title: Integrated system for image restoration and augmentation using deep transformer-based neural networks.

Author 1: Nikita Igorevich Berezhnov, Assistant Lecturer, Department of Information Processing and Information Security Technologies, Faculty of Computer Sciences, Voronezh State University, 1 Universitetskaya Square, 394018 Voronezh, Russia, email: berezhnovnikita@yandex.ru, ORCID: 0000-0002-3532-1002, Web of Science Researcher ID: OPN-6966-2025.

Author 2: Alexander Anatolyevich Sirota, Doctor of Technical Sciences, Professor, Head of the Department of Information Processing and Information Security Technologies, Faculty of Computer Sciences, Voronezh State University, 1 Universitetskaya Square, 394018 Voronezh, Russia, email: sir@cs.vsu.ru, ORCID: 0000-0002-5785-8513, Web of Science Researcher ID: JED-6201-2023, Scopus Author ID: 7005577069.

Abstract: The paper presents a system for automatic restoration and augmentation of scene images distorted by various types of noise, primarily atmospheric precipitation and other weather-related degradations, based on deep neural network transformer models. Unlike traditional approaches, where the tasks of image restoration and the generation/synthesis of artificial distortions are solved separately, the proposed integrated system treats them as mutually inverse processes and combines them within a single architecture. The complex includes an augmentation module that synthesizes weather and other distortions in images and a restoration module that removes these distortions; both modules are built on transformer models with a modified attention mechanism. The system is trained in a cyclic scheme: the augmentation module produces distorted images that are then restored by the restoration module via an inverse transformation, while loss functions and perceptual error metrics are jointly optimized to ensure the realism of both generated and restored images. Experimental results demonstrate that the proposed approach outperforms many state-of-the-art models in terms of quality metrics such as PSNR, SSIM, FID, and others, while its novelty lies in the simultaneous solution of image augmentation and restoration tasks within a single framework capable of effectively modeling diverse weather conditions and distortions and improving images captured under unstable shooting conditions.

Keywords: neural networks, transformers, attention mechanism, data augmentation, image restoration, image augmentation.