

ОТЗЫВ

официального оппонента на диссертацию Бережного Никиты Игоревича «Совершенствование механизмов внимания в глубоких нейронных сетях – трансформерах в задачах восстановления и аугментации изображений», представленную на соискание ученой степени кандидата технических наук по специальности 1.2.1. Искусственный интеллект и машинное обучение.

Актуальность. Современные системы компьютерного зрения предъявляют высокие требования к качеству визуальной информации, обрабатываемой в автоматическом режиме. Однако в реальных условиях изображения, поступающие на вход алгоритмов, часто оказываются искаженными шумами, погодными явлениями, ограничениями сенсоров или ошибками передачи данных. Задачи восстановления и улучшения качества изображений, а также эффективной аугментации обучающих выборок приобретают особую значимость на фоне роста сложности прикладных задач и разнообразия условий съемки. Методы глубокого обучения, особенно архитектуры на основе трансформеров, в последние годы успешно адаптированы для задач компьютерного зрения и демонстрируют высокую устойчивость результатов к сложным типам искажений. Базовый механизм внимания (self-attention), лежащий в основе трансформеров, позволяет учитывать локальные и глобальные взаимосвязи между элементами изображения, что крайне важно для устранения артефактов, шума, перекосов освещения и других отклонений от исходного вида изображения.

Вместе с тем, практическое применение трансформерных моделей в задачах восстановления изображений сопряжено с рядом ограничений. Во-первых, механизм внимания может избыточно фокусироваться на отдельных деталях, снижая эффективность устранения шумов и искажений. Во-вторых, большие глубины архитектуры трансформеров делают их склонными к переобучению при ограниченных объемах данных. В-третьих, современные трансформерные модели для обработки изображений отличаются высокой вычислительной сложностью. Наконец, сбор обширного и разнообразного датасета реальных изображений затруднен, особенно для редких или сложных сценариев, что усложняет обучение глубоких нейронных сетей.

Решение проблемы нехватки данных видится в использовании методов аугментации, включая генеративные нейросетевые модели, способные формировать новые условно-реалистичные изображения. Такой подход особенно актуален для случаев сложных, нелинейных или локальных помех, где традиционные методы фильтрации малоэффективны. Генеративные модели (GAN, VAE, диффузионные и др.) позволяют синтезировать изображения с имитацией атмосферных осадков, цифровых дефектов, шумов различных типов. Это открывает возможности как для создания больших обучающих наборов данных, так и для реализации комплексных архитектур, в которых восстановление и аугментация изображений являются взаимосвязанными этапами единого процесса обработки.

Таким образом, перечисленные проблемы и тенденции указывают на актуальность темы диссертации, посвященной совершенствованию механизмов внимания в глубоких нейросетях-трансформерах и разработке **новых** архитектур для задач восстановления и аугментации изображений.

Краткая характеристика содержания диссертации

Диссертация состоит из введения, четырех глав, заключения, списка литературы из 108 наименований и двух приложений. Работа изложена на 161 странице машинописного текста, включает 28 рисунков и 16 таблиц.

Основные научные результаты диссертации, их новизна и степень обоснованности заключаются в следующем.

1. Предложен метод структурной регуляризации механизма внимания в трансформерных блоках нейросети путем добавления мультипликативной либо эквивалентной ей аддитивной стохастической составляющей при вычислении весов внимания. Данный прием позволяет сглаживать распределение коэффициентов внимания и снижает риск их неконтролируемого роста в процессе обучения. Тем самым повышается устойчивость модели к переобучению и избыточной концентрации внимания на отдельных элементах изображения.

2. Разработана модификация механизма внимания с отдельно обучаемой матрицей масштабирования. В стандартный self-attention введена мультипликативная составляющая в виде специальной матрицы коэффициентов, обучаемых совместно с моделью. Это позволяет компенсировать аномалии внимания, когда отдельные веса доминируют над прочими, а также избежать ситуаций насыщения активаций. Включение

масштабируемых коэффициентов предоставляет дополнительные возможности тонкой настройки распределения внимания и повышает стабильность обучения модели.

3. Показано, что качество восстановления изображений в глубоких нейросетях можно повысить за счет комбинирования предложенных механизмов внимания. В частности, автором реализованы: добавление аддитивной стохастической компоненты в виде выборочных оценок дисперсии признаков, вычисляемых в предшествующих сверточным слоям, совместно с использованием обучаемых матриц масштабных коэффициентов внимания. Эти модификации интегрированы в стандартные архитектуры нейросетей и обеспечивают более эффективное устранение искажений по сравнению с базовыми реализациями без указанных дополнений. Иными словами, внедрение стохастической регуляции внимания и адаптивного масштабирования приводит к существенному улучшению метрик качества восстановленных изображений.

4. Предложена оригинальная архитектура двухвходового трансформера для аугментации изображений с включением погодных эффектов. Разработанная модель объединяет сверточный энкодер-декодер и перекрестный механизм внимания (cross-attention) для переноса эффектов осадков (дождя, снега, тумана) с эталонного изображения на целевое. Тем самым реализован эффективный и малозатратный метод обогащения данных: погодные факторы, ухудшающие восприятие сцен, могут синтетически накладываться на любые исходные изображения.

В целом, основные выводы и результаты диссертации представляются научно обоснованными, так как они базируются на корректном применении математического аппарата, анализе и интерпретации полученных теоретических и экспериментальных результатов. Широкая апробация исследований в научной печати и на международных конференциях подтверждают корректность выводов и правильность представленных в работе результатов.

Достоверность положений, выносимых на защиту, подтверждается использованием комплекса взаимодополняющих методов: строгих математических моделей, вычислительных экспериментов и статистического анализа данных. Все разработанные нейросетевые модели, модифицированные механизмы внимания и алгоритмы аугментации прошли тщательное тестирование в контролируемых условиях на синтетических и реальных

дасетах с применением общепринятых метрик качества. Полученные результаты сопоставлены с базовыми аналогами и альтернативными подходами, что показало преимущество предложенных решений. Многократные вычислительные эксперименты, включая статистическую обработку результатов, подтвердили воспроизводимость работы разработанных алгоритмов. Выводы диссертации имеют убедительную интерпретацию: в частных случаях они согласуются с известными результатами других авторов и не противоречат основным положениям теории машинного обучения и обработки изображений. Таким образом, комплексное применение теоретических и экспериментальных методов исследования, а также согласованность полученных результатов с результатами независимых исследований, свидетельствуют о высокой степени их достоверности.

Значимость диссертации для науки и практики. Основные результаты диссертации обладают определенной теоретической значимостью. Развитые подходы к регуляризации механизма внимания расширяют методологический аппарат глубокого обучения, внося вклад в теорию обучения трансформерных моделей. Предложенные алгоритмы объединения задач восстановления и аугментации открывают новые возможности для создания комплексных систем компьютерного зрения. Работа отвечает актуальным потребностям развития интеллектуальных систем обработки визуальной информации и принятия решений. В частности, результаты диссертации позволяют провести сравнительный анализ альтернативных подходов к обработке изображений, сочетающих статистические методы, трансформеры и генеративные модели, что важно для дальнейшего прогресса в данной области.

Предложенные алгоритмы и модели могут быть непосредственно применены при решении широкого круга прикладных задач, связанных с улучшением качества изображений. Например, разработанные методы актуальны для систем видеонаблюдения и мониторинга обстановки; обработки аэрокосмических снимков, полученных в сложных метеоусловиях; медицинской визуализации (например, томографии) для подавления шумов и артефактов; а также для встроенных систем обработки изображений в мобильных устройствах. Алгоритмы аугментации, особенно модель WeatherTransformer, позволяют автоматически генерировать синтетические обучающие данные, имитирующие реальные условия съемки, без

необходимости их ручного сбора. Это имеет прямое практическое значение при обучении нейросетей на недостаточных или несбалансированных выборках.

Оценивая содержимое диссертации в целом, следует отметить, что она представляет собой законченную работу, в которой достаточно подробно и логично изложены постановки решаемых задач, особенности применяемых методов исследований

Автореферат достаточно полно отражает основное содержание диссертации.

Результаты диссертации нашли апробацию и применение на практике. Основные положения и выводы докладывались и обсуждались на ряде международных и всероссийских научных конференций. По теме диссертации опубликовано 8 научных работ, в том числе 4 статьи в изданиях, рекомендованных ВАК, и 1 статья в материалах международной конференции (IEEE Xplore, индексируемой Scopus); кроме того, получено свидетельство о государственной регистрации программы для ЭВМ. Результаты исследования внедрены при выполнении научно-исследовательских проектов в Воронежском государственном университете в 2022–2025 гг., в которых автор принимал непосредственное участие, а также используются в учебном процессе вуза.

Соответствие диссертации паспорту научной специальности.

Диссертационная работа соответствует паспорту научной специальности 1.2.1. Искусственный интеллект и машинное обучение по пунктам п.4 и п.14.

Вместе с тем по диссертации можно сделать следующие **замечания**.

1. Параметры стохастической регуляризации внимания. В работе предложено добавление случайной составляющей при вычислении весов внимания для их сглаживания. Однако не приведены явные рекомендации по выбору параметров такой регуляризации, например, масштаба добавляемого шума или вида распределения. Более подробный анализ чувствительности модели к этим параметрам повысил бы воспроизводимость метода на практике.

2. Сравнение с существующими подходами к оптимизации внимания. Предложенный подход канального сжатия и внедрения обучаемых коэффициентов внимания направлен на повышение эффективности трансформеров. Было бы полезно дополнительно сопоставить эти решения с известными методами оптимизации self-attention (например, механизмами снижения размерности или приближенными вычислениями внимания, такими как Linformer, Swin и др.), чтобы количественно показать преимущества

выбранного подхода по скорости и требуемым ресурсам. В диссертации основное внимание уделено сравнению качества восстановления изображений, тогда как оценка выигрыша в вычислительной сложности представлена менее подробно.

3. Обобщаемость архитектуры WeatherTransformer. Разработанная модель для переноса погодных эффектов показала свою эффективность в задачах генерации дождя, снега и тумана. Однако в работе практически не обсуждается, возможно ли расширение данного подхода на другие виды искажений вне погодных явлений. Обсуждение потенциала предложенной архитектуры для более широкого класса аугментаций укрепило бы прикладную ценность результатов. Также следовало бы привести более детальное сравнение модели с альтернативными способами генерации погодных эффектов, чтобы подчеркнуть сильные стороны и ограничения предложенного решения.

4. Ресурсоемкость предложенных моделей. Несмотря на заявленное снижение вычислительной сложности отдельных модулей, в диссертации не приведены количественные оценки требуемых ресурсов оперативной памяти, времени обучения или инференса для реализованной архитектуры трансформера до и после оптимизации. Включение подобных данных позволило бы более наглядно продемонстрировать практическую эффективность предложенных усовершенствований, особенно учитывая актуальность темы оптимизации и снижения ресурсоемкости больших нейросетевых моделей.

Указанные замечания существенным образом не снижают общего положительного впечатления о работе.

Вывод. Диссертационная работа «Совершенствование механизмов внимания в глубоких нейронных сетях – трансформерах в задачах восстановления и аугментации изображений» является завершенным научным исследованием, в котором получены новые, обоснованные и практически значимые результаты. По своему содержанию и оформлению диссертация полностью соответствует требованиям, предъявляемым к работам на соискание ученой степени кандидата наук (в том числе положениям п. 9–14 Положения о присуждении ученых степеней, утвержденного Постановлением Правительства РФ № 842, в редакции от 16.10.24), а ее автор, Бережнов Никита Игоревич, заслуживает присуждения ученой степени кандидата технических наук по специальности 1.2.1. Искусственный интеллект и машинное обучение.

