ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

на диссертацию Ушакова Егора Николаевича
«Метод условной генерации изображений в диффузионных моделях со структурированным текстовым условием»,
представленную на соискание ученой степени кандидата физико-математических наук по специальности 2.3.5 – «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей»

Актуальность темы

Диссертация Е.Н. Ушакова направлена на создание алгоритмических решений для контролируемого синтеза визуальных объектов на основе диффузионных моделей со структурированным текстовым условием. В работе рассматривается важная для мультимодального анализа проблема: семантическая избыточность текстовых описаний, в которых одновременно с ключевыми признаками присутствуют синтаксический шум, повторы и контекстные элементы, не имеющие непосредственного визуального соответствия. Это затрудняет эффективное использование текстовой информации в задачах управления генерацией, особенно в условиях ограниченности длины контекста современных текстовых кодировщиков.

В диссертации эта задача исследуется на примере медицинских данных. Радиологические отчеты содержат сведения о наблюдаемых изменениях, их локализации и выраженности, а также отрицания, предположения и ссылки на предшествующие исследования. Учет этих особенностей необходим при формировании условия генерации. Разработка компактного представления с явно выделенными признаками и их атрибутами имеет значение для подготовки синтетических обучающих выборок и развития программных систем генерации изображений.

Достоверность и обоснованность научных положений и выводов

Обоснованность научных результатов подтверждается сформулированными и доказанными математическими результатами, ориентированными на анализ свойств предлагаемого метода. В частности, в диссертации доказаны три теоремы и семь вспомогательных лемм. Е.Н. Ушаковым исследована вычислительная сложность предварительного формирования условия, которое используется в диффузионной модели.

Экспериментальная проверка построена как сопоставление двух вариантов текстового условия при использовании идентичных массивов изображений, разбиении данных, архитектуре генератора и протоколе обучения. Для выбора итоговых состояний моделей используется единый валидационный критерий. Оценка качества генерации

дополнена проверкой классификаторов на независимых реальных данных CheXpert, анализом сходства с обучающими изображениями и экспериментами по оценке возможности определения принадлежности к обучающей выборке и повторной идентификации пациента.

Научная новизна результатов исследования

Научная новизна работы заключается в получении следующих результатов:

1. Разработан метод декомпозиции неструктурированного текстового описания на семантически значимые компоненты с последующим формированием условия для диффузионной генерации изображений. На примере медицинских данных предложено формализованное представление на основе упорядоченных признаков медицинской визуализации, включающих тип признака, локализацию, статус и степень выраженности.

2. Разработаны алгоритмы формирования структурированного условия и получены аналитические оценки их вычислительной и пространственной сложности. Показано, что использование заранее сформированного условия при неизменной архитектуре и общем ограничении длины текстового входа сохраняет асимптотический порядок затрат на обучение и генерацию латентной диффузионной модели.

3. Разработаны архитектура и программная реализация комплекса условной генерации, объединяющего подготовку текстовых описаний, формирование структурированного условия, обучение и применение модели, а также средства экспериментальной оценки. Реализация позволяет сопоставлять исходное и структурированное представления текста в рамках единого конвейера обработки данных.

Теоретическая и практическая значимость

Теоретическая значимость диссертации заключается в формализации процедур преобразования слабоструктурированной текстовой информации в дискретные наборы признаков, а также в аналитическом обосновании асимптотических свойств соответствующих алгоритмов. Полученные представления и оценки могут использоваться при исследовании методов контекстного управления генеративными архитектурами и разработке оптимизированных алгоритмов обработки текстовой компоненты в мультимодальных системах.

Практическая значимость определяется созданием программного комплекса, автоматизирующего весь цикл условной генерации, включающей этапы предобработки текстов, обучения генеративных моделей и валидации синтетических изображений. Продемонстрировано применение разработанного метода на примере датасета MIMIC-CXR и независимого набора данных CheXpert.

Оценка содержания и завершенности диссертации

Диссертация состоит из введения, трех глав, заключения и списка литературы из 125 позиций. Общий объем работы составляет 119 страниц, включая 9 рисунков и 22 таблицы. Структура работы соответствует поставленной цели и отражает переход от анализа проблемы к разработке метода и его экспериментальной проверке.

Во введении обоснована актуальность темы, сформулированы цель и задачи исследования, положения, выносимые на защиту, научная новизна и значимость результатов.

В первой главе рассмотрены основные классы генеративных моделей и способы задания условия. Проанализированы особенности радиологических описаний, подходы к структурированию текста и методы оценки синтетических изображений. По итогам обзора сформулированы требования к разрабатываемому методу.

Во второй главе изложено основное содержание предложенного метода. Введены формальные представления отчета и клинической находки, описаны алгоритмы преобразования текста и сериализации результата, приведены доказательства оценок сложности. Рассмотрено использование структурированного условия в Stable Diffusion 1.5 и архитектура программного комплекса.

В третьей главе представлены данные, протоколы и результаты экспериментов. Сопоставлены два способа задания текстового условия, проверены полезность синтетических данных для классификации и отдельные риски сохранения информации об обучающих примерах в сгенерированных изображениях.

В заключении сформулированы основные результаты, соответствующие поставленным задачам.

Публикации и соответствие автореферата диссертации

По теме диссертации опубликованы четыре статьи в журналах. Результаты докладывались на Открытой конференции ИСП РАН, конференции молодых ученых Центра диагностики и телемедицины Департамента здравоохранения города Москвы и международной конференции «Иванниковские чтения». В автореферате в список публикаций добавлена ссылка на еще один сборник трудов конференции «Annual Computers, Software, and Applications Conference», однако в диссертации данная работа не упоминается. При этом автореферат в целом отражает основное содержание диссертации, полученные результаты и личный вклад соискателя.

Замечания по диссертационной работе

По содержанию диссертации имеются следующие замечания:

1. Для сформулированных в главе 2 лемм и теорем об асимптотической сложности алгоритмов разложения текста (раздел 2.3) отсутствует обсуждение их влияния на сложность и эффективность построенного решения в целом. Требуется обсуждение компромисса между повышением точности и дополнительными затратами на предобработку, предлагаемую соискателем в работе.

2. В разделе 2.4 не приводится аналитическое обоснование алгоритма ранжирования клинических находок по степени их диагностической приоритетности, что создает риск потери редких патологий при жесткой обрезке длинных отчетов из-за ограничений на длину контекста (77 токенов) у используемого в архитектуре текстового кодировщика CLIP.

3. В главе 3 отсутствует прямое экспериментальное сопоставление разработанного метода на основе ресурсоемких диффузионных архитектур с альтернативными решениями в исследуемой задаче, например, на основе генеративно-состязательных сетей.

4. Представленный в разделе 3.6 анализ рисков повторной идентификации пациентов выполнен на ограниченном объеме примеров. Предложенный подход не содержит количественной оценки вероятности утечки конфиденциальной информации, что затрудняет объективную оценку уровня защищенности разработанного метода от утечек данных.

5. Ограничения исследования в разделе 3.8 сформулированы достаточно поверхностно: не рассматриваются вопросы минимально необходимого объема обучающих данных для разработанного метода, не обсуждается возможность масштабирования решения на другие типы медицинских изображений и патологии органов, не рассмотренных в диссертации.

6. В Таблицу 21 (раздел 3.7) при оценке практической полезности синтетических изображений для обучения прикладных классификаторов анализируются показатели (метрики) различных моделей, однако в них не применяются разработанные соискателем методы.

7. В тексте диссертации отсутствует необходимая экспертная оценка врачами-радиологами реалистичности и диагностической точности изображений, синтезированных с помощью предложенного соискателем метода.

Заключение

Диссертация Ушакова Егора Николаевича «Метод условной генерации изображений в диффузионных моделях со структурированным текстовым условием» является научно-

квалификационной работой, в которой решена задача разработки метода формирования структурированного текстового условия для диффузионной генерации изображений, имеющая значение для развития математического и программного обеспечения вычислительных систем.

Работа в целом соответствует требованиям пункта 9 Положения о присуждении ученых степеней, утвержденного постановлением Правительства Российской Федерации от 24 сентября 2013 г. № 842 (в действующей редакции), предъявляемым к диссертациям на соискание ученой степени кандидата наук, а ее автор, Ушаков Егор Николаевич, заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 2.3.5 – «Математическое и программное обеспечение вычислительных систем, комплексов и компьютерных сетей».

Официальный оппонент

д.ф.-м.н., доцент Горшенин Андрей Константинович

«27» октября 2026 г.