

Ушаков Егор Николаевич

**Метод условной генерации изображений в диффузионных моделях
со структурированным текстовым условием**

Специальность 2.3.5 –
«Математическое и программное обеспечение вычислительных
систем, комплексов и компьютерных сетей»

Автореферат
диссертации на соискание ученой степени
кандидата физико-математических наук

Москва – 2026

Работа выполнена в Федеральном государственном бюджетном учреждении науки Институте системного программирования им. В. П. Иванникова Российской Академии Наук

Научный руководитель:

Карпулевич Евгений Андреевич,
кандидат физико-математических наук

Официальные оппоненты:

Горшенин Андрей Константинович,
доктор физико-математических наук, доцент,
руководитель отдела «Вероятностное моделирование
и статистический анализ сложных систем»
Федерального государственного учреждения
«Федеральный исследовательский центр
«Информатика и управление» Российской академии
наук»

Кулеев Рамиль Фуатович,
кандидат технических наук, директор Института
искусственного интеллекта Автономной
некоммерческой организации высшего образования
«Университет Иннополис»

Ведущая организация:

Федеральное государственное автономное
образовательное учреждение высшего образования
«Национальный исследовательский ядерный
университет «МИФИ»

Защита состоится 12 ноября 2026 г. в 13-30 на заседании диссертационного совета 24.1.120.01 при Федеральном государственном бюджетном учреждении науки Институте системного программирования им. В.П. Иванникова Российской академии наук по адресу: 115035, г. Москва, ул. Садовническая, д. 41, ст. 2.

С диссертацией можно ознакомиться в библиотеке и на сайте Федерального государственного бюджетного учреждения науки Института системного программирования им. В.П. Иванникова Российской академии наук.

Автореферат разослан «_____» _____ 20__ г.

Ученый секретарь

диссертационного совета 24.1.120.01,
кандидат физико-математических наук

Турдаков Д.Ю.

Общая характеристика работы

Актуальность проблемы. Современная прикладная математика и математическая статистика располагают развитым аппаратом для работы с неопределенностью и сложными распределениями: байесовским выводом, методами максимального правдоподобия, вариационными приближениями, стохастической оптимизацией, а также построением параметрических и непараметрических моделей данных. Существенный прогресс последних лет связан с развитием генеративного моделирования, в рамках которого задача формулируется как восстановление (аппроксимация) неизвестного распределения данных и последующая генерация новых примеров из этого распределения.

Значительный вклад в развитие методов анализа изображений, машинного обучения и генеративного моделирования внесли как российские, так и зарубежные научные школы. В России данное направление развивается в рамках школ компьютерного зрения, анализа изображений, обработки больших данных. К ним относятся, в частности, школа Ю. В. Визильтера, связанная с развитием методов обработки и анализа изображений, распознавания визуальных объектов, машинного и технического зрения, а также школа Д. С. Ватолина, ориентированная на методы обработки изображений и видео, компьютерного зрения, машинного обучения, сжатия и оценки качества цифрового видео. Близкие по тематике исследования ведутся также в научных коллективах AIRI, ФИЦ ИУ РАН и Сколтеха, где развиваются методы глубокого обучения, мультимодального анализа данных, генеративного моделирования и прикладного искусственного интеллекта. Отдельно следует отметить направление, связанное со школой И. А. Макарова в AIRI, в рамках которого развиваются методы промышленного искусственного интеллекта, компьютерного зрения, машинного обучения на графах и прикладного анализа данных. В зарубежной научной традиции существенный вклад в становление генеративного моделирования внесли работы Д. Румельхарта, Д. Хинтона и Р. Уильямса по обучению многослойных нейронных сетей, исследования Д. Кингмы и М. Веллинга, предложивших вариационные автокодировщики, И. Гудфеллоу и соавторов, разработавших генеративно-состязательные сети, а также Я. Сохла-Дикстейна и соавторов, заложивших основы диффузионного подхода к генерации. В дальнейшем Дж. Хо, А. Джейн и П. Аббил развили этот подход в модели диффузионного вероятностного моделирования, а Р. Ромбах и соавторы показали его высокую практическую эффективность в латентном пространстве изображений.

Одной из важных постановок задачи генерации является условная генерация изображений. При условной генерации на вход модели помимо исходных данных передается дополнительная информация - условие (condition), которое задает требуемые свойства генерируемого изображения. В качестве такого условия могут выступать текстовое описание, набор признаков, метка класса, контекст или иная дополнительная информация, определяющая требуемые свойства результата. В этой постановке требуется построить генеративную модель, способную синтезировать объекты данных, в частности изображения, по заданному условию. С математической точки зрения здесь возникают две взаимосвязанные задачи: во-первых, необходимо выбрать такое представление условия, которое сохраняет существенную информацию, уменьшает неоднозначность и допускает эффективную вычислительную обработку; во-вторых, необходимо включить это условие в процесс генерации таким образом, чтобы обеспечить не только высокое качество синтезируемых объектов, но и их устойчивое соответствие заданному условию. На практике условие нередко задается в неструктурированном виде - например, в форме текста или набора слабо формализованных

признаков, - что затрудняет его интерпретацию моделью и может приводить к снижению управляемости генерации и ухудшению воспроизводимости результатов.

Диффузионные модели представляют собой один из наиболее успешных современных классов генеративных моделей, опирающийся на вероятностную интерпретацию постепенного зашумления данных и обратного стохастического процесса восстановления. Их практическая эффективность обусловлена устойчивостью обучения и способностью аппроксимировать сложные распределения в высоких размерностях. Однако в задаче условной генерации с точки зрения корректности особенно важно, каким образом задано условие и как оно «встраивается» и влияет на обратный процесс: недостаточно получить реалистичные примеры, необходимо обеспечить стабильное выполнение ограничений и уменьшить неопределенность интерпретации условия.

Описанные методы и алгоритмы нашли широкое применение в прикладных областях, где данные имеют сложную структуру, высокую размерность и существенную внутреннюю вариативность. К ним относятся анализ изображений и видеоданных, обработка текстовой и мультимодальной информации, моделирование сложных объектов и процессов, а также построение интеллектуальных систем поддержки принятия решений. Во всех этих задачах возникает необходимость не только в извлечении закономерностей из имеющихся данных, но и в построении моделей, способных воспроизводить структуру исходного распределения, формировать новые примеры и управляемо изменять их свойства в соответствии с заданными условиями.

С математической точки зрения генерация синтетических данных связана с задачами аппроксимации сложных многомерных распределений, оценки неопределенности, построения устойчивых представлений данных и разработки алгоритмов оптимизации, пригодных для обучения моделей с большим числом параметров. При этом особую роль играет условная генерация, поскольку в прикладных задачах требуется не просто синтезировать реалистичные объекты, а получать данные, удовлетворяющие заданным ограничениям: классовым меткам, текстовым описаниям, наборам признаков, структурным характеристикам или иным формализованным условиям. Поэтому качество генерации определяется не только архитектурой модели, но и тем, насколько полно, компактно и однозначно задано условие.

Практическое применение генеративных моделей часто ограничено дефицитом размеченных данных, дисбалансом классов, редкостью отдельных состояний или конфигураций объектов, а также трудоемкостью получения экспертной разметки. В таких условиях синтетические данные могут использоваться для расширения и балансировки обучающих выборок, моделирования редких случаев, анализа устойчивости алгоритмов и тестирования интеллектуальных систем. Однако использование синтетических данных требует обеспечения их соответствия заданному условию: если условие задано неоднозначно или содержит нерелевантную информацию, модель может воспроизводить формально правдоподобные, но плохо управляемые и слабо воспроизводимые результаты.

На практике условие нередко задается в неструктурированном виде - например, в форме свободного текстового описания, набора слабо формализованных признаков или сообщения эксперта. Такие представления могут содержать избыточные фрагменты, неоднозначные формулировки, отрицания, ссылки на контекст и признаки, не имеющие непосредственного соответствия в генерируемом объекте. Это затрудняет интерпретацию условия моделью, снижает управляемость генерации и ограничивает возможность формального анализа получаемого результата. В связи с этим актуальной является разработка методов структурированного задания условия диффузионных моделей, основанных на разложении

исходного текстового описания на значимые компоненты и представлении его в форме, пригодной для алгоритмической обработки.

В настоящей работе указанные задачи рассматриваются на примере генерации медицинских изображений, где проблема управляемой условной генерации имеет практическую значимость. Медицинские изображения характеризуются высокой размерностью, сложной визуальной структурой, ограниченной доступностью, высокой стоимостью экспертной разметки и необходимостью строгого соответствия синтезируемого изображения заданному клиническому описанию. При этом текстовые описания исследований и врачебные заключения содержат сложную клиническую семантику, допускают вариативность формулировок и представлены в неструктурированном виде. Поэтому применение диффузионных моделей со структурированным заданием условия в данной области позволяет исследовать общий математический и алгоритмический подход к управляемой генерации на практически важном и сложном классе данных.

Целью данной работы является разработка и экспериментальная валидация метода условной генерации синтетических изображений на основе диффузионных моделей со структурированным заданием условия, формируемым из текстового описания, а также создание программной реализации, обеспечивающей повышение качества генерации и практической полезности синтетических данных.

Для достижения поставленной цели необходимо решить следующие **задачи**:

1. Разработать метод разложения текстового описания для задачи диффузионной генерации изображений, позволяющий выделять семантически значимые компоненты и представлять их в виде структурированного условия, и исследовать его на примере медицинских данных
2. Разработать способ использования структурированного условия в латентной диффузионной модели и оценить его влияние на метрики качества генерации по сравнению с использованием неструктурированного текста.
3. Выполнить экспериментальную проверку предложенного метода формирования структурированного условия на медицинских данных, включающую сопоставление с базовым способом задания условия, анализ сходства синтетических и обучающих изображений, оценку возможности определения принадлежности данных к обучающей выборке и риска повторной идентификации пациентов, а также оценку практической полезности синтетических данных в downstream-задачах классификации.

Основные положения, выносимые на защиту:

1. Новый метод разложения текстового условия для диффузионной генерации изображений, обеспечивающий выделение значимых признаков и формирование структурированного условия.
2. Алгоритмы в составе метода разложения текстового условия и аналитические оценки их вычислительной и пространственной сложности через доказательство соответствующих теорем.
3. Архитектура и программная реализация комплекса условной генерации изображений со структурированным текстовым условием, объединяющего методы разложения текстового условия, формирования структурированного условия, обучение и применение латентной диффузионной модели, а также средства комплексной экспериментальной оценки на примере медицинских данных.

Научная новизна. Разработан новый метод структурированного представления текстового условия в задаче диффузионной генерации изображений, основанный на разложении

неструктурированного текстового описания на семантически значимые компоненты, их нормализации и формировании компактного условия генерации; метод разработан и исследован на примере медицинских данных. Разработаны алгоритмы формирования структурированного условия, включающие выделение значимых сущностей и их атрибутов, нормализацию, фильтрацию нерелевантной информации и разрешение противоречий. Для разработанных алгоритмов получены и доказаны оценки вычислительной и пространственной сложности. Показано, что замена исходного текстового условия его структурированным представлением не изменяет асимптотическую сложность обучения и генерации латентной диффузионной модели.

Теоретическая и практическая значимость. Теоретическая значимость работы заключается в развитии подходов к условному генеративному моделированию в диффузионных моделях за счет формализации процесса преобразования неструктурированного текстового описания в структурированное условие и анализа вычислительных свойств входящих в него алгоритмов. Полученные формальные представления, алгоритмическая схема и аналитические оценки могут использоваться как научная и методическая основа в последующих исследованиях и диссертационных работах, посвященных условной генерации, структурированию мультимодальных данных и оценке синтетических изображений. Практическая значимость заключается в программной реализации воспроизводимого конвейера, который может служить основой для разработчиков аналогичных систем: от подготовки текста и формирования условия до обучения и применения латентной диффузионной модели и комплексной оценки синтетических данных. Разработанные решения могут применяться при создании систем расширения и балансировки обучающих выборок, моделирования редких состояний и тестирования алгоритмов анализа медицинских данных.

Апробация работы. Результаты работы докладывались на следующих конференциях:

1. Открытая конференция ИСП РАН, Москва, Россия, декабрь 2021.
2. Открытая конференция ИСП РАН, Москва, Россия, декабрь 2023.
3. IV Открытая конференция молодых ученых ГБУЗ “НПКЦ ДиТ ДЗМ”, Москва, Россия, апрель 2024.
4. Международная конференция «Иванниковские чтения», Калужская область, Россия, май 2026

Личный вклад. Все выносимые на защиту результаты получены лично автором.

Публикации. Основные результаты по теме диссертации изложены в четырех работах, опубликованных в изданиях, рекомендованных ВАК.

В статье [1, 4] совместно с соавторами сформулирована задача. Автором выполнена основная часть работы: разработка архитектуры и обучение модели, проведение вычислительных экспериментов, анализ результатов и подготовка к публикации.

В статье [2] автором была разработана постановка задачи исследования. Автор принимал участие в обсуждении методологии и интерпретации результатов.

В статье [3] автором была разработана постановка задачи исследования и план экспериментов. Автор участвовал в постановке и планировании экспериментальных исследований, осуществлял проверку и валидацию полученных результатов, а также выполнял подготовку текста публикации.

Объем и структура работы. Диссертационная работа состоит из введения, трех глав и заключения. Работа изложена на 119 страницах, содержит 9 рисунков и 22 таблицы.

Содержание работы

Во **введении** обосновывается актуальность темы диссертационной работы, формулируются цель и задачи исследования, определяются объект и предмет исследования, приводятся положения, выносимые на защиту, а также излагается научная новизна и практическая значимость полученных результатов.

В **первой главе** выполнен обзор современных методов генерации изображений и сформулированы предпосылки разработки метода условной генерации изображений со структурированным заданием условия на примере медицинских изображений. В начале главы рассматривается задача генеративного моделирования как задача аппроксимации неизвестного распределения данных и последующего синтеза новых объектов, обладающих свойствами исходной выборки. Приводится краткая характеристика основных классов генеративных моделей, включая вариационные автокодировщики, генеративно-состязательные сети, нормализующие потоки, авторегрессионные модели и диффузионные модели.

Далее рассматриваются диффузионные модели как вероятностный подход к генерации изображений, основанный на прямом процессе постепенного зашумления данных и обратном процессе восстановления изображения из шума. Пусть x_0 обозначает исходное изображение, а x_t - его зашумленную версию на шаге t . Прямой диффузионный процесс задается соотношением:

$$q(x_t | x_{t-1}) = \mathcal{N}(x_t; \sqrt{1 - \beta_t} x_{t-1}, \beta_t I),$$

где β_t коэффициент шума на шаге t , I - единичная матрица. Тогда распределение x_t при фиксированном x_0 может быть записано в виде:

$$q(x_t | x_0) = \mathcal{N}(x_t; \sqrt{\bar{\alpha}_t} x_0, (1 - \bar{\alpha}_t)I),$$

где $\alpha_t = 1 - \beta_t$, $\bar{\alpha}_t = \prod_{s=1}^t \alpha_s$. В процессе обучения нейросетевая модель ε_θ учится предсказывать добавленный шум ε , а стандартная функция потерь имеет вид:

$$L = \mathbb{E}_{x_0, t, \varepsilon} [\|\varepsilon - \varepsilon_\theta(x_t, t)\|^2]$$

Описываются основные варианты реализации диффузионных моделей, включая модели в пиксельном пространстве и модели латентной диффузии. Отдельное внимание уделяется архитектурным компонентам современных text-to-image моделей: вариационному автоэнкодеру для перехода в латентное пространство, U-Net-подобной сети для обратного диффузионного процесса, текстовому энкодеру и механизмам внимания, обеспечивающим передачу условия в генеративную модель. Показано, что в моделях латентной диффузии качество генерации определяется не только архитектурой и объемом обучающих данных, но и способом представления условия, поступающего на вход текстового энкодера и далее в cross-attention блоки U-Net.

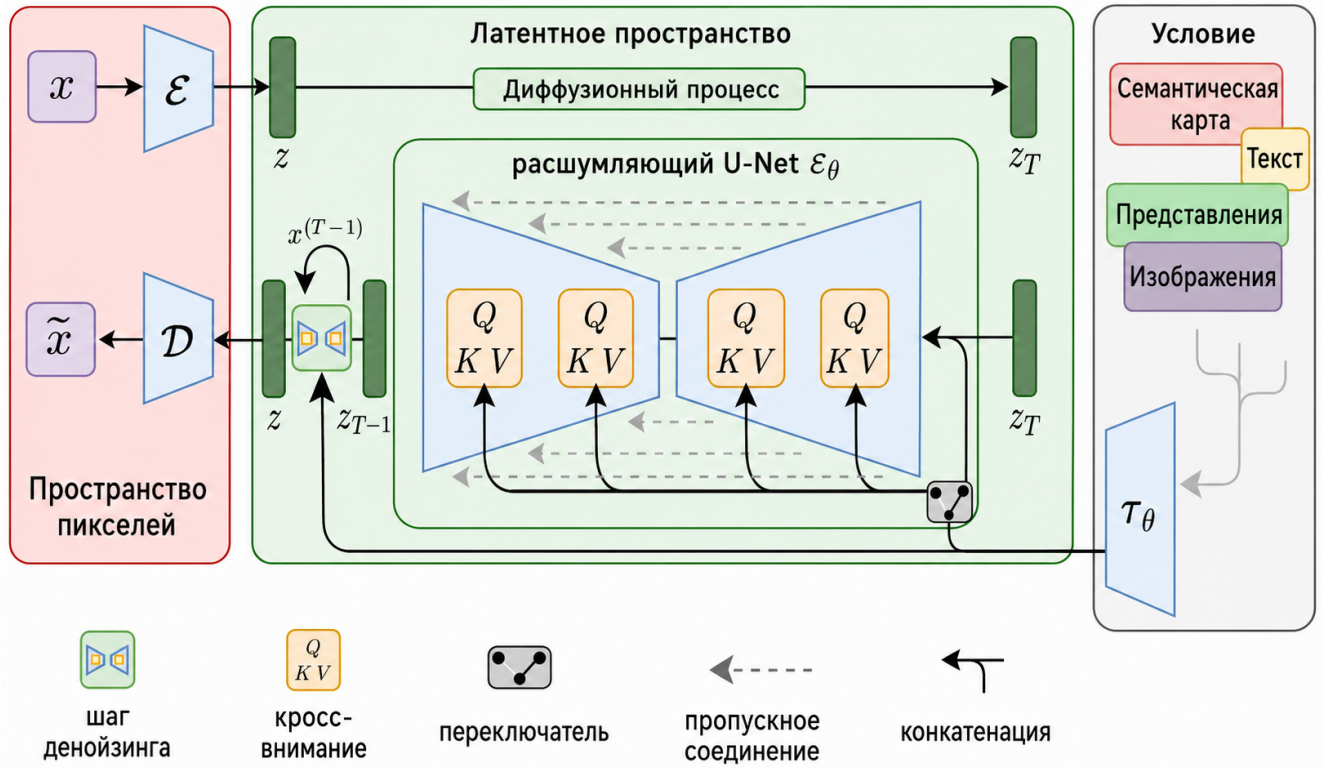


Рисунок 1 - Общая схема условной генерации изображения в модели латентной диффузии.

Затем анализируется задача условной генерации изображений. Для условной генерации модель предсказания шума дополнительно получает условие c , поэтому функция потерь записывается в виде:

$$L = \mathbb{E}_{x_0, t, \varepsilon} [\|\varepsilon - \varepsilon_\theta(x_t, t, c)\|^2],$$

где c обозначает условие генерации. Условие генерации рассматривается как внешняя информация, задающая требуемые свойства синтезируемого объекта: категориальную метку, набор признаков, текстовое описание, маску, изображение-пример или их комбинацию. Рассматриваются основные способы задания условия диффузионных моделей, включая использование дискретных меток, векторных признаков, текстовых описаний и мультимодальных представлений. Общая схема условной латентной диффузии приведена на Рисунке 1. В данной схеме текстовое условие кодируется текстовым энкодером и передается в U-Net через механизмы cross-attention. В работе рассматриваются два варианта такого условия: исходная секция FINDINGS и структурированный список находок.

Отдельный раздел главы посвящен особенностям медицинских изображений как объекта генерации. Рассматриваются высокая стоимость экспертной разметки, неоднородность протоколов получения изображений, дисбаланс классов, редкость отдельных патологических состояний, а также ограничения, связанные с доступом к данным пациентов. В существующих исследованиях медицинских изображений рассматриваются также ограничения, связанные с низкой вариативностью обучающих данных: например, в таких работах исследуется влияние выбора функции потерь на качество Vision Transformer при диагностике повреждений костей по

рентгенограммам¹. В задачах обработки несбалансированных изображений высокого разрешения применяются и графовые подходы, учитывающие локальные и глобальные зависимости между элементами изображения². На примере рентгенограмм органов грудной клетки показано, что синтетические изображения могут использоваться для расширения обучающих выборок, моделирования редких клинических случаев и исследования устойчивости алгоритмов анализа медицинских изображений. На Рисунке 2 показан пример изображения и текстового условия из открытого датасета MIMIC-CXR.

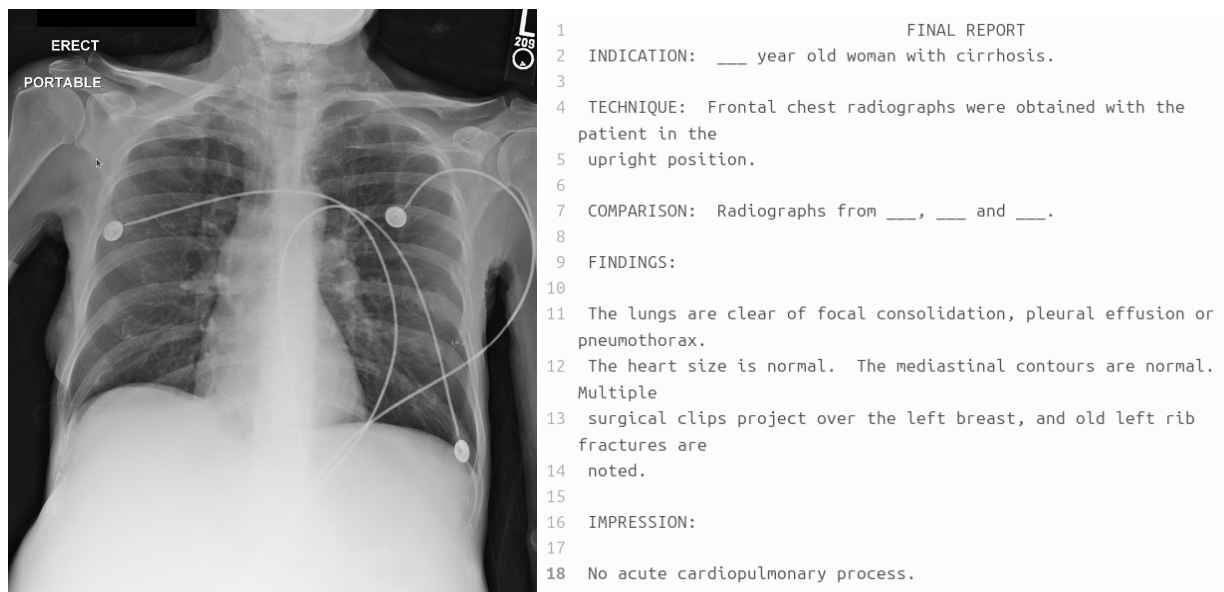


Рисунок 2 - Пример пары “изображение-аннотация” из открытого датасета MIMIC-CXR.

Далее рассматриваются существующие подходы к генерации медицинских изображений с использованием диффузионных моделей. Анализируются работы, посвященные синтезу рентгенограмм органов грудной клетки, маммографических изображений и других типов медицинских данных, а также применению синтетических изображений для аугментации, балансировки выборок и защиты конфиденциальности. Обсуждаются преимущества диффузионных моделей по сравнению с более ранними генеративными подходами, а также их ограничения.

Отдельно рассматривается текстовое задание условия медицинских диффузионных моделей, включая использование радиологических отчетов и их отдельных секций, в частности секции FINDINGS, в качестве текстового условия для генерации рентгенограмм органов грудной клетки. Показано, что такие тексты содержат клинически значимую информацию, однако представлены в свободной форме и могут включать шаблонные фразы, ссылки на предыдущие исследования, технические замечания, отрицания, неопределенные формулировки и фрагменты, не имеющие непосредственного визуального соответствия на изображении. В результате исходный текст отчета может неоднозначно интерпретироваться моделью, что снижает управляемость генерации и затрудняет воспроизводимое получение изображений с заданными признаками.

¹ Gorshenin A. K., Antonenko S. V. Loss Function Selection for Vision Transformers in the Automatic Diagnosis of Bone Damage Based on X-ray Images with Low Variability in Training Data //Pattern Recognition and Image Analysis. – 2026. – Т. 36. – №. 1. – С. 16-25.

² Горшенин А. К., Достовалова А. М. MMRFIGN: ансамблевая графовая модель сегментации несбалансированных изображений высокого разрешения, информированная мультикомпонентными марковскими случайными полями //Доклады Российской академии наук. Математика, информатика, процессы управления. – 2025. – Т. 527. – №. 0. – С. 156-170.

Затем рассматриваются подходы к структурированию медицинского текста. Рассматриваются возможные формы представления извлеченной информации: набор признаков, список находок и граф клинических сущностей и связей. Переход от свободного текстового описания к структурированному условию позволяет уменьшить влияние нерелевантных фрагментов отчета, повысить интерпретируемость условия и сформировать более устойчивую связь между клиническим описанием и генерируемым изображением.

Для сопоставления неструктурированного и структурированного представления условия приводится сравнительная характеристика двух подходов, которая представлена в Таблице 1.

Таблица 1 - Сравнение неструктурированного и структурированного текстового условия

Характеристика	Свободный текст	Структурированное условие
Наличие нерелевантных фрагментов	Возможно	Минимизировано
Интерпретируемость	Ограниченная	Более высокая
Учет локализации и тяжести	Неявный	Явный
Управляемость генерации	Снижена из-за неоднозначности	Выше за счет явного задания признаков
Пригодность для формального алгоритмического анализа	Ограниченная	Высокая

Из приведенного сравнения следует, что структурированное представление условия обладает рядом преимуществ по сравнению с исходным текстовым описанием, поскольку позволяет сделать клинически значимую информацию более компактной, интерпретируемой и пригодной для дальнейшей алгоритмической обработки.

В заключительной части первой главы формулируется содержательная постановка задачи. В качестве основы условной генерации выбирается латентная диффузионная модель, а условием служит радиологическое текстовое описание. Основная проблема связана с формой этого условия: секция FINDINGS содержит клинически значимую информацию вместе с отрицаниями, сравнительными, историческими и техническими фрагментами и допускает вариативность формулировок. Поэтому задача формулируется как разработка метода, который преобразует клиническое текстовое описание изображения в компактное структурированное условие, сохраняющее визуально обоснованные клинические признаки, и использует его в качестве условия в латентной диффузионной модели. Таким образом, уменьшение неоднозначности и исключение нерелевантных фрагментов текстового условия позволит повысить управляемость условной генерации и качество соответствия синтезируемых изображений заданным клиническим признакам.

Во **второй главе** приводится описание предложенного метода условной генерации изображений со структурированным заданием текстового условия, разработанного и исследованного на примере медицинских данных. В основе метода лежит преобразование исходного радиологического описания в формализованное представление клинически значимой информации с последующим использованием этого представления в качестве условия латентной диффузионной модели. Таким образом, метод связывает два последовательных этапа:

формирование структурированного условия из исходного текста и использование полученного условия для управления генерацией изображения.

На первом этапе из радиологического отчета выделяется секция FINDINGS, после чего содержащаяся в ней информация разлагается на клинически значимые компоненты и преобразуется в структурированный список находок. Каждая находка содержит сведения о типе визуально наблюдаемого признака, его наличии или отсутствии, анатомической локализации и степени выраженности. Полученный список приводится к текстовому представлению фиксированного формата и используется в качестве структурированного условия генерации. На втором этапе это условие передается через штатный текстовый энкодер латентной диффузионной модели и используется механизмами условной генерации при синтезе изображения.

Общая схема предложенного метода представлена на Рисунке 3. Левая часть схемы соответствует формированию структурированного условия из исходного радиологического описания, а правая - его использованию в стандартном механизме условной латентной диффузии.

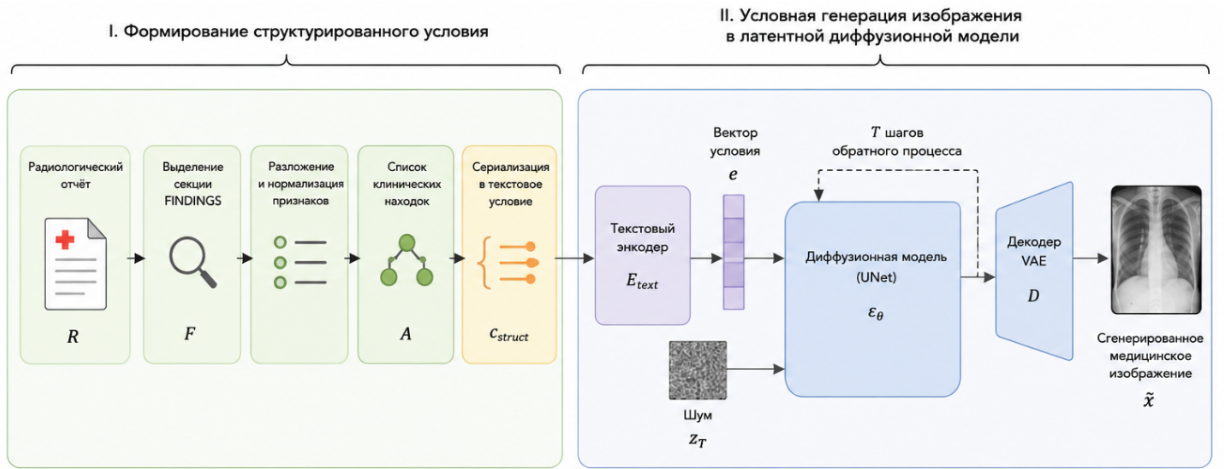


Рисунок 3 - Общая схема предложенного метода условной генерации изображений со структурированным текстовым условием.

Далее формализуется математическая постановка задачи. Радиологический отчет рассматривается как последовательность токенов, из которой требуется выделить множество клинически значимых утверждений. Для одного исследования структурированное условие представляется списком клинических находок:

$$A = \{a_1, a_2, \dots, a_m\},$$

где m - число найденных клинических находок. Каждая находка задается кортежем:

$$a_j = (\tau_j, p_j, \lambda_j, \rho_j),$$

где τ_j обозначает тип находки, p_j - признак наличия или отсутствия, λ_j - локализацию или латеральность, ρ_j - степень выраженности. Отдельно рассматриваются случаи отрицания, неопределенности и нерелевантных фрагментов текста, которые не должны усиливать условие

генерации. В результате формируется список находок, который служит новым условием для диффузионной модели.

Для иллюстрации различия между исходным текстовым условием и структурированным представлением в главе приводится пример преобразования секции FINDINGS в список находок в Таблице 2.

Таблица 2 - Пример преобразования секции FINDINGS в список находок

Исходная секция FINDINGS	Список находок
Hyperexpanded lungs with flattened diaphragms are unchanged. Previously described right tracheal deviation is not seen on the current study. Unchanged bilateral apical bronchial thickening. The lungs are otherwise clear without focal consolidation, pneumothorax, or effusion. The cardiomeastinal and hilar silhouettes are normal. Stable descending tortuous aorta.	hyperexpanded lungs bilateral lungs present none ; flattened diaphragm bilateral diaphragm present none ; bronchial thickening bilateral apical present none ; tortuous aorta descending aorta present none
PA and lateral views the chest provided. Biapical pleural parenchymal scarring noted. No focal consolidation concerning for pneumonia. No effusion or pneumothorax. No signs of congestion or edema. Cardiomeastinal silhouette is stable with an unfolded thoracic aorta and top-normal heart size. Bony structures are intact.	pleural parenchymal scarring bilateral apices present none
Portable AP chest radiograph. The lungs are relatively well expanded without focal consolidation, pleural effusion or pneumothorax. The heart is normal in size.	No findings

Из приведенного примера видно, что список находок сохраняет клинически значимую информацию, но устраняет часть вариативности свободного текста. Такое представление делает условие более компактным, интерпретируемым и пригодным для дальнейшей алгоритмической обработки.

Далее описывается алгоритм формирования списка находок из радиологического отчета. Алгоритм включает выделение секции FINDINGS из полного текста отчета, токенизацию и нормализацию текста, выделение клинически значимых сущностей и их атрибутов, фильтрацию нерелевантных фрагментов и формирование итогового списка находок. Полученное представление может быть использовано либо как линейное текстовое условие специального формата, либо как промежуточная структурированная форма, допускающая дальнейшее кодирование и передачу в генеративную модель. В текущей реализации в качестве структурированного условия используется именно список находок, что позволяет удалить из исходного текста нерелевантные фрагменты и сделать условие генерации более компактным и интерпретируемым. Псевдокод алгоритма формирования списка клинических находок приведен в Листинге 1.

Листинг 1 - Алгоритм формирования списка находок из радиологического отчета

```

Вход:
R - радиологический отчет

Выход:
A - список клинических находок

1. Function BuildFindingsList(R) begin
2.   F ← ExtractSection(R, "FINDINGS")
3.   T ← Normalize(Tokenize(F))
4.   E ← ExtractEntities(T)
5.   A ← ∅
6.   for each e in E do
7.     a ← BuildFinding(e) // тип, наличие, локализация, выраженность
8.     if IsValid(a) then
9.       A ← A ∪ {a}
10.    end if
11.  end for
12.  A ← MergeDuplicates(A)
13.  A ← ResolveConflicts(A)
14.  return A
15. end

```

Далее в главе приводится формальное описание вычислительной сложности предложенного алгоритма формирования списка находок из радиологического отчета.

Теорема 1. Пусть R - радиологический отчет, состоящий из L токенов. Пусть из отчета R выделяется секция FINDINGS, содержащая l токенов, где $l \leq L$. Для каждого токена секции FINDINGS выполняется не более d элементарных операций сопоставления, нормализации, классификации или фильтрации, а число типов клинических признаков, используемых в схеме представления списка находок, ограничено константой C . Тогда вычислительная сложность алгоритма формирования списка клинических находок из одного радиологического отчета:

$$O(L + l \cdot d + C).$$

Если параметры d и C фиксированы выбранной схемой алгоритма и не зависят от длины отчета, то вычислительная сложность формирования списка находок для одного отчета составляет:

$$O(L)$$

Затем рассматривается вопрос хранения полученного структурированного условия.

Теорема 2. Пусть список находок для одного исследования имеет вид:

$$A = \{a_1, a_2, \dots, a_m\},$$

где m - число найденных клинических находок. Предполагается, что каждая находка кодируется конечным набором атрибутов фиксированного формата и суммарный размер кодирования одной находки не превосходит a машинных слов.

Тогда пространственная сложность хранения списка находок для одного исследования оценивается как:

$$O(m \cdot a)$$

Если размер кодирования одной находки фиксирован, то есть $a = \text{const}$, то пространственная сложность хранения структурированного условия для одного исследования составляет:

$$O(m).$$

Данный результат показывает, что переход от свободного текста FINDINGS к списку находок не приводит к асимптотическому росту объема памяти, необходимого для хранения условий генерации.

После формального описания первой составляющей предложенного метода - формирования структурированного условия - рассматривается его вторая составляющая: использование полученного условия для управления генерацией изображения. В рамках экспериментальной реализации в качестве базовой генеративной модели используется латентная диффузионная модель типа Stable Diffusion, обучаемая на парах (x, c) , где x - медицинское изображение, а c - текстовое условие. В базовом варианте в качестве условия используется исходная секция $c_{base} = \text{FINDINGS}$, а в предлагаемом варианте - структурированное текстовое представление в виде списка клинических находок $c_{struct} = \text{List}(a_1, a_2, \dots, a_m)$.

При этом архитектура модели не изменяется: сохраняются текстовый энкодер, вариационный автоэнкодер, U-Net, scheduler и механизм текстового задания условия. Изменяется только содержательная форма текстового условия, подаваемого на вход модели.

Далее описывается использование структурированного условия при обучении латентной диффузионной модели. Для каждого изображения из обучающего набора формируется соответствующий список находок, который сериализуется в текстовое представление и подается через тот же текстовый энкодер, что и исходная секция FINDINGS. Последующее обучение выполняется стандартным для Stable Diffusion способом с функцией потерь, основанной на среднеквадратичной ошибке предсказания шума. Таким образом, предложенный метод не требует изменения базового механизма обучения Stable Diffusion, но изменяет информационное содержание условия, за счет чего ожидается повышение метрик качества генерации.

В заключительной части главы формулируется и доказывается результат о сложности этапов обучения и генерации. Предполагается, что структурированное условие уже сформировано и представлено в формате, совместимом с текстовым энкодером, а максимальная длина текстового входа ограничена сверху константой S , то есть $|c_{base}| \leq S, |c_{struct}| \leq S$.

Теорема 3. Пусть в модели латентной диффузии типа Stable Diffusion исходная секция FINDINGS заменяется на структурированное текстовое представление в виде списка клинических находок, причем список находок подается через тот же текстовый энкодер и тот же механизм текстового задания условия, а длина текстового условия ограничена сверху константой S . Обозначим через FV сложность операций VAE, через FU - сложность одного прохода U-Net, через $FE(s)$ - сложность текстового энкодера для условия длины s , через FB - сложность обратного распространения ошибки, а через T - число шагов генерации. Тогда сложность одного шага обучения имеет порядок:

$$O(F_V + F_E(S) + F_U + F_B),$$

а сложность генерации одного изображения:

$$O(F_E(S) + T \cdot F_U + F_V).$$

Тогда замена секции FINDINGS на список находок не изменяет асимптотическую сложность обучения и генерации модели Stable Diffusion.

В конце главы приводится описание программной реализации предложенного метода. Реализация включает модуль извлечения релевантной секции радиологического отчета, модуль формирования списка находок, модуль преобразования списка в формат условия для модели и модуль обучения диффузионной модели на подготовленных парах «изображение - структурированное условие». Отдельно указывается, что предложенный метод допускает расширение: вместо списка находок может использоваться более сложное структурированное представление, например граф клинических сущностей и отношений, однако в рамках данной работы основной реализацией является именно список находок как компактная и практически применимая форма структурированного условия.

Таким образом, во второй главе предложенный метод условной генерации изображений со структурированным текстовым условием рассмотрен на примере медицинских данных как единая последовательность преобразований от исходного текстового описания до использования сформированного структурированного условия в латентной диффузионной модели. В первой части метода секция FINDINGS преобразуется в список клинических находок, содержащий нормализованные сведения о клинически значимых признаках, их наличии, локализации и степени выраженности. Для этого описаны этапы выделения и нормализации сущностей, обработки отрицаний, связывания признаков с анатомическими областями, объединения дубликатов и разрешения конфликтов.

Во второй части метода сформированный список находок приводится к текстовому представлению и используется вместо исходной секции FINDINGS в стандартном механизме условной генерации Stable Diffusion. Для входящих в метод алгоритмов получены аналитические оценки вычислительной и пространственной сложности, а также показано, что переход к структурированному условию не изменяет асимптотическую сложность обучения и генерации Stable Diffusion.

Третья глава посвящена экспериментальной оценке разработанного метода условной генерации изображений со структурированным заданием условия на медицинских данных. Для медицинских изображений наиболее непосредственной формой проверки результата является экспертная оценка врачами-специалистами, позволяющая определить не только визуальную реалистичность синтетического изображения, но и корректность воспроизведения клинически значимых признаков и их соответствие заданному условию. Вместе с тем проведение полномасштабной экспертной оценки связано со значительными временными и организационными затратами, необходимостью привлечения квалифицированных врачей и формирования достаточно представительной выборки исследований для анализа.

В связи с этим в настоящей работе экспериментальная валидация выполняется посредством совокупности взаимодополняющих количественных и прикладных экспериментов. Основная экспериментальная постановка строится как сравнение двух вариантов text-to-image генерации: базовой модели, обученной на парах «CXR-изображение - секция FINDINGS», и предлагаемой модели, обученной на парах «CXR-изображение - список клинических находок».

Для их сопоставления используются стандартные количественные метрики генеративных моделей и downstream-оценка практической полезности синтетических изображений. Такие вычислительные оценки не являются полной заменой клинической экспертизы, однако позволяют воспроизводимо сравнить два способа задания условия и оценить отдельные свойства получаемых синтетических данных.

Дополнительно исследуются свойства синтетических данных, существенные для их практического использования: сходство с обучающими изображениями, возможность определения принадлежности реального изображения к обучающей выборке генеративной модели и риск повторной идентификации пациента. Выделение этих экспериментов в отдельный блок обусловлено известной для генеративных моделей возможностью запоминания элементов обучающей выборки и наличием в медицинских изображениях индивидуальных признаков пациента. Наконец, отдельный эксперимент на маммографических изображениях используется для оценки переносимости принципа структурированного задания условия на другой тип медицинских данных.

В начале главы описываются используемые данные и процедура их подготовки. В качестве экспериментальной базы используется открытый набор данных MIMIC-CXR, содержащий рентгенограммы органов грудной клетки в формате DICOM и соответствующие им радиологические отчеты. Изображения извлекаются из DICOM-файлов, нормализуются и приводятся к формату, пригодному для обучения модели латентной диффузии. Основные характеристики экспериментального набора данных приведены в Таблице 3. В качестве исходного текстового условия используется секция FINDINGS, содержащая описание визуально наблюдаемых признаков и патологических изменений. На этапе подготовки данных из полного радиологического отчета автоматически выделяется секция FINDINGS с помощью разработанной процедуры предобработки текстовых отчетов. Данный этап входит в состав общего метода подготовки текстового условия для обучения модели и обеспечивает переход от полного радиологического отчета к стандартизованному текстовому представлению, используемому далее для формирования как базового, так и структурированного условия. Для предлагаемого варианта секция FINDINGS далее преобразуется в список клинических находок по алгоритму, описанному во второй главе. В результате для каждого изображения формируются два типа условий: исходное текстовое условие и структурированное условие.

Таблица 3 - Характеристики экспериментального набора данных

Характеристика	Значение
Тип изображений	Рентгенограммы органов грудной клетки
Формат исходных изображений	DICOM
Число исследований	227 835
Число изображений	377 110
Обучающая выборка	368 960
Валидационная выборка	2 991
Тестовая выборка	5 159

Далее описывается общий экспериментальный протокол. В обоих вариантах используется Stable Diffusion 1.5; различие между постановками заключается в форме текстового условия. В базовом варианте модель обучается на парах «CXR-изображение - секция FINDINGS», а в предлагаемом варианте - на парах «CXR-изображение - список находок». Такая постановка позволяет напрямую оценить вклад структурирования условия, не смешивая его с влиянием архитектурных изменений или изменения обучающего набора изображений. Общая схема экспериментального протокола представлена на Рисунке 4.

Схема экспериментального протокола



Рисунок 4 - Схема экспериментального протокола сравнения двух вариантов условной генерации. Полный радиологический отчет преобразуется в секцию FINDINGS; далее формируются две ветви обучения: базовая модель использует FINDINGS, а предлагаемая модель - список клинических находок. Полученные модели сравниваются по стандартным метрикам генерации и по downstream-задачам.

Экспериментальная оценка включает несколько взаимодополняющих групп экспериментов, каждая из которых характеризует отдельное свойство предложенного метода и получаемых с его помощью синтетических данных. Первая группа основана на стандартных количественных метриках MSE, FID и KID и предназначена для воспроизводимого сопоставления двух вариантов генерации по общим статистическим характеристикам синтезируемых изображений. Следует отметить, что такие метрики не являются прямой мерой клинической достоверности и не позволяют полностью заменить экспертную оценку врачами, однако широко применяются для количественного сравнения генеративных моделей и позволяют исключить субъективность при сопоставлении рассматриваемых вариантов.

Вторая группа экспериментов представляет собой downstream-оценку практической полезности синтетических данных. В этой постановке проверяется не только внешнее сходство синтетических изображений с реальными, но и сохранение в них признаков, пригодных для обучения прикладной модели анализа медицинских изображений. Такой тип оценки используется в исследованиях генеративных моделей медицинских изображений как функциональный критерий полезности синтетических данных.

Третья группа экспериментов связана с оценкой конфиденциальности и включает поиск ближайших обучающих изображений, оценку возможности определения принадлежности

реального изображения к обучающей выборке генеративной модели и оценку риска повторной идентификации пациента. Эти эксперименты характеризуют безопасность практического использования сформированных синтетических наборов. Отдельный экспериментальный блок на маммографических изображениях предназначен для проверки переносимости принципа структурированного задания условия на другой тип медицинских данных. Таким образом, выбранный протокол позволяет рассматривать предложенный метод с точки зрения качества генерации, практической полезности, рисков конфиденциальности и переносимости.

Сначала проводится количественное сравнение базовой и предлагаемой моделей по стандартным метрикам MSE, FID и KID. Данный эксперимент используется как воспроизводимая вычислительная оценка общих характеристик синтезируемых изображений и не рассматривается как самостоятельная оценка их клинической достоверности. В рамках эксперимента обе модели оцениваются на одной и той же тестовой выборке, что позволяет сопоставить влияние формы текстового условия при неизменных остальных компонентах экспериментальной постановки. Результаты сравнения представлены в Таблице 4.

Таблица 4 - Сравнение моделей по стандартным метрикам качества генерации

Метод	MSE	FID	KID
FINDINGS	0.0901	64.156	0.0185
Список находок	0.0921	61.245	0.0159

Результаты количественного сравнения показывают преимущество структурированного условия по метрикам FID и KID при близких значениях MSE для двух вариантов модели. Полученный результат свидетельствует о том, что переход от исходной секции FINDINGS к списку клинических находок не приводит к ухудшению общих статистических характеристик генерации и обеспечивает улучшение показателей, характеризующих близость распределений реальных и синтетических изображений.

Следующая группа экспериментов направлена на downstream-оценку их практической полезности. Подобная постановка применяется в исследованиях синтетических медицинских данных для функциональной проверки генеративных моделей: если изображения содержат устойчивые признаки, связанные с требуемыми клиническими состояниями, модель, обученная на таких данных, должна сохранять способность решать соответствующую задачу на независимых реальных изображениях. Для этого с помощью каждой из сравниваемых генеративных моделей формируется синтетический набор изображений по заданному набору клинических условий. Затем на каждом синтетическом наборе обучается отдельная классификационная модель, а качество классификаторов оценивается на независимом наборе реальных CXR-изображений с разметкой. Такая постановка позволяет косвенно оценить сохранение в синтетических данных клинически значимой информации и сопоставить практическую полезность изображений, полученных двумя вариантами метода. Результаты сравнения представлены в Таблице 5.

Для классификатора, обученного на изображениях, сформированных со структурированным условием, получены несколько более высокие значения всех трех метрик. Результат показывает небольшое, но последовательное преимущество структурированного условия и подтверждает сохранение практической полезности синтетических данных. Тем

самым подтверждается, что структурированное задание условия не только улучшает интерпретируемость условия генерации, но и сохраняет практическую ценность синтетических изображений для последующего применения в задачах анализа медицинских данных.

Таблица 5 - Downstream-оценка полезности синтетических данных на датасете CheXpert

Метод	Macro ROC AUC	Macro PR AUC	Macro F1
FINDINGS	0.6205	0.2145	0.1507
Список находок	0.6252	0.2172	0.1637

Дополнительно проведена проверка наличия среди синтетических данных практически точных копий обучающих изображений. Для каждого синтетического изображения выполнялся поиск ближайшего изображения из соответствующей обучающей выборки в пространстве признаков представлений. В качестве контрольного сравнения аналогичный поиск проводился для реальных изображений тестовой выборки, не использованных при обучении генеративной модели. Таким образом, сопоставлялись направления «синтетическое изображение → ближайшее обучающее изображение» и «реальное тестовое изображение → ближайшее обучающее изображение». Результаты представлены в Таблице 6. Примеры полученных распределений представлены на Рисунке 5.

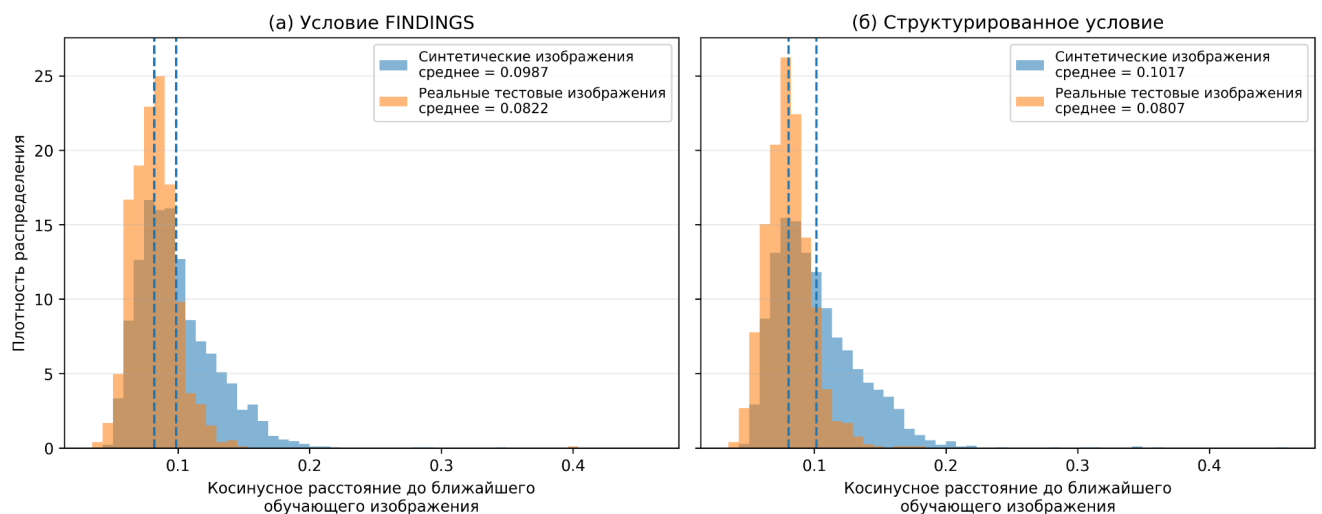


Рисунок 5 - Распределения косинусного расстояния до ближайшего обучающего изображения для синтетических и реальных тестовых изображений при использовании условия FINDINGS (а) и структурированного списка клинических находок (б).

При обоих вариантах условия распределение для синтетических изображений было смещено в сторону больших расстояний по сравнению с распределением для независимых реальных изображений. В рамках использованной подвыборки и выбранного признакового представления признаки прямого воспроизведения отдельных обучающих изображений не выявлены.

В следующем эксперименте проверяется, можно ли по близости к синтетическим изображениям определить, входило ли конкретное реальное изображение в обучающую выборку генеративной модели. Для каждого реального изображения из обучающей и тестовой

выборка находится ближайшее синтетическое изображение. Сопоставляются два направления поиска: «реальное обучающее изображение → ближайшее синтетическое изображение» и «реальное тестовое изображение → ближайшее синтетическое изображение». Расстояние до ближайшего синтетического изображения используется в качестве оценки принадлежности реального изображения к обучающей выборке, после чего рассчитывается площадь под ROC-кривой. Результаты оценки возможности определения принадлежности изображений к обучающей выборке представлены в Таблице 7.

Таблица 6 - Результаты поиска ближайших обучающих изображений.

Метод	Косинусное расстояние, synthetic / test	SSIM, synthetic / test
FINDINGS	0.0987 / 0.0822	0.445 / 0.509
Список находок	0.1017 / 0.0807	0.436 / 0.508

Распределения расстояний для обучающих и тестовых изображений оказались практически одинаковыми, а значения AUC были близки к 0.5. Следовательно, использованные меры близости не позволяют надежно отличить изображения, входившие в обучающую выборку генеративной модели, от независимых тестовых изображений. Полученный результат не выявляет выраженного следа обучающей выборки в сформированном синтетическом наборе.

Таблица 7 - Результаты оценки возможности определения принадлежности изображений к обучающей выборке по различным мерам близости. В столбцах «обучающая / тестовая» последовательно приведены средние значения меры близости для реальных изображений обучающей и тестовой выборок. ROC AUC характеризует возможность определения принадлежности изображения к обучающей выборке по соответствующей мере близости к синтетическим данным.

Условие генерации	Косинусное расстояние		SSIM		MSE	
	Среднее обучающая / тестовая	ROC AUC	Среднее обучающая / тестовая	ROC AUC	Среднее обучающая / тестовая	ROC AUC
FINDINGS	0.1011 / 0.1024	0.5149	0.4900 / 0.4897	0.5002	0.0621 / 0.0625	0.5082
Список находок	0.1043 / 0.1020	0.4641	0.4874 / 0.4888	0.4924	0.0625 / 0.0638	0.5068

Отдельный эксперимент посвящен оценке риска повторной идентификации пациента по синтетическим CXR-изображениям. Такая проверка имеет самостоятельное значение для медицинских синтетических данных, поскольку в литературе показано, что рентгенограммы органов грудной клетки содержат устойчивые анатомические признаки, позволяющие алгоритмически сопоставлять изображения одного пациента. Следовательно, отсутствие прямого копирования обучающего изображения само по себе не гарантирует отсутствия переноса индивидуальных особенностей пациента в синтетические данные.

Для проведения эксперимента использовалась специализированная модель повторной идентификации на основе ResNet50, обученная методом контрастивного обучения с учителем и формирующая признаковые представления рентгенограмм. Работоспособность модели предварительно проверялась в постановке поиска изображений одного пациента среди реальных данных, результаты которой подтвердили наличие в рентгенограммах устойчивых признаков, позволяющих сопоставлять исследования одного пациента.

Далее в качестве запросов использовались изображения, сгенерированные по секции FINDINGS и по структурированному списку клинических находок. Для каждого синтетического изображения выполнялся поиск в реальной галерее и проверялось, присутствуют ли изображения исходного пациента среди первых результатов выдачи. Для каждого варианта условия оценивались синтетические изображения, соответствующие исходным пациентам. Результаты поиска представлены в Таблице 8.

При использовании синтетических изображений показатели повторной идентификации оказались существенно ниже, чем в контрольной постановке поиска между реальными изображениями. Для структурированного условия значения Top-1, Top-5 и mAP были ниже, чем для условия FINDINGS. Парный бутстрэп-анализ показал устойчивость наблюдаемой тенденции к снижению показателей повторной идентификации. Таким образом, использование структурированного условия связано с дополнительным ослаблением выраженности признаков, позволяющих сопоставить синтетическое изображение с исходным пациентом.

Для проверки переносимости предложенного подхода на другой тип медицинских изображений проведен дополнительный эксперимент на наборе VinDr-Mammo. С использованием диффузионной модели сформированы синтетические наборы маммографических изображений различного объема. Генерация выполнялась по структурированному условию, включавшему проекцию исследования, сторону, категорию плотности молочной железы, оценку BI-RADS и типы выявленных находок. В отличие от экспериментов на рентгенограммах органов грудной клетки, условие генерации формировалось непосредственно из набора клинических и технических атрибутов маммографического исследования.

Таблица 8 - Результаты оценки риска повторной идентификации пациента.

Тип запроса	Top-1	Top-5	mAP
Реальное изображение → реальная галерея	0.7700	0.9110	0.7040
Синтетическое изображение, FINDINGS → реальная галерея	0.0135	0.0296	0.0155
Синтетическое изображение, список находок → реальная галерея	0.0091	0.0236	0.0128

Для оценки практической полезности синтетических изображений сравнивались классификаторы, обученные только на реальных данных, на смешанных наборах с объемом синтетической части 25%, 50% и 100% относительно обучающей выборки диффузионной модели, а также только на синтетических данных. Все модели оценивались на одной и той же независимой реальной тестовой выборке. Сравнение проводилось по метрикам ROC AUC, Accuracy, Precision, Recall и Specificity. Результаты представлены в Таблице 9.

Результаты показали немонотонную зависимость качества классификации от объема синтетической части. Добавление синтетических данных в объеме 25% и 50% не обеспечило улучшения относительно модели, обученной только на реальных изображениях. При равном объеме реальных и синтетических данных были получены наибольшие значения ROC AUC, Accuracy, Precision и Specificity, однако Recall оказался ниже базового уровня. Обучение только на синтетических изображениях привело к существенному снижению всех основных показателей, что, однако, показывает возможность обучения моделей только на синтетических данных, хотя полученное качество существенно уступает результатам на реальных данных. Таким образом, синтетические маммограммы могут быть полезны как дополнение к реальной обучающей выборке при соответствующем выборе их объема, но не являются полноценной заменой реальных данных. Эксперимент подтверждает переносимость принципа структурированного задания условия на другой тип медицинских изображений и возможность применения полученных данных для расширения обучающих выборок.

В завершение главы обсуждаются ограничения проведенной экспериментальной оценки. Наиболее полная проверка медицинской достоверности синтетических изображений должна включать экспертную оценку врачами-специалистами, способными непосредственно определить анатомическую правдоподобность изображения, корректность отображения патологических изменений и соответствие результата заданному клиническому условию. Проведение такой оценки в объеме, достаточном для статистически обоснованного сравнения моделей, требует привлечения квалифицированных специалистов и значительных временных и организационных ресурсов. Поэтому в настоящей работе сравнение моделей основано на совокупности количественных метрик и downstream-оценки. Эти методы позволяют воспроизводимо оценить свойства синтетических изображений.

Таблица 9 - Результаты классификации на наборе VinDr-Mammo при различных составах обучающей выборки

Обучающая выборка	Размер синтетической части	ROC AUC	Accuracy	Precision	Recall	Specificity
Реальные	0%	0.9406	0.9918	0.7	0.6087	0.9967
Реальные и синтетические	25%	0.933	0.9896	0.6	0.5217	0.9956
Реальные и синтетические	50%	0.9252	0.9904	0.6222	0.6087	0.9953
Реальные и синтетические	100%	0.9494	0.9921	0.7742	0.5217	0.9981
Синтетические	100%	0.8138	0.9738	0.1429	0.2174	0.9834

Дополнительные ограничения связаны с тем, что итоговое качество генерации зависит от корректности промежуточных этапов подготовки данных, включая выделение секции FINDINGS и формирование списка находок. Оценки конфиденциальности основаны на выбранных признаковых представлениях, моделях атак и алгоритмах поиска ближайших изображений, поэтому отсутствие выявленного риска не является доказательством его отсутствия для всех возможных способов анализа. Эксперимент на маммографических данных дополнительно ограничен выраженным дисбалансом классов и зависимостью результата от выбранного соотношения реальных и синтетических изображений. Для уменьшения влияния данного фактора сравнение проводилось для нескольких вариантов балансировки обучающей выборки, а итоговая оценка выполнялась на единой независимой реальной тестовой выборке.

Несмотря на указанные ограничения, проведенный комплекс экспериментов позволил оценить предложенный метод на нескольких взаимодополняющих уровнях. Количественное сравнение показало улучшение FID и KID при использовании структурированного условия, а downstream-эксперимент подтвердил сохранение практической полезности синтетических данных и показал небольшое повышение показателей классификатора, обученного на изображениях, сформированных по списку клинических находок. В совокупности эти результаты позволяют воспроизводимо сопоставить структурированное и неструктурированное задание условия.

Дополнительная группа экспериментов позволила охарактеризовать риски, связанные с практическим использованием синтетических медицинских данных. Анализ ближайших изображений не выявил практически точного воспроизведения обучающих примеров, а использованные меры близости не позволили надежно отличить изображения обучающей выборки от независимых тестовых изображений. Эксперимент по повторной идентификации показал низкие значения соответствующих метрик для синтетических данных и дополнительное снижение этих показателей при использовании структурированного условия.

Наконец, эксперимент на маммографических изображениях подтвердил переносимость принципа структурированного задания условия на другой класс медицинских изображений. Синтетические данные показали возможность использования в качестве дополнения к реальным обучающим выборкам, при этом их эффективность зависела от соотношения реальной и синтетической частей. Таким образом, третья глава позволяет охарактеризовать предложенный метод не только по количественным показателям генерации, но также с точки зрения практической полезности получаемых синтетических данных, рисков конфиденциальности и переносимости подхода на другой тип медицинских изображений.

В **заключении** приведены основные результаты диссертационной работы. В работе решена задача разработки метода условной генерации изображений на основе диффузионных моделей со структурированным заданием условия, формируемым из текстового описания. Разработка и экспериментальная валидация метода выполнены на примере медицинских данных. Основные результаты работы заключаются в следующем:

1. Разработан метод формирования структурированного условия для диффузионной генерации изображений на основе разложения неструктурированного текстового описания на семантически значимые компоненты, исследованный на примере медицинских данных.
2. Разработан алгоритм формирования структурированного условия, реализованный для медицинских данных в виде формирования списка клинических находок из радиологического отчета, и его формальное представление, пригодное для использования в text-to-image модели.

3. Получены и доказаны теоретические оценки сложности метода: вычислительной сложности формирования списка находок, пространственной сложности его хранения и неизменности асимптотической сложности обучения и генерации Stable Diffusion при замене FINDINGS на список находок.
4. Разработан программный конвейер подготовки условий для генерации, включающий выделение секции FINDINGS, формирование списка находок и подготовку пар «изображение - структурированное условие».
5. Проведена экспериментальная валидация подхода на медицинских данных. Использование структурированного условия обеспечило снижение FID на **4.5%**, KID - на **14.1%** и повышение Macro F1 в downstream-задаче на **8.6%**. Дополнительные эксперименты не выявили признаков прямого воспроизведения обучающих изображений; значения ROC AUC определения принадлежности находились в диапазоне близком к случайному уровню **0.5**, а показатели повторной идентификации снизились более чем на **96%** по сравнению с поиском между реальными изображениями.
6. Проверена переносимость подхода на маммографические изображения набора VinDr-Mammo. При равном объеме реальных и синтетических данных ROC AUC увеличился с **0.9406 до 0.9494**, а Precision - с **0.7000 до 0.7742**, что подтверждает возможность использования синтетических данных для расширения обучающих выборок.

Публикации автора в журналах по теме диссертации

1. Ushakov E. et al. Endonet: A model for the automatic calculation of H-score on histological slides //Informatics. – MDPI, 2023. – Т. 10. – №. 4. – С. 90.
2. Litvinov A. et al. Clinically aware learning: Ordinal loss improves medical image classifiers //Journal of Clinical Medicine. – 2026. – Т. 15. – №. 1. – С. 365.
3. Shodiev D. et al. Privacy-preserving synthetic mammograms: a generative model approach to privacy-preserving breast imaging datasets //Informatics. – MDPI, 2025. – Т. 12. – №. 4. – С. 112.
4. Ushakov E. et al. Optimizing the Mammography AI Pipeline: From Data Filtering to Vision-Language Models //Informatics. – MDPI, 2026. – Т. 13. – №. 8. – С. 125.

Публикации автора на конференциях по теме диссертации

5. Ibragimov A. et al. MamT 4: Multi-view Attention Networks for Mammography Cancer Classification //2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC). – IEEE, 2024. – С. 1965-1970.

Ушаков Егор Николаевич

Метод условной генерации изображений в диффузионных моделях со
структурированным текстовым условием

Автореф. дис. на соискание ученой степени канд. физико-математических наук

Подписано в печать __.__.____. Заказ № _____

Формат 60×90 / 16. Усл. печ. л. 1. Тираж 100 экз.

Типография _____