

**РЕЗАИАН НАИМ**

**МЕТОДИКА СЕМАНТИЧЕСКОГО АНАЛИЗА МУЛЬТИМОДАЛЬНЫХ  
БОЛЬШИХ ДАННЫХ НА ОСНОВЕ ПРИМЕНЕНИЯ МЕТОДОВ МАШИННОГО  
ОБУЧЕНИЯ**

Научная специальность

2.3.1. Системный анализ, управление и обработка информации, статистика

**Автореферат диссертации**  
на соискание ученой степени кандидата технических наук

Работа выполнена в инженерной академии ФГАОУ ВО «Российский университет дружбы народов имени Патриса Лумумбы» (РУДН)

Научный руководитель: Разумный Юрий Николаевич  
доктор технических наук, профессор, заведующий кафедрой  
механики и процессов управления инженерной академии  
РУДН

Официальные оппоненты: Игнатьев Александр Владимирович  
доктор технических наук, доцент, профессор кафедры  
«Цифровые технологии в урбанистике, архитектуре и  
строительстве» института архитектуры и строительства  
ФГБОУ ВО «Волгоградский государственный технический  
университет»

Клюшников Валерий Юрьевич  
доктор технических наук, главный ученый секретарь АО  
«Центральный научно-исследовательский институт  
машиностроения»

Петров Андрей Владимирович  
кандидат технических наук, советник по научным  
исследованиям и разработкам ООО «МОНИТОР СОФТ»

Защита диссертации состоится 29.12.2025 в 13:00 на заседании диссертационного совета ПДС 2022.016 по адресу: 115419, г. Москва, ул. Орджоникидзе, д. 3

С диссертацией можно ознакомиться в научной библиотеке РУДН по адресу: 117198, г. Москва, ул. Миклухо-Маклая, д. 6

Объявление о защите и автореферат диссертации размещены на сайтах  
<http://vak2.ed.gov.ru/> и <https://www.rudn.ru/science/dissovet>

Автореферат разослан 27.11.2025.

Ученый секретарь диссертационного совета ПДС 2022.016  
кандидат технических наук

О.Е. Самусенко

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

### **Актуальность темы исследования**

Стремительный рост объемов и разнообразия данных делает одной из центральных проблем современного искусственного интеллекта обеспечение эффективной интеграции, анализа и интерпретации информации из разных источников: изображений, звуковых сигналов, текстов, временных рядов и других типов данных. Эти источники образуют мультимодальные большие данные, характеризующиеся высокой размерностью, гетерогенностью, наличием шумов, пропусков и артефактов, что приводит к искажению признаков и снижению качества принимаемых решений.

Традиционные матричные модели и классические алгоритмы машинного обучения плохо масштабируются в условиях «проклятия размерности», не обеспечивают сохранение межмодальных связей и недостаточно устойчивы к шумам и неполноте данных. В отличие от традиционных матричных представлений, тензорные структуры позволяют описывать данные в многомерных пространствах, выявляя латентные зависимости и семантические закономерности. Это делает их эффективным инструментом для задач слияния мультимодальной информации, что особенно актуально в областях компьютерного зрения, обработки речи, биомедицинских исследований и интеллектуальных систем мониторинга.

Одновременно с задачей интеграции данных сохраняет актуальность проблема устранения шумов и восстановления информации в многомерных массивах. С ростом сложности и размерности данных увеличивается доля случайных искажений, обусловленных аппаратными и методологическими факторами. Эти искажения существенно влияют на точность обучения моделей и достоверность принимаемых решений. Поэтому создание методов, способных эффективно выделять полезные сигналы на фоне шумов и корректно восстанавливать утраченные фрагменты информации, имеет не только теоретическую, но и прикладную значимость.

Это определяет актуальность разработки методики семантического анализа мультимодальных больших данных, основанной на тензорных представлениях и современных методах машинного обучения, обеспечивающей одновременное слияние модальностей, уменьшение размерности, подавление шумов и восстановление недостающей информации.

### **Степень разработанности темы исследования**

Исследования, посвященные вопросам интеграции и анализа мультимодальных данных, имеют значительную историю и в последние годы получили активное развитие благодаря расширению вычислительных возможностей. В диссертации проведен подробный обзор современных подходов к работе с мультимодальными данными, включая классические и нейросетевые методы слияния признаков, а также тензорные модели представления. Рассмотрены классические схемы Data Fusion, применяемые при интеграции гетерогенных данных дистанционного зондирования и геоинформатики (Du et al., 2021; Феоктистов и др., 2017; Бондур и др., 2023), а также тензорный слой слияния и основанные на нем архитектуры Tensor Fusion Network и ее модификации для анализа эмоций и настроений (Zadeh et al., 2017). Особое внимание удалено низкоранговым методам тензорного объединения признаков (Low-rank Multimodal Fusion, LMF и др.), позволяющим существенно снизить параметрическую сложность моделей за счет разложения веса тензорного слоя на набор модально-специфичных факторов (Liu et al., 2018). Проанализированы нелинейные полиномиальные модели слияния (High-Order Polynomial Fusion, полиномиальное тензорное объединение, HPFN), обеспечивающие учет взаимодействий высоких порядков между модальностями при умеренном росте числа параметров (Hou et al., 2019). Рассмотрены attention- и gating-архитектуры (Gated Multimodal Unit, GMU; Gated Multimodal Embedding LSTM with Temporal Attention, GME-LSTM(A) и др.), реализующие адаптивное взвешивание вкладов модальностей и подавление шумовых каналов (Arevalo et al., 2017; Chen et al., 2018), а также трансформерные модели для последовательностей мультимодальных признаков, в том числе Multimodal Transformer (MuIT) и High-Modality Multimodal Transformer (HighMMT), ориентированные на моделирование сложных временных и межмодальных зависимостей (Tsai et al.,

2019; Liang et al., 2022). Отдельное внимание уделено динамическим схемам слияния, в которых структура взаимодействий между модальностями адаптируется к контексту входных данных (Dynamic Multimodal Fusion, DynMM; Recurrent Attended Variation Embedding Network, RAVEN и др.) (Wang et al., 2019; Xue et al., 2023), а также методам оптимизации мультимодальных остатков (Multimodal Residual Optimization, MRO), позволяющим явно декомпозировать унимодальные, бимодальные и тримодальные вклады (Wörtwein et al., 2022). Наконец, проанализированы подходы на основе мультимодальных автоэнкодеров и вариационных автоэнкодеров, направленные на совместное обучение общих и модально-специфичных представлений и реконструкцию пропусков в данных (Gao et al., 2020; Xu et al., 2022).

Анализ показал, что, несмотря на существенный прогресс, большинство существующих подходов ориентированы преимущественно на повышение точности в рамках конкретных прикладных задач (распознавание эмоций, анализ настроений и т.п.) и в меньшей степени направлены на выявление скрытых взаимосвязей между модальностями. При этом мало внимания уделяется уровню шума и задаче восстановления исходных модальностей: мультимодальные данные, как правило, содержат значительные помехи и пропуски, и выполнение слияния без предварительного удаления шума и восполнения недостающей информации существенно ограничивает возможность корректного выявления межмодальных зависимостей. Использование тензорного слоя слияния сопровождается эффектом «проклятия размерности», когда экспоненциальный рост числа параметров делает обучение таких моделей практически недоступным на реальных вычислительных ресурсах. Методы низкорангового слияния, напротив, обычно предполагают фиксированный ранг для всех модальностей, что не позволяет учитывать их индивидуальные характеристики: при малых рангах теряются важные взаимосвязи между модальностями, а при больших – резко возрастает число параметров и сложность модели. В совокупности это свидетельствует о недостаточной разработанности комплексных решений, которые одновременно обеспечивали бы семантически осмысленное слияние мультимодальных представлений, подавление шумов, восстановление исходной информации и оценку качества реконструкции в единой тензорной модели, что и обуславливает выбор в диссертационной работе подхода, основанного на тензорных разложениях с end-to-end-обучением и рандомизированными алгоритмами тензорного восстановления.

**Объектом исследования** является система интеграции и представления мультимодальных данных на основе тензорного разложения, функционирующая в условиях многомерности, избыточности и зашумленности исходных данных.

**Предмет исследования** – тензорные модели представления и алгоритмы обработки многомерных и мультимодальных данных, направленные на повышение достоверности, устойчивости и эффективности анализа информации.

**Цель работы** заключается в разработке нового вычислительного метода для решения проблемы обработки мультимодальных больших данных с представлением данных в формате тензора и также нового подхода для удаления шума и восстановления информации в мультимодальных данных для устойчивости работы с нейронными сетями.

Для достижения поставленной цели решены следующие научно-технические задачи:

- Выполнен анализ предметных областей, в которых используются мультимодальные большие данные, и выявлены типичные сценарии их обработки.
- Исследованы возможности и ограничения существующих систем обработки мультимодальных данных.
- Проанализированы современные алгоритмы интеграции многомерных данных на уровне признаков и на уровне принятия решений.
- Разработан и обоснован метод интеграции, удаления шума и восстановления мультимодальных данных на основе мультирангового тензорного разложения.
- Проведены вычислительные эксперименты, подтверждающие эффективность предложенных подходов.

- Разработана программная библиотека на языке Python, реализующая предложенную методику.

**Методология и методы исследования.** Для решения поставленных в работе задач использовались методы линейной и мультилинейной алгебры, математической статистики, а также алгоритмы машинного обучения и глубокого обучения. Основу математического аппарата составляют операции над тензорами и различные виды их разложения, включая CPD, разложение Таттера, Tensor Train и Tensor Ring. Для повышения устойчивости и достоверности анализа применялись методы регуляризации, рангового ограничения и низкорангового восстановления данных. Для реализации и экспериментальной проверки предложенных решений разработано программное обеспечение на языках Python и Matlab. Реализованные модули обеспечивают выполнение тензорных операций, моделирование алгоритмов интеграции и восстановления мультимодальных данных, а также визуализацию полученных результатов.

**Научная новизна** работы заключается в следующем:

- Разработана структурная модель интеграции мультимодальных данных на основе тензорного представления, в которой различные модальности объединяются в едином многомерном пространстве признаков. Предложенный подход обеспечивает сохранение межмодальных связей и позволяет повысить информативность объединенных данных при анализе сложных объектов.
- Предложен метод построения тензорных представлений с использованием декомпозиции по низкоранговым компонентам (CPD, Tucker, TT), обеспечивающий уменьшение размерности данных без потери значимых признаков. В отличие от традиционных подходов, метод позволяет учитывать корреляционные зависимости между модальностями.
- Разработан алгоритм удаления шумов и восстановления недостающих элементов в мультимодальных данных с использованием тензорного формализма и адаптивных регуляризаторов. Алгоритм сочетает принципы низкорангового восстановления и адаптивной фильтрации, что обеспечивает повышение достоверности реконструированных данных и их пригодности для последующего анализа.
- Разработана обобщенная архитектура интеллектуальной системы анализа мультимодальных данных, включающая модули тензорного слияния, подавления шумов и реконструкции пропусков. Архитектура обеспечивает масштабируемость и возможность адаптации под различные типы данных и прикладные задачи (видеонаблюдение, биомедицинская диагностика, распознавание эмоций и др.).
- Проведен вычислительный эксперимент на реальных и синтетических наборах мультимодальных данных, доказавший преимущество предложенных методов по сравнению с существующими по критериям точности, устойчивости и вычислительной эффективности.

**Практическая значимость полученных результатов** заключается в разработке вычислительных методов и программных средств для эффективной интеграции и анализа мультимодальных данных. Формализовано мультиранговое слияние, обеспечивающее объединение разнородных источников информации с сохранением взаимных зависимостей между модальностями и повышением точности и устойчивости анализа при работе с реальными данными. Использование факторизованных тензорных представлений позволяет избежать явного формирования высокоразмерных тензоров признаков и, тем самым, уменьшить вычислительные затраты. Разработанные методы удаления шумов и восстановления недостающих элементов данных делают возможным применение предложенной методики при обработке неполных и зашумленных мультимодальных выборок. В рамках работы создана программная библиотека на языке Python, реализующая предложенные алгоритмы мультирангового тензорного слияния и низкорангового восстановления, что облегчает их интеграцию в существующие программные комплексы интеллектуального анализа данных.

**Теоретическая значимость** результатов заключается в развитии методов тензорного анализа и моделирования мультимодальных данных, направленных на решение фундаментальных проблем представления, интеграции, удаления шума и восстановления информации.

## Положения, выносимые на защиту

- Структурная модель интеграции мультимодальных данных на основе тензорного представления, обеспечивающая сохранение семантических связей между модальностями и уменьшение размерности признакового пространства.
- Метод мультирангового тензорного слияния, позволяющий эффективно интегрировать мультимодальные данные и решать задачу «проклятия размерности» в задачах машинного обучения.
- Алгоритм подавления шума и восстановления недостающих элементов в тензорных моделях данных, основанный на минимизации тензорного ранга и аддитивной регуляризации.
- Методика количественной оценки качества тензорного слияния и восстановления на основе совокупности метрик согласованности модальностей, ранговых характеристик и точности реконструкции.
- Результаты вычислительных экспериментов, демонстрирующие преимущество предложенных алгоритмов по сравнению с существующими методами по точности, устойчивости и вычислительной эффективности.

## Апробация работы

Основные концепции и результаты исследований обсуждались и представлялись на следующих международных научных конференциях:

- V International Conference on Information Technologies in Engineering Education (Inforino), 2020.
- Международную конференцию Сбера по искусственному интеллекту "AI Journey 2024", Москва.
- Международную научно-практическую конференцию «Образовательная трансформация в условиях цифровой экономики», организованную Государственным гуманитарно-технологическим университетом (ГГТУ), Московская область, 2025.
- Конференцию «AI-Горизонты», Москва, 2025.

По теме диссертации опубликовано 5 печатных работ в рецензируемых научных изданиях, индексируемых в международных базах Scopus и Web of Science.

## Степень достоверности результатов

Достоверность результатов обеспечивается строгой математической постановкой задач, использованием общепринятых методов мультилинейной алгебры и статистики, а также сопоставлением разработанных алгоритмов с существующими решениями на множестве тестовых примеров. Степень достоверности подтверждается согласованностью теоретических выводов и полученных экспериментальных данных, а также повторяемостью результатов при вариации параметров модели.

## ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

**В первой главе** подробно рассматриваются основные понятия мультимодальных данных и методов их слияния, под которыми понимается совокупность наблюдений, описывающих один и тот же объект или явление через несколько принципиально различных способов представления информации. Каждое такое представление называется модальностью и соответствует определенному «каналу» восприятия или измерения.

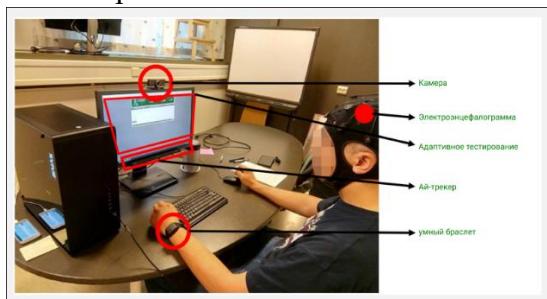


Рисунок 1. Экспериментальная установка – участник подключен ко всем устройствам сбора данных

Существенной характеристикой мультимодальных данных является их гетерогенность, причем различия между модальностями носят многомерный характер. В диссертации выделяются шесть измерений такой гетерогенности.

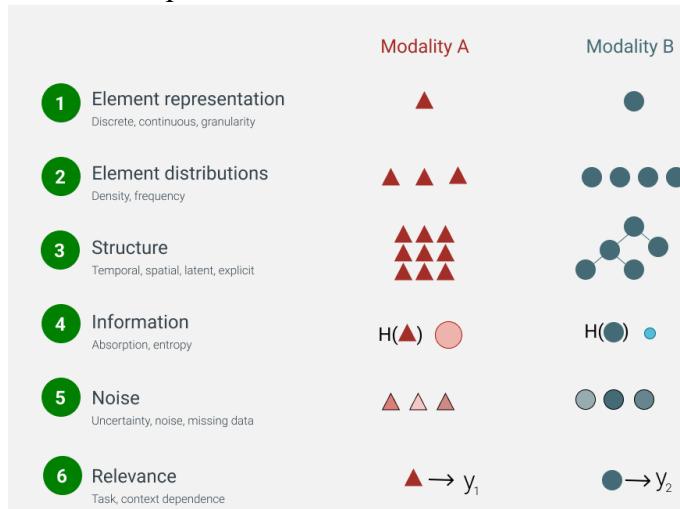


Рисунок 2. Шесть измерений гетерогенности между модальностями

Во-первых, модальности различаются по способу представления элементов: пиксели изображения, спектральные коэффициенты аудиосигнала и токены текста имеют принципиально разную природу и структуру. Во-вторых, различаются статистические распределения данных: плотность вероятности значений в пространстве изображений, как правило, существенно отличается от распределений в пространстве текстовых эмбеддингов или акустических признаков. В-третьих, важна структурная организация: одни модальности обладают ярко выраженной временной структурой (речь, сенсорные временные ряды), другие пространственной (изображения, трехмерные облака точек), третьи графовой или последовательной (текст, сцены с отношениями между объектами). В-четвертых, модальности могут нести различное информационное содержание: одна модальность может быть более информативной относительно целевой задачи, а другая содержать избыточную либо комплементарную информацию. В-пятых, уровни шума и искажений в разных модальностях могут существенно различаться, что необходимо учитывать при совместной обработке. Наконец, в-шестых, модальности различаются по релевантности для решаемой задачи: например, для эмоционализации речи визуальный канал (мимика, жесты) часто оказывается более информативным, чем текстовая транскрипция, тогда как в задачах семантического анализа доминирует именно текст. Важной составляющей теоретической основы является типология связей и взаимодействий между модальностями. На уровне вероятностного моделирования рассматриваются статистические связи: ассоциации, условные зависимости, совместные распределения, соответствие элементов одной модальности элементам другой.

На этой теоретической базе рассматриваются стратегии слияния данных (data fusion), применяемые в современных мультимодальных системах. Классически выделяют раннее слияние (early fusion), позднее слияние (late fusion) и гибридные схемы.

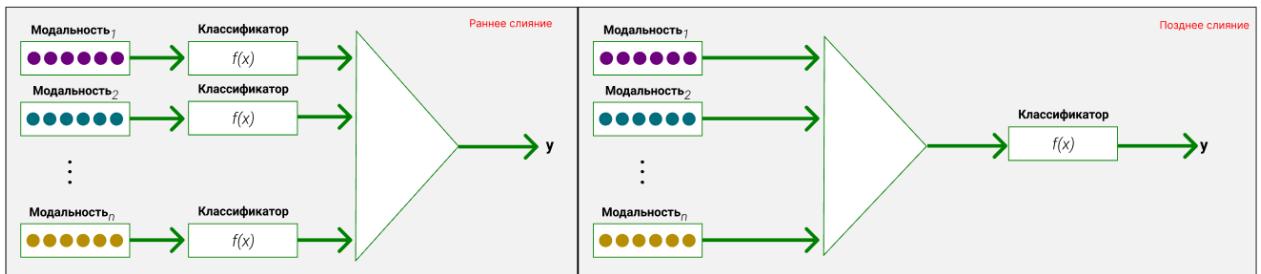


Рисунок 3. Архитектура раннего и позднего слияния

При раннем слиянии признаки, извлеченные из разных модальностей, конкатенируются или объединяются на начальных этапах обработки, после чего единый вектор признаков поступает в общую модель. Преимуществом такого подхода является возможность явно моделировать

межмодальные зависимости и получать представления, учитывающие корреляции между каналами. Однако простая конкатенация приводит к резкому росту размерности признакового пространства, что усиливает проблему «проклятия размерности» и делает модель чувствительной к шуму, дисбалансу и пропущенным данным в отдельных модальностях. При позднем слиянии каждая модальность обрабатывается собственной моделью, выдающей частичное решение (оценку, вероятность, распределение), а объединение происходит на уровне решений, например, путем взвешенного усреднения или мета-классификации. Этот подход более устойчив к отсутствию отдельных модальностей, но хуже использует их тонкие взаимодействия. Гибридные стратегии пытаются совместить достоинства обоих подходов, вводя промежуточные уровни слияния и механизмы адаптивного взвешивания модальностей.

Далее в главе проводится обзор современных методов мультимодального слияния. Существенное внимание уделяется моделям, в первую очередь основанным на тензорных и нейросетевых представлениях. Отдельно рассматриваются подходы, в которых совместное представление модальностей строится с помощью тензорного слоя слияния (Tensor Fusion Network, TFN) и его модификаций, позволяющих явно моделировать попарные и более высокоранговые взаимодействия между признаками различных модальностей. Обсуждаются методы низкорангового тензорного слияния, направленные на уменьшение размерности и числа параметров без потери ключевых межмодальных зависимостей, а также полиномиальные схемы слияния, реализующие нелинейные комбинации признаков.

Завершая теоретический обзор, в главе формулируется выявленный научный разрыв. Анализ литературы показывает, что существующие методы либо слабо учитывают высокоранговые взаимодействия между модальностями, ограничиваясь попарными связями и простыми схемами слияния признаков, либо демонстрируют недостаточную устойчивость к реальным условиям, когда данные содержат шум, выбросы, а отдельные модальности могут частично или полностью отсутствовать. Кроме того, многие модели оказываются слишком ресурсоемкими из-за высокой размерности признаковых тензоров и большого числа параметров, что затрудняет их применение в практических системах реального времени.

**Во второй главе** формализована задача интеграции мультимодальных данных и предложена методика их тензорного представления и обработки. Постановка задачи слияния мультимодальных данных связана с построением модели, способной эффективно объединять информацию, поступающую из различных модальностей, где каждая модальность имеет собственное пространство признаков, размерность и статистические свойства. Цель состоит в построении отображения:

$$\mathcal{F}: \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \dots \times \mathbb{R}^{d_M} \rightarrow \mathbb{R}^h \quad (1)$$

такого, что результирующее представление  $H=F(D_1, D_2, \dots, D_M)$  одновременно сохраняет внутримодальные и межмодальные зависимости, обеспечивая при этом низкую вычислительную сложность и устойчивость к избыточности данных. Для достижения этой цели рассматриваются следующие подходы тензорного разложения.

**Разложение Такера.** Основная идея заключается в аппроксимации исходного многомерного тензора  $X$  с помощью меньшего по размерности тензора ядра  $g$  и набора матриц-факторов, которые описывают линейные подпространства для каждой модальности данных. Задан трехмерный тензор  $X \in \mathbb{R}^{I \times J \times K}$ . Тогда разложение Такера записывается в виде:

$$X \approx g \times_1 B^{(1)} \times_2 B^{(2)} \times_3 B^{(3)} \quad (2)$$

где  $g \in \mathbb{R}^{R_1 \times R_2 \times R_3}$  ядро (core tensor), хранящее взаимодействия между компонентами разных модальностей  $B^{(1)} \in \mathbb{R}^{I \times R_1}$ ,  $B^{(2)} \in \mathbb{R}^{J \times R_2}$ ,  $B^{(3)} \in \mathbb{R}^{K \times R_3}$  матрицы факторов, отображающие исходные признаки каждой модальности в пониженное пространство размерности  $\mathbb{R}_n$ . Для каждого элемента тензора  $X$  аппроксимация записывается как

$$\mathcal{X}_{ijk} \approx \sum_{p=1}^{R_1} \sum_{q=1}^{R_2} \sum_{r=1}^{R_3} \mathcal{G}_{pqr} b_{ip}^{(1)} b_{jq}^{(2)} b_{kr}^{(3)} \quad (3)$$

Мы можем использовать подход разложения Такера для слияния мультимодальных данных и мультимодальное представление выражается как

$$\widehat{H} \approx \mathcal{G} \times_1 B^{(1)} \times_2 B^{(2)} \times_3 \dots \times_M B^{(M)} \quad (4)$$

Полученное латентное представление  $\widehat{H}$  подается в классификатор или регрессионный блок для выполнения целевой задачи

$$\widehat{Y} = \sigma(W_H \widehat{H} + b) \quad (5)$$

Процесс оптимизации параметров направлен на минимизацию функции потерь:

$$\mathcal{L} = \mathcal{L}_{\text{task}}(Y, \widehat{Y}) + \lambda_1 \sum_{i=1}^M \|B^{(i)}\|_F^2 + \lambda_2 \|\mathcal{G}\|_F^2 \quad (6)$$

Одним из ключевых преимуществ Tucker-разложения является резкое снижение числа параметров по сравнению с традиционным построением тензора с помощью внешнего произведения вектора признаков. Вычислительная сложность метода оценивается как

$$\mathcal{O}\left(\sum_{i=1}^M d_i r_i + \prod_{i=1}^M r_i h\right) \quad (7)$$

**Tensor Train разложение.** В предыдущем подходе основная идея заключается в построении многомерного тензора, однако разложение Такера масштабируется плохо при большом числе модальностей, а ядро  $\mathcal{G}$  быстро становится высокоразмерным, что ограничивает применение в более сложных мультимодальных задачах. TT-разложение, предложенное Оседецем (Oseledets, 2011), представляет собой эффективный метод аппроксимации высокомерных тензоров через последовательную цепочку низкоранговых компонент. Для тензора  $X \in \mathbb{R}^{I_1 \times I_2 \times \dots \times I_N}$ , TT-разложение определяется как

$$X(i_1, i_2, \dots, i_N) \approx \sum_{R_0=1}^{R_0} \sum_{R_1=1}^{R_1} \dots \sum_{R_{N-1}=1}^{R_{N-1}} G_1(r_0, i_1, r_1) G_2(r_1, i_2, r_2) \dots G_N(r_{N-1}, i_N, r_N) \quad (8)$$

Структура TT-разложения адаптируется для представления разнородных данных путем последовательной цепочки ядер, где каждое ядро соответствует моде тензора. Ядро Такера может быть заменено TT-представлением, поскольку TT аппроксимирует полный тензор через последовательные матричные произведения, минимизируя параметры и контролируют степень сжатия и выразительность, обеспечивая баланс между точностью и сложностью. Число параметров в TT-формате составляет  $\mathcal{O}(d_y R_1 + \sum_{m=1}^M (d_m + 1) R_m R_{m+1})$ . TT обеспечивает гибкость через независимые  $R_k$ , лучше адаптируясь к разнородным модальностям. TT - разложение обеспечивает оптимальный компромисс между точностью и вычислительной эффективностью, что делает его предпочтительным для систем с большим числом модальностей и высокоразмерными признаками. Однако TT имеет линейную структуру, что может ограничивать захват циклических зависимостей между модами.

**Tensor Ring разложение.** TT-разложение аппроксимирует тензор  $X$  в виде последовательной цепочки ядер  $X(i_1, \dots, i_N) = G_1[i_1] G_2[i_2] \dots G_N[i_N]$ , где каждое ядро  $G_N \in \mathbb{R}^{r_{N-1} \times I_N \times r_N}$  с граничными условиями  $R_0 = R_N = 1$ . TR-разложение снимает эти условия, позволяя  $R_0 = R_N = R$ , замыкая структуру в кольцо:

$$X(i_1, \dots, i_N) = TR(G_1[i_1] G_2[i_2] \dots G_N[i_N]) \quad (9)$$

где  $W$  хранится в TR-формате. Таким образом:

$$\widehat{y} = TR\left(\prod_{n=1}^N G_n^{(W)}[i_n]\right) \prod_{n=1}^N x_n[i_n] \quad (10)$$

Обучение осуществляется end-to-end через дифференцируемую схему, аналогичную ТТ-рекурсии, с обновлением ядер  $G_n^{(W)}$  посредством градиентного спуска. Общее количество параметров  $P_{TR} = \sum_{n=1}^N I_n r_{n-1} r_n$ . Ожидается, что TR-подход позволит работать с данными на 2-3 порядка большей размерности при сокращении потребления памяти на 15-25% по сравнению с ТТ, с повышением точности на 5-10% в задачах с циклическими зависимостями.

В подразделе, посвященном методу снижения шума и восстановления информации в мультимодальных данных обобщены стандартные матричные разложения на тензорный случай с использованием Т-произведения. Разработаны рандомизированные однопроходные алгоритмы и методы с фиксированной точностью, ориентированные на работу с большими данными в условиях ограниченных вычислительных ресурсов. Показано, что предложенные решения позволяют уменьшить вычислительную сложность при сохранении требуемой точности восстановления. В предлагаемом методе предпринята попытка использовать тензорные подходы для снижения уровня шума и восстановления утраченных данных, что позволяет обеспечить согласованность между процессами мультимодального слияния данных и шумоподавления. В частности, анализируется возможность восстановления изображения при удалении до 90% пикселей, что актуально для обработки сигналов и изображений.



Рисунок 4. Восстановление изображения при 90% пропущенных данных

Формулировка оптимизационной задачи имеет вид:

$$\min_X \text{rank}(X), \text{ Subject to } X_{i,j} = M_{i,j} \quad (i,j) \in \Omega \quad (11)$$

Однако минимизация ранга является NP-трудной задачей из-за невыпуклости функции ранга. Для приближенного решения используется релаксация с заменой ранга на ядерную норму  $\|X\|_*$ , определяемую как сумма сингулярных значений матрицы:

$$\|X\|_* = \sum_{k=1}^{\min(m,n)} \sigma_k(X) \quad (12)$$

Задача tensor completion формулируется как поиск низкорангового тензора  $X$ , совпадающего с  $M$  в  $\Omega$ . Одна из распространенных формулировок использует разложение Такера, где тензор аппроксимируется как:

$$X = \mathcal{G} \times_1 U_1 \times_2 U_2 \times_3 \dots \times_d U_d \quad (13)$$

где  $\mathcal{G} \in \mathbb{R}^{R_1 \times R_2 \times \dots \times R_d}$  ядерный тензор,  $U_k \in \mathbb{R}^{I_k \times R_k}$  факторные матрицы, а  $\times_k$  произведение по моде  $k$ . Многомерный ранг определяется кортежем  $(R_1, R_2, \dots, R_d)$ . Оптимизационная задача может быть сформулирована как:

$$\min_{X,C} \|X - C\|_F^2, \text{ Subject to } \text{rank}(X) = R, P_\Omega(C) = P_\Omega(M) \quad (14)$$

где  $\|X\|_F$  норма Фробениуса,  $R$  заданный многомерный ранг,  $P_\Omega$  проектор на множество наблюдаемых элементов и  $C$  вспомогательный тензор.

Для решения применяется алгоритм чередующихся наименьших квадратов (ALS), чередуя обновления  $X$  и  $C$ :

- Обновление  $X^{(n)} \approx \mathcal{L}(C^{(n)})$ , где  $\mathcal{L}$  оператор низкоранговой аппроксимации
- Обновление  $C^{(n+1)} = \Omega \odot M + (1 - \Omega) \odot X^{(n)}$

Метод поддерживает различные тензорные разложения, такие как CP (Canonical Polyadic), Tensor Train или Block Tensor разложение. Однако метод обладает ограничениями, такими как высокая вычислительная сложность ALS-алгоритма, особенно при большом числе итераций, и

зависимость от выбора разложения тензора. Рандомизированные алгоритмы представляют собой класс методов, которые обрабатывают поток данных за один проход, без необходимости хранить весь набор данных в оперативной памяти. Это свойство является крайне важным при работе с крупномасштабными данными, которые не помещаются в память целиком.

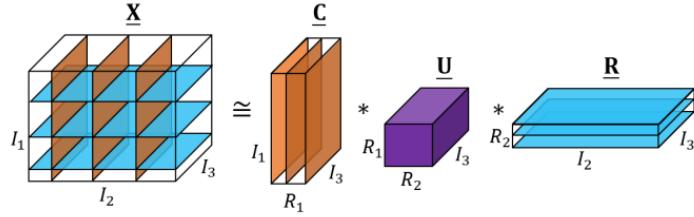


Рисунок 5. Рандомизированная аппроксимация с низким трубным рангом

В модифицированной версии алгоритма T-QR-разложение заменяется на T-произведение тензоров и обращение малых тензоров, что обеспечивает более высокую степень параллелизма. Описание этого усовершенствованного подхода представлено в алгоритме.

Пусть  $\underline{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$  и  $\underline{\Omega} \in \mathbb{R}^{I_1 \times K \times I_3}$  случайный тензор. Определим  $\underline{Y} = \underline{X} * \underline{\Omega}$ ,  $\underline{W} = \underline{X}^T * \underline{Y}$  а экономичное T-SVD-разложение тензора  $\underline{Y}$  задается как  $\underline{Y} = \underline{U} * \underline{S} * \underline{V}^T$ . Тогда

$$\underline{Q} = \underline{Y} * \underline{\hat{V}} * \underline{\hat{S}}^{-1}, \underline{B} = (\underline{W} * \underline{\hat{V}} * \underline{\hat{S}}^{-1})^T \quad (15)$$

получаем аппроксимацию  $\underline{X} \approx \underline{Q} * \underline{B}$ , которая имеет такую же точность, как и базовые рандомизированные алгоритмы. Кроме того, выполняется равенство

$$\|\underline{B}\|_F^2 = \text{trace}(\underline{H}_1) \quad (16)$$

где  $\underline{H}_1 = \underline{H}(:, :, 1)$  это первый фронтальный срез тензора  $\underline{H} = \underline{W} * \underline{W}^T * (\underline{Y} * \underline{Y})^{-1}$ .

Рассмотрим тождества

$$\underline{Q} = \text{orth}(\underline{X} * \underline{\Omega}) = \text{orth}(\underline{Y}) = \underline{\hat{U}} = \underline{Y} * \underline{\hat{V}} * \underline{\hat{S}}^{-1} \quad (17)$$

Подставляя (15) в  $\underline{B} = \underline{Q}^T$ , получаем:

$$\underline{B} = \underline{Q}^T * \underline{X} = (\underline{W} * \underline{\hat{V}} * \underline{\hat{S}}^{-1})^T \quad (18)$$

Поскольку  $\underline{Q}$  представляет собой ортонормализацию тензора  $\underline{X} * \underline{\Omega}$  аппроксимация  $\underline{Q} * \underline{B}$  обеспечивает такую же точность, как и базовые рандомизированные алгоритмы, но без итераций по степеням и без передискретизации. Теперь, используя тот факт, что  $\|\underline{B}\|_F^2 = \text{trace}(\underline{G}_1) = \underline{G}(:, :, 1)$ , где  $\underline{G}_1$  первый фронтальный срез тензора  $\underline{G} = \underline{B}^T * \underline{B}$  и получаем:

$$\begin{aligned} \|\underline{B}\|_F^2 &= \text{trace}((\underline{\hat{S}}^{-1} * \underline{\hat{V}}^T * \underline{W}^T * \underline{W} * \underline{\hat{V}} * \underline{\hat{S}}^{-1})_1) \\ &= \text{trace}((\underline{W}^T * \underline{W} * \underline{\hat{S}}^{-2} * \underline{\hat{V}})_1) \\ &= \text{trace}((\underline{W}^T * \underline{W} * (\underline{Y}^T * \underline{Y})^{-1})_1) \end{aligned} \quad (19)$$

Определим  $\underline{T} = \underline{W}^T * \underline{W}$  и  $\underline{Z} = \underline{Y}^T * \underline{Y}$ . Тогда критерий остановки  $\|\underline{X} - \underline{Q} * \underline{B}\|_F^2 = \|\underline{Q}\|_F^2 - \|\underline{B}\|_F^2$  и можно переписать в следующем виде:

$$\left\| \underline{X} - \underline{Q} * \underline{B} \right\|_F^2 = \left\| \underline{Q} \right\|_F^2 - \text{trace} \left( (\underline{T} * \underline{Z}^{-1})_1 \right) \quad (20)$$

**Input** : The data tensor  $\underline{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ , a block size  $b$ , a power iteration  $q$  and an approximation error bound.

**Output**: The QB approximation of the tensor  $\|\underline{X} - \underline{Q} * \underline{B}\|_F \leq \epsilon$

```

1  $\underline{Y} = []$ ,  $\underline{W} = []$ ;
2  $E = \|\underline{X}\|_F$ ,  $tol = \epsilon^2$ ;
3 for  $i = 1, 2, \dots$  do
4   Generate a random tensor  $\underline{\Omega}_i$  of size  $I_2 \times b \times I_3$ ;
5   for  $j = 1, 2, \dots, q$  do
6      $\underline{W}_i = \underline{X}^T * \underline{X} * \underline{\Omega}_i - \underline{W} * \underline{Z}^{-1} * \underline{W}^T * \underline{\Omega}_i$ ;
7      $\underline{\Omega}_i = \text{orth}(\underline{W}_i)$ ;
8   end
9    $\underline{Y}_i = \underline{X} * \underline{\Omega}_i$ ,  $\underline{W}_i = \underline{X}^T * \underline{Y}_i$ ;
10   $\underline{Y} = \underline{Y} \boxplus_2 \underline{Y}_i$ ,  $\underline{W} = \underline{W} \boxplus_2 \underline{W}_i$ ;
11   $\underline{Z} = \underline{Y}^T * \underline{Y}$ ,  $\underline{T} = \underline{W}^T * \underline{W}$ ;
12  if  $E - (\text{trace}((\underline{T} * \underline{Z}^{-1})_1)) < tol$  then
13    break
14  end
15 end
16  $[\hat{\underline{V}}, \hat{\underline{D}}] = \text{T-EIG}(\underline{Z})$ ;
17  $\underline{Q} = \underline{Y} * \hat{\underline{V}} * \text{sqrt}(\hat{\underline{D}})^{-1}$ ,  $\underline{B} = (\underline{W} * \hat{\underline{V}} * \text{sqrt}(\hat{\underline{D}})^{-1})^T$ ;

```

Таким образом, используя эти преобразования, можно уменьшить время вычислений, при этом точность снижается незначительно.

**В третьей главе** представлены результаты математического моделирования. Анализируются экспериментальные результаты метода слияния на основе тензорного представления (LMF, Tucker, Tensor Train, Tensor Ring) и сравнительный анализ. Далее оценивается метод снижения шума на синтетических тензораах, сжатии изображений и видео, повышении разрешения изображений и применении в глубоком обучении.

### Метод слияния мультимодальных данных на основе тензорного представления

Эксперименты выполнялись на наборе данных POM, который состоит из 903 видеороликов с обзорами фильмов. Каждое видео сопровождается аннотациями со следующими характеристиками говорящего: уверенный, страстный, приятный голос, доминирующий, заслуживающий доверия, яркий, опытный, развлекательный, сдержанный, доверчивый, расслабленный, общительный, тщательный, нервный, убедительный и юмористический.

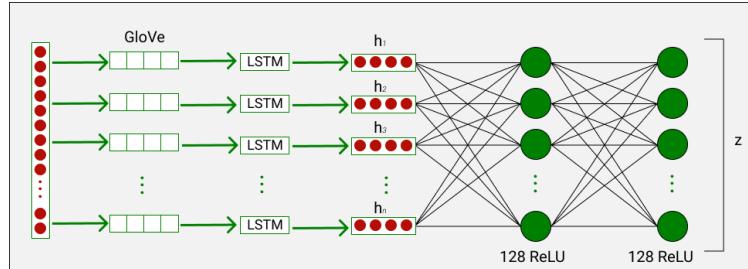


Рисунок 6. Графическое представление извлечения признаков из текста

На первом этапе текстовой обработки каждое высказывание приводится к последовательности токенов, для которых строятся векторные представления на основе предобученных эмбеддингов GloVe размерности 300. Такие векторы отражают семантическую близость и ассоциативные связи между словами, что позволяет модели учитывать скрытые семантические отношения даже для слов, которые редко встречаются в рассматриваемом корпусе. В результате каждое высказывание представляется в виде временного ряда 300-мерных векторов. Далее временная последовательность слов подается на вход сети типа LSTM (Long Short-Term Memory).

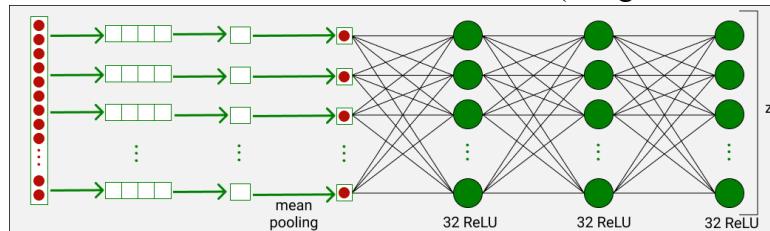


Рисунок 7. Графическое представление извлечения признаков из изображения

Визуальная модальность в корпусе РОМ представлена видеорядом, на котором отчетливо видны лицо, мимика и движения говорящего. Для извлечения информативных визуальных признаков из каждого видеоролика применяется двухэтапный подход. Сначала из сырых видеоданных извлекаются набор высокоровневых дескрипторов выражения лица с помощью специализированного инструментария FACET, а также координаты лицевых ориентиров при помощи библиотеки OpenFace.

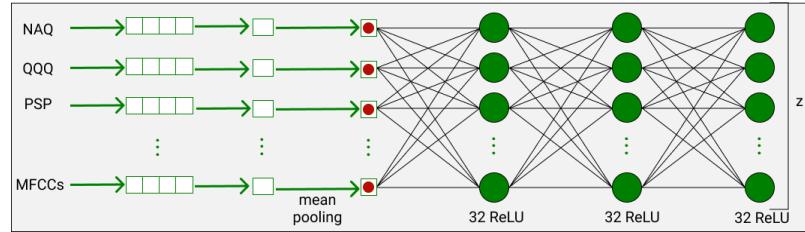


Рисунок 8. Графическое представление извлечения признаков из акустических данных

Акустическая модальность корпуса РОМ отражает просодические и голосовые характеристики речи, тесно связанные с эмоциональным состоянием говорящего и его убеждающим стилем. Для аудиоанализа используется специализированный набор низкоровневых дескрипторов, извлекаемых с помощью пакета COVAREP.

Таким образом, результирующее мультимодальное пространство характеризуется размерностью  $\mathcal{X} \in \mathbb{R}^{128 \times 64 \times 128}$ .

Для обеспечения стабильности обучения все модели обучались в единой среде на GPU NVIDIA RTX 3090 с одинаковыми параметрами оптимизатора (Adam,  $\eta = 0.001$ ,  $batch size = 32$ ).

Таблица 1 – Ключевые результаты экспериментов

Модель	Оптимальные ранги	MAE	Время обучения (мин)	Относительное улучшение
<b>LMF</b>	R=4	0.47	40	базовая модель
<b>Tucker</b>	(6,8,5)	0.415	80-90	+13% точности к LMF
<b>Tensor Train</b>	(6,4,3)	0.358	50	+24% к LMF, +14% к Tucker
<b>Tensor Ring</b>	(4,5,3)	0.318	30-35	+32% к LMF, +11% к TT

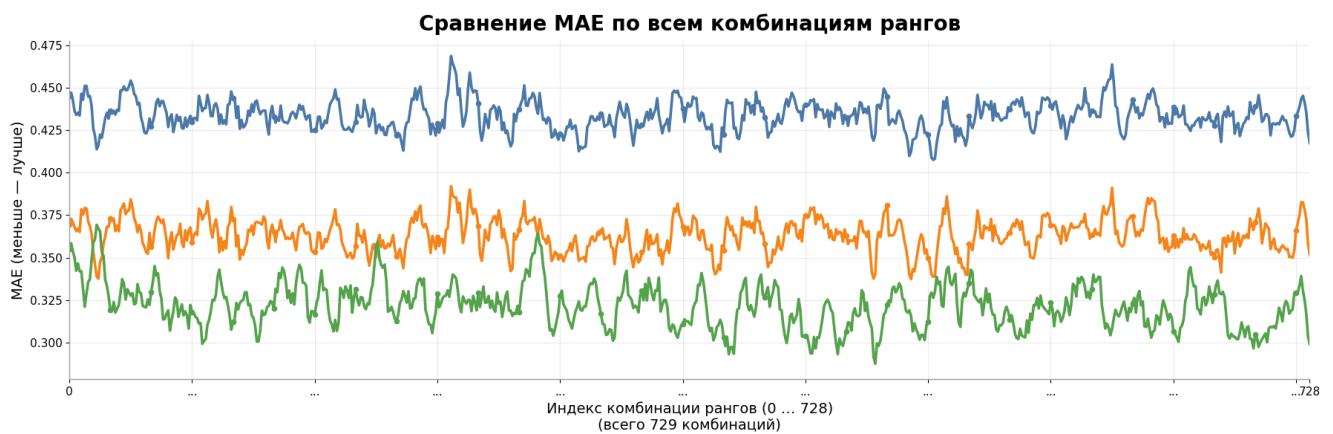


Рисунок 9. Сравнение MAE по всем комбинациям рангов для различных методов тензорного разложения. Зеленая линия (Tensor Ring), Оранжевая линия (Tensor Train), Синяя линия (Tucker)

По результатам экспериментов (см. зеленую линию на графике MAE), Tensor Ring демонстрирует наилучшие показатели точности. Среднее значение MAE  $\approx 0.325$ , минимальное 0.250 при рангах (4, 5, 3). Это улучшение на  $\approx 10\%$  по сравнению с Tensor Train и на  $\approx 23\%$  по сравнению с Tucker. Важно отметить, что TR сохраняет устойчивость ошибки даже при изменении рангов, что говорит о его низкой чувствительности к гиперпараметрам и высокой обобщающей способности.

Визуальное сравнение показало, что колебания MAE для TR минимальны, что свидетельствует о высокой стабильности модели. Более того, TR позволяет использовать различные комбинации рангов для каждой модальности (multi-rank), что делает модель адаптивной к неоднородной структуре данных. Таким образом, экспериментальные результаты подтверждают, что Tensor Ring демонстрирует оптимальное сочетание точности, устойчивости и вычислительной эффективности, превосходя не только Tucker и Tensor Train, но и базовый LMF.

### Метод снижения шума и восстановление информации в мультимодальных данных

В этой части работы мы оцениваем эффективность алгоритмов на синтетических и реальных тензорных данных. Эксперименты проводились в среде MATLAB на компьютере с процессором Intel(R) Core(TM) i7-5600U с тактовой частотой 2.60 ГГц и оперативной памятью 8 ГБ. Первый эксперимент выполнен на синтетических данных. Второй и третий эксперименты посвящены задачам сжатия изображений и видео. Два последних эксперимента демонстрируют применение предложенных подходов к задачам сверхразрешения изображений и глубокого обучения.

**Синтетические тензоры данных.** В этом эксперименте проводится сравнение эффективности предложенных алгоритмов и базовых методов на синтетических тензорных данных. Пусть необходимо сгенерировать случайный тензор с низким трубным рангом (low tubal rank approximation). Для этого рассмотрим безошибочный (чистый) тензор  $\underline{X} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$  с трубным рангом  $R$ , который определяется как

$$\underline{X}_{clean} = randn(I_1, R, I_3) * randn(R, I_1, I_3) \quad (21)$$

и добавим к нему шумовой член для генерации зашумленного тензора  $\underline{X}_{perturb} = \underline{X}_{clean} + \delta \frac{\underline{Y}}{\|\underline{Y}\|_F} \|\underline{X}_{clean}\|_F$ , где  $\underline{Y}$  стандартный гауссовский тензор того же размера, что и исходный тензор  $\underline{X}$ .

Таблица 2 – Сравнение времени вычислений и относительных ошибок предложенных алгоритмов и усеченного T-SVD. Значения в таблице представлены в виде (время вычислений, относительная ошибка)

Размер тензора $n$	Алгоритм 9	Алгоритм 10	Алгоритм 11	Усечённый T-SVD
$n=200$	$(2.94, 2.95e^{-10})$	$(2.71, 2.58e^{-09})$	$(1.18, 4.72e^{-09})$	$(11.43, 1.34e^{-09})$
$n=300$	$(8.74, 6.47e^{-10})$	$(4.21, 6.95e^{-09})$	$(3.07, 9.20e^{-09})$	$(36.81, 1.17e^{-09})$
$n=400$	$(20.86, 1.15e^{-9})$	$(7.98, 1.27e^{-08})$	$(6.65, 1.63e^{-08})$	$(81.83, 2.11e^{-09})$
$n=500$	$(45.89, 1.90e^{-9})$	$(20.59, 2.34e^{-08})$	$(19.32, 1.61e^{-08})$	$(195.25, 3.43e^{-09})$

Из таблицы видно, что предложенные алгоритмы (особенно алгоритм 11) значительно снижают время вычислений по сравнению с усеченным T-SVD, при этом точность аппроксимации остается на том же уровне или снижается незначительно.

**Сжатие изображений.** В данном эксперименте мы оцениваем эффективность предложенных рандомизированных одношаговых алгоритмов при решении задачи сжатия изображений. Для экспериментов использовался набор данных Kodak и были рассмотрены четыре изображения: Kodim15, Kodim17, Kodim18 и Kodim23. Два первых изображения имеют размер  $512 \times 768 \times 3$ , а два последних  $768 \times 512 \times 3$ . Мы применили предложенные алгоритмы и сравнили их с базовыми методами.

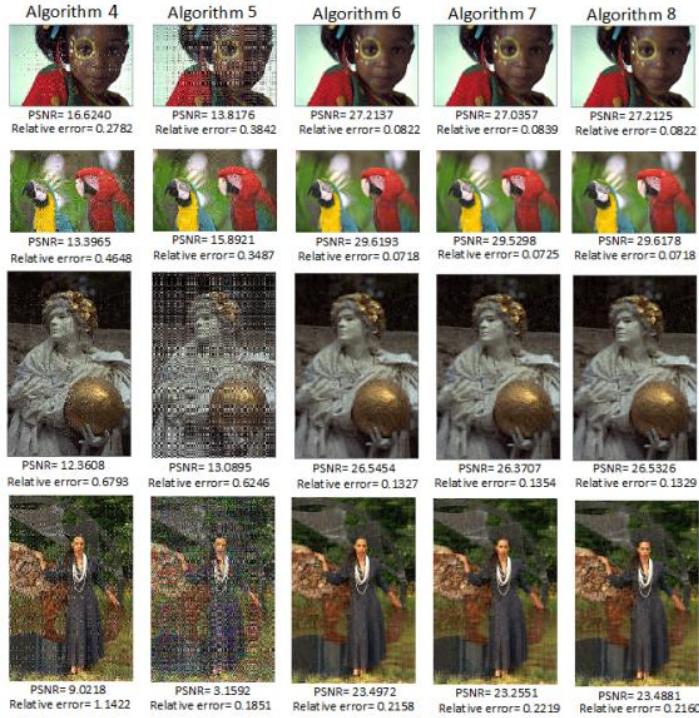


Рисунок 10. Восстановленные изображения, полученные с использованием различных одношаговых (single-pass) алгоритмов

Таким образом, результаты показывают, что предложенные алгоритмы обеспечивают более точное восстановление изображений, сохраняя при этом высокую вычислительную эффективность и способность к обработке данных большого размера.

**Сжатие видео.** В данном эксперименте исследуется производительность предложенных рандомизированных одношаговых алгоритмов при решении задачи сжатия видео.

В качестве исходных данных использовались видеодатасеты Foreman и News.

Размер каждого видеоролика представлен в виде тензора третьего порядка размером  $144 \times 176 \times 300$ . Сначала была протестирована эффективность одношаговых алгоритмов для низкоранговых тубальных аппроксимаций указанных видеоданных при следующих параметрах скетча  $L=90$ ,  $K=90$ ,  $H=20$ ,  $R=20$ . Значения PSNR всех кадров видеороликов Foreman и News, полученные с помощью предложенных одношаговых алгоритмов и базовых методов, представлены на рисунках 11 и 12.

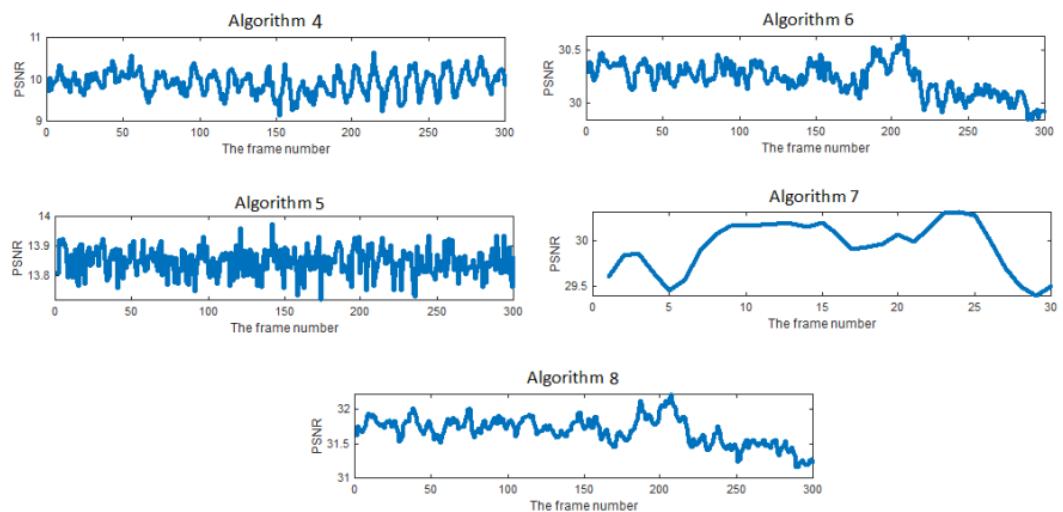


Рисунок 11. PSNR всех кадров видео News, вычисленный при тех же параметрах скетча  $L = 90$ ,  $K = 90$ ,  $H = 20$ ,  $R = 50$ . Предложенные алгоритмы показывают высшую производительность по сравнению с базовыми методами

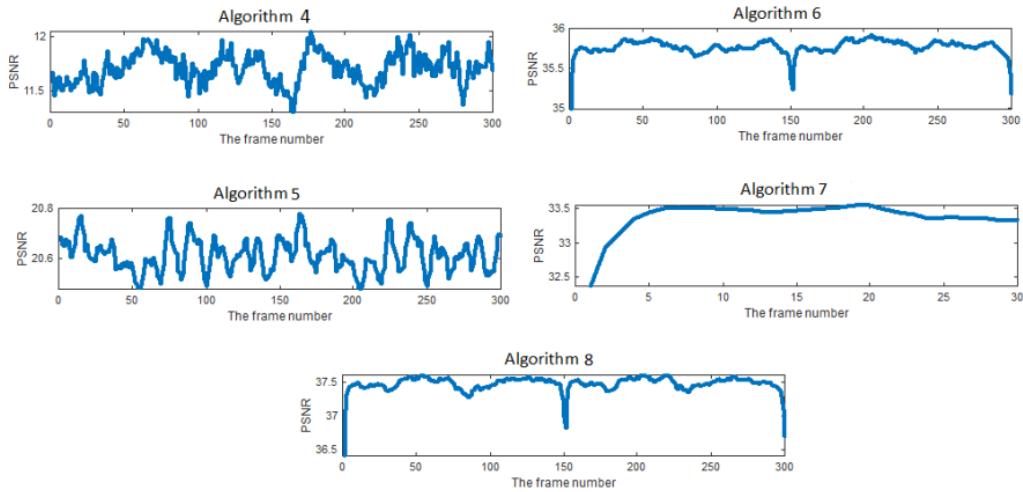


Рисунок 12. PSNR всех кадров видео News, вычисленный при тех же параметрах скетча  $L = 90, K = 90, H = 20, R = 50$ . Предложенные алгоритмы показывают высшую производительность по сравнению с базовыми методами

Также реконструированные кадры некоторых фреймов указанных видео приведены на рисунках 13 и 14.



Рисунок 13. Восстановленные отдельные кадры видео Foreman, полученные при тех же параметрах. Наблюдается лучшее качество реконструкции у предложенных одношаговых алгоритмов



Рисунок 14. Восстановленные отдельные кадры видео News при тех же параметрах

Предложенные методы демонстрируют наиболее точную реконструкцию и устойчивость к шумам. Полученные результаты показывают, что предложенные алгоритмы обладают высокой устойчивостью к выбору параметров скетча и демонстрируют высокую точность и стабильность при сжатии видео. Таким образом, данный пример подтверждает надежность и эффективность предложенных рандомизированных одношаговых алгоритмов для задач видеокомпрессии.

**Повышение разрешения изображений.** В данном эксперименте мы исследуем эффективность и применимость предложенных подходов к задаче повышения разрешения изображений (super-resolution). В эксперименте использовались пять изображений: Peppers, Airplane, Kodim01, Kodim02 и Kodim03. Первые два изображения имеют размер  $256 \times 256 \times 3$  и остальные три  $512 \times 768 \times 3$ . Все изображения увеличивались в 4 раза по осям  $x$  и  $y$ , для всех изображений использовался тубальный ранг  $R=60$ .

Таблица 3 – Сравнение времени вычислений и значений PSNR (в скобках указано: время, PSNR)

Изображение	Одношаговое восстановление	Детерминированное восстановление
Peppers	(26.99, 22.01)	(42.70, 22.01)
Airplane	(27.69, 22.13)	(44.58, 22.11)
Kodim01	(50.34, 20.44)	(142.42, 20.56)
Kodim02	(44.58, 26.70)	(137.30, 26.91)
Kodim03	(48.05, 26.96)	(146.76, 27.42)



Рисунок 15. Результаты суперразрешения, полученные с использованием алгоритма завершения тензора и предложенных рандомизированных одношаговых методов для низкорангового приближения оператора  $\mathcal{L}$  из уравнения

Результаты показывают, что предложенные рандомизированные алгоритмы фиксированной точности восстанавливают изображения так же точно, как и детерминированный подход (усеченный T-SVD), но требуют значительно меньше вычислительных затрат. Это демонстрирует преимущество предложенных одношаговых рандомизированных алгоритмов в задаче повышения разрешения изображений.

**Применение в глубоком обучении.** В данном эксперименте рассматривается применение предложенного метода завершения тензора к задаче точного обнаружения объектов (object detection), одной из ключевых задач компьютерного зрения в контексте глубокого обучения.

Для демонстрации берутся два изображения, показанные на рисунке 16 (первый столбец). В некоторых частях изображений вручную удаляются пиксели, что иллюстрируется на рисунке 16 (второй столбец).

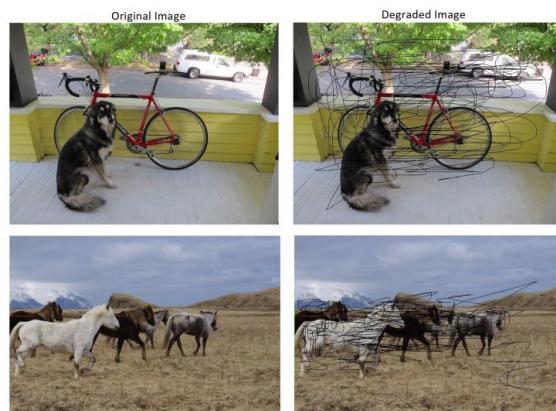


Рисунок 16. Оригинальные и искаженные изображения, использованные в эксперименте

Выполненные эксперименты показали, что при обработке искаженных изображений (с утратой части пикселей) качество работы детектора объектов YOLOv3 существенно снижается: сеть допускала ошибки классификации (например, собака распознавалась как кошка, лошадь – как жираф) и пропускала часть объектов на сцене. После применения предложенного метода завершения тензора в качестве предобработки, позволившего восстановить испорченные изображения, работа YOLOv3 заметно улучшилась: на восстановленном изображении с собакой корректно были обнаружены три объекта (велосипед, собака и грузовик) с точным определением границ, а на изображении с лошадьми – все четыре лошади без ошибок классификации. Тем самым продемонстрирована эффективность предложенного метода завершения тензора для повышения устойчивости и точности детекции объектов в условиях повреждения или частичной утраты пиксельной информации.

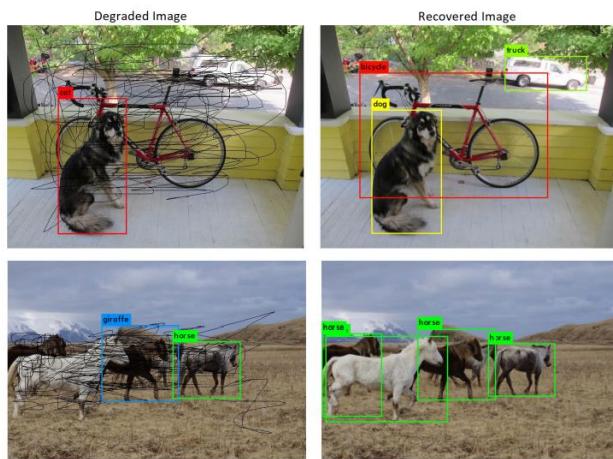


Рисунок 17. Результаты обнаружения объектов на искаженных изображениях (слева) и на восстановленных изображениях (справа), полученные с помощью метода завершения тензора, основанного на предложенных рандомизированных одношаговых алгоритмах

В данной работе предложены эффективные одношаговые (single-pass) и фиксированной точности (fixed-precision) алгоритмы для вычисления низкоранговых тубальных аппроксимаций тензоров третьего порядка. На первом этапе были разработаны три новых одношаговых алгоритма для низкорангового тубального приближения, а также проведено исследование их устойчивости и точности при аппроксимации изображений и видео. Результаты моделирования подтвердили, что предложенные одношаговые алгоритмы обладают большей устойчивостью, чем существующие базовые методы. Кроме того, была продемонстрирована их эффективность в задачах сжатия данных, повышения разрешения изображений и глубокого обучения. Во второй части работы были предложены два новых алгоритма фиксированной точности для низкоранговой тубальной аппроксимации.

Проведенные эксперименты подтвердили, что предложенные методы обеспечивают лучшие результаты по сравнению с современными аналогами (state-of-the-art) при меньших затратах вычислительного времени.

## ЗАКЛЮЧЕНИЕ

В ходе диссертационного исследования получены следующие основные результаты

- Сформирована и обоснована структурная модель интеграции мультимодальных данных в виде тензорного представления, обеспечивающая сохранение межмодальных связей и семантической целостности данных.
- Разработан метод мультирангового тензорного слияния мультимодальных данных, позволяющий уменьшить размерность признакового пространства, снизить риск переобучения и повысить устойчивость алгоритмов машинного обучения.
- Предложен алгоритм подавления шумов и восстановления недостающих элементов дан-

ных в тензорной форме, обеспечивающий повышение достоверности и полноты исходной информации.

- Применены рандомизированные однопроходные алгоритмы для устойчивости глубокой нейронной сети, повышения разрешения изображений, сжатия видео и изображений.
- Создана программа на языке Python, реализующая предложенные методы и обеспечивающая их применение в прикладных задачах обработки мультимодальных больших данных.
- Проведен комплекс вычислительных экспериментов на реальных и синтетических данных, подтвердивший преимущество предложенных подходов по сравнению с существующими методами по критериям точности, устойчивости и вычислительной эффективности.

## СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ АВТОРОМ ПО ТЕМЕ ДИССЕРТАЦИИ

### Публикации

1. Rezaeian N., Novikova G.M. Detecting Near-duplicates in Russian Documents through Using Fingerprint Algorithm Simhash / Procedia Computer Science, 103 (2017), P.421-425.
2. Rezaeian N., Novikova G. Persian Text Classification using naive Bayes algorithms and Support Vector Machine algorithm / Indonesian Journal of Electrical Engineering and Informatics (IJEI), 8(1) (2020), P.178-188.
3. Rezaeian N., Novikova G.M. An Approach to Solve the Problem of Creating an Automatic Knowledge Control System Based on Text Analysis / 2020 V International Conference on Information Technologies in Engineering Education (Inforino 2020) 2020.
4. Rezaeian N., Novikova G.M. Morphological and syntactic analysis of Persian text with conditional random fields / International Research Journal, 5(47) 2016.
5. Rezaeian N., Gurina R., Saltykova O.A., Lokmane Hezla, Mammetnazar Nohurov, Kazem Reza Kashyzadeh. Novel GA-Based DNN Architecture for Identifying the Failure Mode with High Accuracy and Analyzing Its Effects on the System / Applied Sciences, 14(8), article 3354.
6. Rezaeian N., L. Hezla, R. Gurina, M. Hezla, M. Nohurov, S. Aouati. The Role of Artificial Intelligence in Improving Failure Mode and Effects Analysis (FMEA) Efficiency in Construction Safety Management / AI Technologies and Virtual Reality (AIVR 2023), Vol.382, P.397-411.
7. Rezaeian N., Kazem Reza Kashyzadeh, Gene Expression Programming to High-Cycle and Low-Cycle Fatigue Life Prediction of Polymer-Reinforced Concrete in the Vicinity of Room, Water, and Seawater Considering Different Immersion Times / Journal of Failure Analysis and Prevention, 25(4), P.1125-1137.
8. Rezaeian N., Ahmadi-Asl S., Ugwu U.O. Randomized Algorithms for Computing the Generalized Tensor SVD Based on the Tensor Product / Communications on Applied Mathematics and Computation. 2025.
9. Rezaeian, N., Salman Ahmadi-Asl, Roman Garayev, Lukmanov R.A., Asad Masood Khattak, Manuel Mazzara. Efficient Smooth Tensor Train and Tensor Ring Completion for Image Classification Enhancement / IEEE Access. 2025.
10. Резаиан Н., Высоцкий Р.Н., Нажмутдинов Р.А., Клюкин А.В. Упрощенное обучение моделей сегментации экземпляров / Научно-практический печатный журнал «Нефть и газ: опыт и инновации». Т.9. №2, 2025.

### Результаты интеллектуальной деятельности

1. Резаиан Н., Поддубский А.А., Носов А.А. Программа интеграции на основе мультимодальных данных RGB и тепловизора. Свидетельство о государственной регистрации программы для ЭВМ № 2025681907 от 20.08.2025.
2. Резаиан Н., Поддубский А.А, Гречаников Н.Д., Князев А.М. Цифровой аватар. Свидетельство о государственной регистрации программы для ЭВМ № 2025685026 от 19.09.2025.
3. Резаиан Н., Поддубский А.А., Высоцкий Р.Н. Интеллектуальный ассистент преподавателя. Свидетельство о государственной регистрации программы для ЭВМ № 2025615880 от

11.03.2025.

4. Резаиан Н., Реза Каши Заде Казем, Горбани Сиамак Новая программа для расчета сил и моментов на подвеску автомобиля при различных маневрах и различных условиях вождения на основе динамики многотельной модели: FVD-MBD. Свидетельство о государственной регистрации программы для ЭВМ № 2023667494 от 15.08.2023.

5. Резаиан Н., Нохуров М., Гурина Р.Р., Хезла Локман Разработка методов поощрения работников путем внедрения геймификации для поддержки мотивации для дальнейшего анализа данных с применением методов искусственного интеллекта. Свидетельство о государственной регистрации программы для ЭВМ № 2023663626 от 27.06.2023.

6. Резаиан Н., Реза Каши Заде Казем Программа HMFD оптимизации регулировки углов колес транспортных средств с целью уменьшения повреждений подвески. Свидетельство о государственной регистрации программы для ЭВМ № 2023618409 от 24.04.2023.

7. Резаиан Н., Должикова А.В., Завьялов А.А., Должиков П.К., Калинина Ю.М., Никишин А.Н. Лингвокультурологический интерактивный тренажер для иностранцев «Знакомлюсь с Россией». Свидетельство о государственной регистрации программы для ЭВМ № 2023665211 от 12.07.2023.

8. Резаиан Н., Нохуров М., Гурина Р.Р., Хезла Локман Разработка методов отслеживания показателей присутствия и других ключевых показателей работников на отделах предприятия в полуавтоматическом режиме для дальнейшего применения методов искусственного интеллекта. Свидетельство о государственной регистрации программы для ЭВМ № 2022683421 от 05.12.2022.

9. Резаиан Н., Нохуров М., Гурина Р.Р., Хезла Локман Разработка метода сбора данных для анализа и принятия решений с использованием искусственного интеллекта. Свидетельство о государственной регистрации программы для ЭВМ № 2022665750 от 19.08.2022.

10. Резаиан Н., Нохуров М., Гурина Р.Р., Хезла Локман Совершенствование архитектуры нейронной сети на основе генетических алгоритмов, используемых в анализе причин и последствий потенциальных несоответствий. Свидетельство о государственной регистрации программы для ЭВМ № 2022662896 от 07.07.2022.

11. Резаиан Н., Месбах Мохсен, Сури Камбиз Автоматизированный расчет однократного абсорбционного чиллера бромид лития вода, соединенного с солнечным коллектором. Свидетельство о государственной регистрации программы для ЭВМ № 2022611464 от 25.01.2022.

12. Поддубский А.А., Высоцкий Р.Н. Платформа многоязычной вопросно-ответной системы на основе языковой модели с использованием глубокого обучения. Свидетельство о государственной регистрации программы для ЭВМ № 2022611463 от 25.01.2022.

13. Резаиан Н., Поддубский А.А., Гречаников Н.Д., Князев А.М., Рождественский Е.В. Умный Отчёт: интеллектуальная система извлечения и визуализации информации. Свидетельство о государственной регистрации программы для ЭВМ № 2025687762 от 15.10.2025.

14. Резаиан Н., Новикова Г.М. Программа тестирования знаний на основе анализа текста. Свидетельство о государственной регистрации программы для ЭВМ № 2020614754 от 24.04.2020.

15. Резаиан Н., Новикова Г.М. Платформа для электронной библиотеки цифрового сотрудничества (ЭБЦС). Свидетельство о государственной регистрации программы для ЭВМ № 2020618462 от 29.07.2020.

16. Резаиан Н., Новикова Г.М. Синтаксическая и морфологическая разметка текста на персидском языке с помощью случайных полей. Свидетельство о государственной регистрации программы для ЭВМ № 2016615758 от 20.06.2016.

## АННОТАЦИЯ

Резаиан Найм

«Методика семантического анализа мультимодальных больших данных на основе применения методов машинного обучения»

Диссертационная работа посвящена разработке методов представления, слияния и восстановления мультимодальных данных на основе тензорных представлений. Актуальность исследования обусловлена ограничениями традиционных подходов к интеграции разнородных источников информации, которые не сохраняют многомерную корреляционную структуру признаков и приводят к потере информативности. Целью работы является создание тензорного аппарата, позволяющего объединять и реконструировать данные различной природы при сохранении их внутренней структуры и сниженных вычислительных затратах. Обоснована необходимость перехода к тензорному представлению мультимодальных данных и выявлены ограничения классических разложений, связанные с вычислительной сложностью и чувствительностью к шуму. Предложен метод тензорного слияния, обеспечивающий формирование общего латентного пространства без вычисления полного декомпозиционного базиса и снижающий риск переобучения. Экспериментальные исследования на модельных и реальных наборах данных показали, что использование разложения Tensor Ring обеспечивает прирост точности мультимодальной аппроксимации до 32% по сравнению с базовыми решениями и почти двукратное ускорение вычислений относительно традиционных методов. Возможность выбора различных рангов для отдельных модальностей обеспечивает адаптивность при анализе асимметричных данных. Показано, что рандомизированные одношаговые процедуры восстановления достигают точности детерминированных методов при меньших временных затратах и сохраняют качество реконструкции изображений и видео, что подтверждает их применимость в задачах сжатия, суперразрешения и предобработки данных для глубинных нейронных сетей, включая повышение точности детекции объектов.

## ABSTRACT

Rezaian Naeim

«Methodology of Semantic Analysis of Multimodal Big Data Based on the Application of Machine Learning Methods»

The dissertation is devoted to the development of methods for representation, fusion, and reconstruction of multimodal data based on tensor models. The relevance of the research is determined by the limitations of traditional approaches to the integration of heterogeneous information sources, which fail to preserve the multidimensional correlation structure of features and lead to a loss of informativeness. The aim of the work is to develop a tensor framework that makes it possible to integrate and reconstruct data of different nature while preserving their internal structure and reducing computational costs. The necessity of switching to tensor representations of multimodal data is substantiated, and the limitations of classical decompositions associated with computational complexity and sensitivity to noise are identified. A tensor fusion method is proposed that provides the formation of a common latent space without computing a full decomposition basis and reduces the risk of overfitting. Experimental studies on simulated and real-world data sets have shown that the use of Tensor Ring decomposition yields an increase in the accuracy of multimodal approximation of up to 32% compared to baseline solutions and almost a twofold speed-up of computations relative to traditional methods. The ability to choose different ranks for individual modalities ensures adaptability when analyzing asymmetric data. It is shown that randomized reconstruction procedures achieve the accuracy of deterministic methods with lower time costs and preserve the quality of image and video reconstruction, which confirms their applicability to problems of compression, super-resolution, and data pre-processing for deep neural networks, including improving object detection accuracy.