

Федеральное государственное автономное  
образовательное учреждение высшего образования  
«Казанский (Приволжский) федеральный университет»  
Институт вычислительной математики и информационных технологий

На правах рукописи

Першин

## **Глубокие модели в медицинской диагностике на основе внимания рентгенолога**

Першин Илья Андреевич

Специальность 1.2.1

Искусственный интеллект и машинное обучение

Диссертация на соискание учёной степени

кандидата технических наук

Руководитель: канд. физ.-мат. наук, доцент

Тумаков Дмитрий Николаевич

Москва, Россия

2025

## Оглавление

Стр.

|                                                                                                                                                       |           |
|-------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| <b>Введение</b>                                                                                                                                       | <b>7</b>  |
| <b>Глава 1. Глубокая модель, учитывающая семантическую информацию, для предсказания траектории взгляда на рентгеновском снимке</b>                    | <b>16</b> |
| 1.1 Введение                                                                                                                                          | 16        |
| 1.2 Постановка задачи                                                                                                                                 | 19        |
| 1.3 Извлечение семантической информации с использованием метода логит-линзы                                                                           | 20        |
| 1.4 Архитектуры нейронных сетей                                                                                                                       | 23        |
| 1.5 Метрики качества предсказания взгляда                                                                                                             | 26        |
| 1.6 Эксперименты и сравнительный анализ моделей                                                                                                       | 27        |
| 1.7 Обсуждение                                                                                                                                        | 32        |
| <b>Глава 2. Внедрение взгляда рентгенолога в глубокую модель классификации патологий легких</b>                                                       | <b>35</b> |
| 2.1 Введение                                                                                                                                          | 35        |
| 2.2 Постановка задачи                                                                                                                                 | 37        |
| 2.3 Архитектуры нейронных сетей                                                                                                                       | 38        |
| 2.4 Наборы данных                                                                                                                                     | 41        |
| 2.5 Результаты                                                                                                                                        | 43        |
| <b>Глава 3. Разработка и валидация экспериментальной установки для сбора и анализа данных взгляда в условиях, приближенных к клинической практике</b> | <b>47</b> |
| 3.1 Введение                                                                                                                                          | 47        |
| 3.2 Разработка установки для отслеживания взгляда рентгенолога                                                                                        | 49        |
| 3.3 Протокол экспериментов с практикующими рентгенологами                                                                                             | 51        |
| 3.4 Методы оценки усталости и концентрации рентгенолога                                                                                               | 53        |
| 3.5 Методы анализа взгляда рентгенолога                                                                                                               | 56        |
| 3.6 Результаты экспериментов                                                                                                                          | 58        |
| 3.7 Обсуждение                                                                                                                                        | 64        |

|                                                                                               |           |
|-----------------------------------------------------------------------------------------------|-----------|
| <b>Глава 4. Интерактивная сегментация медицинских изображений на основе взгляда . . . . .</b> | <b>69</b> |
| 4.1 Введение . . . . .                                                                        | 69        |
| 4.2 Рабочая станция рентгенолога . . . . .                                                    | 72        |
| 4.3 Постановка задачи . . . . .                                                               | 74        |
| 4.4 Метод сегментации с использованием данных взгляда . . . . .                               | 74        |
| 4.5 Набор данных . . . . .                                                                    | 77        |
| 4.6 Эксперименты и результаты . . . . .                                                       | 77        |
| 4.7 Эксперименты с рентгенологами . . . . .                                                   | 80        |
| 4.8 Обсуждение . . . . .                                                                      | 84        |
| <b>Заключение . . . . .</b>                                                                   | <b>89</b> |
| <b>Словарь терминов . . . . .</b>                                                             | <b>92</b> |
| <b>Список рисунков . . . . .</b>                                                              | <b>94</b> |
| <b>Список таблиц . . . . .</b>                                                                | <b>96</b> |
| <b>Список литературы . . . . .</b>                                                            | <b>98</b> |

## Введение

**Актуальность темы.** В последнее десятилетие в клиническую практику массово внедряются решения на основе глубокого обучения [1—3]. Это связано с двумя факторами. Во-первых, увеличение вычислительных мощностей и количества доступных аннотированных медицинских данных позволило разработать модели глубокого обучения, которые в некоторых задачах сопоставимы или превосходят уровень человека-эксперта. Во-вторых, во время пандемии коронавируса COVID-19 было необходимо снизить избыточную нагрузку на здравоохранительную систему, сохранив или повысив скорость и качество диагностики. Это привело к регуляторной поддержке и появлению большого количества решений на основе глубокого обучения для разнообразных клинических задач [4].

Несмотря на значительные достижения в области глубокого обучения для решения медицинских задач, остается проблема эффективного внедрения таких решений в клиническую практику [5; 6]. В ряде исследований показано, что использование решений на основе глубокого обучения может приводить к ухудшению качества диагностики и снижению производительности врача. Кроме того, сложность интерпретируемости решений на основе глубокого обучения и объяснения полученных предсказаний повышает риски использования (например, из-за галлюцинирования) таких решений в критически важных областях, включая здравоохранение.

Взаимодействие врача с решениями на основе глубокого обучения является качественным шагом вперед от принятой парадигмы автономного использования решений на основе глубокого обучения [7]. При таком подходе целью внедрения решений на основе глубокого обучения является повышение эффективности и качества работы медицинских специалистов, а не их замена. С одной стороны, врач обладает клиническим опытом, интуицией и человеческим интеллектом, с которым человечество умеет работать. С другой стороны, решения на основе глубокого обучения получены на основе автоматизированного анализа большого количества данных, способны находить сложные закономерности, работают быстрее человека и не имеют негативных эффектов от усталости. Взаимодействие врача и машины потенциально может нивелировать недостатки и усилить преимущества человека и машины по-отдельности.

Основным инструментом рентгенолога является визуальный анализ медицинского изображения. Во время визуального анализа можно отслеживать направление взгляда врача на изображении при помощи устройства для отслеживания взгляда. Анализ движения взгляда врача открывает возможности для оценки качества визуального анализа (например, путем оценки соответствия протоколам визуального анализа, оценки усталости врача [8—10]), построения более гибких рекомендательных систем на основе глубоких моделей (например, за счет детекции ошибок при визуальном анализе, их классификации и сопоставлении взгляда врача с предсказанием глубокой модели о наличии той или иной патологии), повышения качества глубоких моделей для анализа медицинских изображений, использования данных взгляда в качестве альтернативного способа ввода данных (например, для интерактивной сегментации [11; 12]).

Проблема использования данных взгляда заключается в небольшом количестве общедоступных наборов данных с отслеживанием взгляда [13], особенно для узких областей, таких как медицина. В рамках исследования было найдено всего 3 публичных набора данных взгляда рентгенолога [14—16], которые суммарно содержат взгляд для 6650 рентгеновских снимков грудной клетки. Это связано с высокой стоимостью проведения экспериментов с отслеживанием взгляда, в которых требуется разработка программного обеспечения и создание аппаратной установки, привлечение специалистов в качестве участников эксперимента. По этой причине применение данных взгляда для глубоких моделей остается относительно малоизученной областью.

Взгляд рентгенолога содержит информацию, отражающую намерения рентгенолога, его экспертизу и психоэмоциональное состояние. Эта информация может быть использована для повышения эффективности рентгенолога и оптимизации рабочих процессов. В то же время визуальное восприятие человека хоть и имеет общие закономерности (например, движение от более темного к светлому или продолжение движения в том направлении, в котором оно уже было начато), но является в высокой степени индивидуальным процессом, зависящим от многих факторов и имеющим естественные флуктуации. Индивидуальность визуального восприятия, ограниченность доступных наборов данных со взглядом рентгенолога и сложность их сбора для каждой клинической задачи представляют наибольшую трудность для использования взгляда в глубоких моделях в медицине.

Одним из способов преодоления нехватки данных в глубоком обучении является создание синтетических или модифицирование реальных данных (например, путем применения методов маскирования текста при обработке естественного языка или аугментации изображений). Это позволяет не только разнообразить обучающую выборку, но и сделать глубокую модель более устойчивой к шуму, лучше понимать контекст. На основе этих рассуждений был выбран фокус диссертации на улучшении качества анализа медицинских изображений, путем использования данных о взгляде рентгенолога в глубоких моделях.

**Объект исследования** – медицинские изображения, в частности рентгеновские изображения грудной клетки и КТ-срезы органов брюшной полости.

**Проблема** – недостаточное качество моделей глубокого обучения для анализа медицинских изображений из-за ограниченного использования информации о визуальном внимании рентгенологов.

**Предмет исследования** – глубокие модели, использующие данные о взгляде, а также методы синтеза и анализа этих данных.

**Целью** диссертационной работы является разработка методов интеграции данных взгляда рентгенолога в глубокие модели для повышения качества решения клинических задач.

#### **Задачи работы:**

1. Разработать глубокую модель, учитывающую семантическую информацию, для прогнозирования взгляда рентгенолога на рентгеновском изображении грудной клетки.
2. Разработать способ повышения качества глубокой модели для классификации патологий на рентгеновском изображении грудной клетки, используя синтетические данные взгляда рентгенолога во время обучения модели.
3. Разработать и апробировать экспериментальную установку для анализа визуального поведения рентгенолога при анализе рентгеновских изображений. Провести валидирующие эксперименты с участием практикующих рентгенологов.
4. Разработать способ и глубокую модель для интерактивной сегментации медицинских изображений на основе взгляда. Провести тестирование модели в экспериментах с привлечением практикующих рентгенологов.



Рисунок 1 — Диаграмма решения исследовательской проблемы, отражающая вклад каждой главы диссертации

5. Разработать комплекс программ для проведения экспериментов по анализу медицинских изображений с отслеживанием взгляда в режиме реального времени и оконтуривания анатомических областей на медицинских изображениях на основе взгляде рентгенолога с использованием технологий искусственного интеллекта.

На рисунке 1 представлена диаграмма решения проблемы, иллюстрирующая вклад каждой из глав диссертации и последовательную связь между ними. Глава 1 направлена на решение проблемы ограниченной доступности данных взгляда для рентгеновских снимков грудной клетки, путем разработки глубокой модели для генерации синтетических данных взгляда. В главе 2 предложен метод для повышения качества классификации патологий на рентгеновских снимках грудной клетки, путем использования синтетических данных взгляда из предыдущей главы. В главе 3 описана экспериментальная установка для сбора данных взгляда рентгенолога, которая была протестирована в экспериментах с практикующими рентгенологами. Разработанная установка с небольшими изменениями далее использовались в главе 4 для интерактивной сегментации анатомических областей на КТ-снимках брюшной полости в режиме реального времени при помощи взгляда рентгенолога.

**Положения выносимые на защиту.**

1. Глубокая модель обогащенная семантической информацией для предсказания взгляда рентгенолога на рентгеновском изображении грудной клетки.
2. Способ повышения качества глубокой модели и глубокая модель для классификации патологий на рентгеновском изображении грудной клетки, используя синтетические данные взгляда рентгенолога во время обучения модели.
3. Экспериментальная установка для сбора и анализа данных визуального восприятия рентгенолога при анализе рентгеновских изображений, позволяющая выявлять устойчивые закономерности взгляда.
4. Способ и глубокая модель для интерактивной сегментации медицинских изображений на основе взгляда.
5. Комплекс программ для проведения экспериментов по анализу медицинских изображений с отслеживанием взгляда в режиме реального времени и оконтуривания анатомических областей на медицинских изображения на основе взгляде рентгенолога с использованием технологий искусственного интеллекта.

### **Научная новизна работы.**

1. Предложен и реализован метод предсказания траектории взгляда рентгенолога, основанный на интеграции в глубокую модель семантической информации, полученной методом логит-линзы из визуально-языковой модели, что обеспечивает улучшение точности по метрике ScanMatch до 48% по сравнению с базовыми моделями (глава 1) (Пункты 2, 4 и 13 паспорта специальности 1.2.1).
2. Предложен и разработан метод повышения точности классификации патологий на рентгеновских изображениях грудной клетки, заключающийся во включении синтетических данных взгляда на этапе обучения сверточной нейронной сети, что позволяет повысить AUROC на 10% по сравнению с моделью, использующей реальные данные взгляда (глава 2) (Пункты 2 и 4 паспорта специальности 1.2.1).
3. Разработана и валидирована экспериментальная установка для сбора и анализа данных визуального восприятия рентгенолога при интерпретации рентгеновских изображений, обеспечивающая выявление устойчивых паттернов взгляда и создающая основу для учета внимания врача

в интеллектуальных медицинских системах (глава 3) (Пункты 6 паспорта специальности 1.2.1).

4. Предложен способ интерактивной сегментации медицинских изображений, использующий набор координат взгляда врача как источник управления генерацией сегментационной маски, позволяющий улучшить качество сегментации и снизить ручные усилия при корректировке границ анатомических областей (глава 4) (Пункты 6 и 7 паспорта специальности 1.2.1).

**Теоретическая и практическая значимость работы** состоит в том, что разработанные глубокие модели и способы предназначены для повышения эффективности решения клинических задач и безопасности использования глубоких моделей в клинической практике.

**Достоверность научных достижений** обеспечивается корректным тестированием разработанных решений, в том числе с привлечением экспертов-рентгенологов; совпадением с результатами, полученными другими авторами; использованием глубоких моделей и способов в рамках реализации гранта РНФ №18-71-10072 «Разработка новых математических методов и алгоритмов анализа рентгеновских изображений органов грудной полости для автоматизации диагностики заболеваний легких с помощью решения задачи классификации изображений с высокой внутриклассовой и низкой межклассовой дисперсией» и гранта на создание центра искусственного интеллекта при поддержке аналитического центра при Правительстве Российской Федерации (Соглашение № 70-2021-00143 от 01.11.2021, ИГК 000000D730321P5Q0002).

**Апробация работы.** Основные результаты работы докладывались на следующих международных научных конференциях уровня А\* по искусственному интеллекту: Conference on Neural Information Processing Systems (Канада, 2024), IEEE/CVF Conference on Computer Vision and Pattern Recognition (Нашвилл, США, Канада), International Conference on Learning Representations (Сингапур, 2025). Также основные результаты работы докладывались на следующих международных конференциях: SPIE Medical Imaging (США, 2022 г.), Intelligent Systems Design and Applications (США, 2022 г.), IEEE Congress on Evolutionary

Computation СЕС (2023, США). Кроме того, основные результаты были представлены и обсуждались на итоговых научных конференция сотрудников Казанского федерального университета (2022-2023 г.), внутренних семинарах университета Иннополис (2020-2025 гг.), во время совместной научной работы с коллективом республиканского клинического онкологического диспансера (РКОД) имени профессора М.З.Сигала (2024 г.).

**Публикации.** Список всех публикаций автора по теме диссертации:

1. AI-based analysis of radiologist's eye movements for fatigue estimation: a pilot study on chest X-rays / I. Pershin [и др.] // Medical Imaging 2022: Image Perception, Observer Performance, and Technology Assessment / под ред. C. R. Mello-Thoms, S. Taylor-Phillips. — SPIE, 04.2022. — С. 39. — URL: <http://doi.org/10.1117/12.2612760>.
2. *Shmykova T., Khaertdinova L., Pershin I.* Zero-Shot Gaze-based Volumetric Medical Image Segmentation // Proceedings of the Computer Vision and Pattern Recognition (CVPR). — 06.2025. — URL: <https://sites.google.com/view/mmfm-biomed-2025/0>.
3. Artificial Intelligence for the Analysis of Workload-Related Changes in Radiologists' Gaze Patterns / I. Pershin [и др.] // IEEE Journal of Biomedical and Health Informatics. — 2022. — Сент. — Т. 26, № 9. — С. 4541—4550. — URL: <http://doi.org/10.1109/JBHI.2022.3183299>.
4. Gaze-based attention to improve the classification of lung diseases / M. Kholiavchenko [и др.] // Medical Imaging 2022: Image Processing / под ред. I. Išgum, O. Colliot. — SPIE, 04.2022. — С. 10. — URL: <http://doi.org/10.1117/12.2612767>.
5. Changes in Radiologists' Gaze Patterns Against Lung X-rays with Different Abnormalities: a Randomized Experiment / I. Pershin [и др.] // Journal of Digital Imaging. — 2023. — Янв. — Т. 36, № 3. — С. 767—775. — URL: <http://doi.org/10.1007/s10278-022-00760-2>.
6. AI-Based Extraction of Radiologists Gaze Patterns Corresponding to Lung Regions / I. Pershin [и др.] // Intelligent Systems Design and Applications. — Springer Nature Switzerland, 2023. — С. 386—393. — URL: [http://doi.org/10.1007/978-3-031-35501-1\\_39](http://doi.org/10.1007/978-3-031-35501-1_39).

7. *Pershin I., Mustafaeu T., Ibragimov B.* Contrastive Learning Approach to Predict Radiologist's Error Based on Gaze Data // 2023 IEEE Congress on Evolutionary Computation (CEC). — IEEE, 07.2023. — С. 1—6. — URL: <http://doi.org/10.1109/CEC53210.2023.10254056>.
8. Gaze Assistance for Efficient Segmentation Correction of Medical Images / L. Khaertdinova [и др.] // IEEE Access. — 2025. — Т. 13. — С. 14199—14213. — URL: <http://doi.org/10.1109/access.2025.3530701>.
9. Gaze-Assisted Medical Image Segmentation / L. Khaertdinova [и др.] // Advances in Neural Information Processing Systems 37. — Vancouver, BC, Canada, 2024. — URL: <https://neurips.cc/virtual/2024/103920>.
10. *Lvov D., Pershin I.* LogitGaze: Predicting Human Attention Using Semantic Information from Vision-Language Models // The Thirteenth International Conference on Learning Representations. — Singapore, 2025. — URL: <https://iclr.cc/virtual/2025/workshop/23968#collapse32765>.
11. *Свидетельство о гос. регистрации программы для ЭВМ.* Программа для оконтуривания анатомических областей на медицинских изображениях на основе взгляда рентгенолога с использованием технологий искусственного интеллекта / И. Першин, Т. Шмыкова, Л. Хаертдинова ; АНО ВО «Университет Иннополис». — № RU 2025619297 ; опубл. 14.04.2025, 2025616768 (ru).
12. *Свидетельство о гос. регистрации программы для ЭВМ.* Программа для проведения экспериментов по анализу медицинских изображений с отслеживанием взгляда рентгенолога в режиме реального времени с использованием технологий искусственного интеллекта / И. Першин, Д. Тумаков ; АНО ВО «Университет Иннополис». — № RU 2025619063 ; опубл. 14.04.2025, 2025616768 (ru).
13. *Першин И., Тумаков Д.* База данных по отслеживанию взгляда рентгенологов во время анализа медицинских изображений и обучения моделей искусственного интеллекта. — 14.04.2025. — Дата регистрации: 14.04.2025, Бюл. № 4. Авторы указаны в заявке на регистрацию. Свидетельство о государственной регистрации базы данных № 2025670018.

14. *Патент Российской Федерации*. Система и способ оценки усталости врача / И. А. Першин, Т. А. О. Мустафаев ; АНО ВО «Университет Иннополис». — № RU 2 800 312 С1 ; заявл. 30.12.2022 ; опубл. 20.07.2023, 2022135393 (ru). — 11 с. — Бюл. №20.

**Структура и объем диссертации.** Диссертация состоит из введения, 4 глав и заключения.

Полный объём диссертации составляет 107 страниц, включая 17 рисунков и 14 таблиц. Список литературы содержит 127 наименований.

# Глава 1. Глубокая модель, учитывающая семантическую информацию, для предсказания траектории взгляда на рентгеновском снимке

## 1.1 Введение

Настоящая глава посвящена разработке метода для повышения качества прогнозирования целенаправленного взгляда и обучению глубокой модели для генерации синтетических данных взгляда на рентгеновских снимках грудной клетки.

Задача моделирования и прогнозирования целенаправленного внимания человека – в частности, предсказания траектории фиксации взгляда во время визуального поиска — остаётся одной из ключевых тем в когнитивных науках, нейронауке и компьютерном зрении [17–19]. В последние годы интерес к этой задаче существенно усилился благодаря росту производительности моделей глубокого обучения, появлению масштабных наборов данных с аннотациями движений глаз и растущему спросу на интерпретируемые ИИ-системы в приложениях с высокой стоимостью ошибки, включая медицинскую диагностику.

Особый интерес вызывает прогнозирование целенаправленного взгляда – траектории визуального поиска, обусловленного заданием, вопросом или целевой категорией. В отличие от свободного просмотра изображения, направленное внимание характеризуется семантически детерминированными шаблонами, формируемыми на основе когнитивных стратегий и экспертного опыта. В медицинской визуализации эти паттерны особенно важны: они отражают диагностическое мышление рентгенолога и могут быть использованы в качестве априорной информации для повышения точности и объяснимости ИИ-моделей [20].

Наиболее ярким примером этого является анализ рентгенограмм грудной клетки. Проблема заключается в том, что визуальные признаки патологий часто выражены неявно, не ограничиваются чётко очерченными областями и требуют интегративного восприятия анатомических и патологических структур. Траектории взгляда опытного рентгенолога при визуальном восприятии

изображения отражают не только последовательность просмотра областей изображения, но и приоритетность областей с диагностической значимостью.

Однако практическое применение информации о взгляде ограничено высокой стоимостью её сбора. Современные наборы данных взгляда, такие как GazeSearch [20] и Eye Gaze Chest X-ray [14], остаются редкими и узкоспециализированными. Это создаёт значительный барьер для обучения глубоких моделей. В последние годы активно развиваются методы синтетической генерации траекторий взгляда, направленные на предсказание фиксации человека при наличии изображения и текстового описания цели.

Ранние подходы прогнозирования траектории взгляда опирались на построение карт заметности, либо на вручную созданных эвристиках [21], не учитывающих семантической информации о целевом объекте. Недавние подходы на основе архитектуры нейронных сетей Трансформер позволили существенно повысить качество прогнозирования целенаправленного взгляда – направленного на заданную цель. Подход GazeFormer [22] использует текстовое описание цели на естественном языке, достигая высокой точности при значительно меньшем времени инференса по сравнению с предыдущими подходами. Подходы НАТ (англ. The Human Attention Transformer) [23] и GazeXplain [24] расширяют этот подход путем генерации обоснований на естественном языке, связывая траектории взгляда с причинно-следственными интерпретациями. LookHear (ART) [25] позволяет предсказывать координаты взгляда в реальном времени в процессе устного упоминания целевых объектов.

В связи с этим возникает потребность в построении доменно-специфичных моделей внимания, способных учитывать как визуальные, так и семантические особенности медицинского изображения. Кроме того, большинство существующих моделей целенаправленного внимания были разработаны и обучены на изображениях общего назначения. Их перенос в медицинскую визуализацию сопряжён с рядом проблем. Например, визуальные паттерны медицинских изображений существенно отличаются от естественных сцен (монохромность, отсутствие явных объектов, высокая плотность информации, а патологии могут быть пространственно распределёнными и не иметь чёткой границы (например, «выпот», «инфильтрат», «кардиомегалия»)).

Концепция визуально-языковых моделей позволила эффективно внедрять изображения в большие языковые модели, например, в такие как LLaMa [26]. В этом случае изображение обычно представляется в виде набора эмбеддингов,

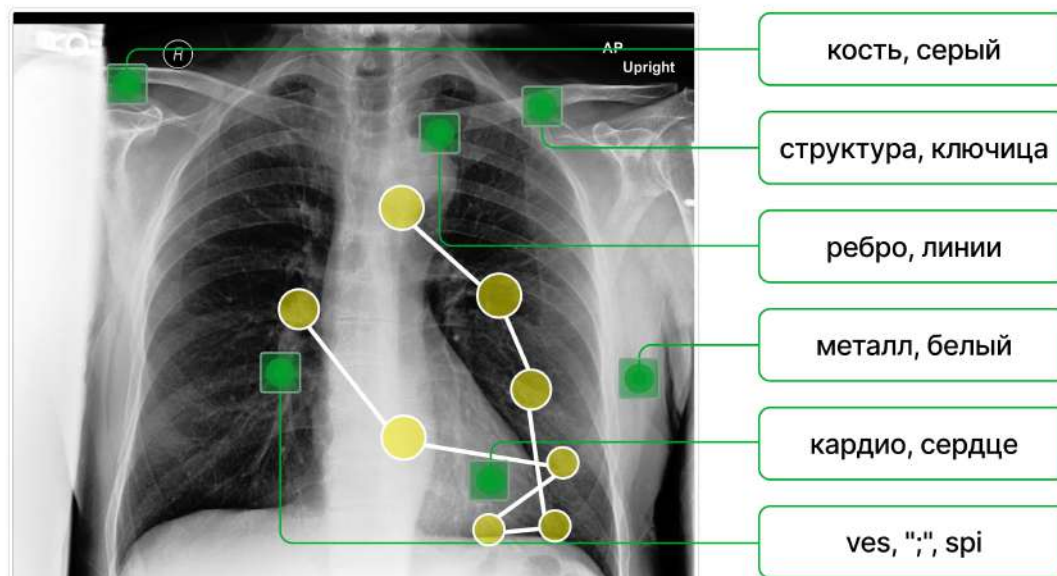


Рисунок 1.1 — Визуализация траектории взгляда на рентгеновском снимке грудной клетки, помеченном как «норма». Жёлтые круги обозначают точки фиксации. Фрагменты изображения декодируются в слова при помощи метода логит-линзы и визуально-языковой модели

извлеченных из фрагментов изображения. Эмбединги изображения проецируются в пространство текстовых токенов (после чего их обычно называют визуальными токенами) и обрабатываются совместно большой языковой моделью. Недавно метод логит-линзы, изначально разработанный для объяснения предсказаний больших языковых моделей, был применен для визуально-языковых моделей [27]. Это позволило сопоставить текстовый токен для каждого визуального токена (см. рисунок 1.1). Таким образом, удалось получить семантическую информацию для каждого фрагмента изображения.

В рамках диссертации **выдвинуто предположение**, что внедрение семантической информации из метода логит-линзы в глубокую модель позволит повысить качество прогнозирования целенаправленного взгляда как на рентгеновских изображениях грудной клетки, так и на общих изображениях. В качестве основы для глубокой модели используется архитектура GazeFormer [22], в которую была интегрирована семантическая информация (см. раздел 1.3).

## 1.2 Постановка задачи

В данной главе рассматривается задача прогнозирования направленного визуального внимания рентгенолога при интерпретации рентгеновского изображения грудной клетки. Формально, она заключается в моделировании последовательности фиксаций взгляда, обусловленной как визуальным содержанием изображения, так и диагностической целью, заданной в текстовой форме.

Каждый объект обучающей выборки определяется тройкой:

- $I \in \mathbb{R}^{H \times W}$  – рентгеновское изображение грудной клетки с высотой  $H$  пикселей и шириной  $W$  пикселей;
- текстовая формулировка  $T$  цели визуального поиска (например, «пневмония», «кардиомегалия» и пр.);
- $G = \{(x_y, y_t, \tau_t)\}_{t=1}^L$  – истинная траектория взгляда, представленная как упорядоченная последовательность координат фиксаций  $(x_t, y_t)$  с длительностью задержки  $\tau_t$  (мс).

Задача заключается в построении модели  $f_w(I, T) = \hat{G}$ , которая аппроксимирует истинную траекторию  $G$  в терминах как координат, так и временных характеристик фиксаций. Здесь  $w$  обозначает параметры модели, подлежащие обучению.

В отличие от традиционного предсказания карт заметности, где моделируется плотность внимания без учёта последовательности, в настоящей постановке требуется генерация временно-упорядоченной структуры, сохраняющей следующие аспекты:

- **Пространственная релевантность:** фиксации должны попадать в клинически значимые области изображения;
- **Временная последовательность:** порядок и длительности фиксаций должны соответствовать паттернам взгляда врача-рентгенолога;
- **Контекстуальная обусловленность:** модель должна учитывать конкретную диагностическую цель, выраженную в тексте;

Модель обучается в режиме автотегрессии, минимизируя следующую составную функцию потерь:

$$\mathcal{L} = \mathcal{L}_{cls} + \lambda_1 \mathcal{L}_{spatial} + \lambda_2 \mathcal{L}_{temporal},$$

где  $\mathcal{L}_{cls}$  – бинарная кросс-энтропия для токена завершения траектории,  $\mathcal{L}_{spatial}$  – L1-потеря по координатам фиксации,  $\mathcal{L}_{temporal}$  – l2-потеря по длительности фиксации,  $\lambda_1, \lambda_2$  – гиперпараметры модели.

Основными целями, поставленными в данной работе, являются:

1. разработка архитектуры LogitGaze-Med, способной учитывать семантическую информацию изображения для генерации правдоподобных траекторий взгляда рентгенолога;
2. исследование влияния доменно-специфичных признаков (медицинских концептов на изображении, извлечённых через метод логит-линзы) на качество предсказаний траектории взгляда;
3. сравнительный анализ с существующими архитектурами общего назначения (GazeFormer, HAT);
4. оценка переносимости модели на немедицинские домены (например, COCO-Search18);
5. подготовка рекомендаций по использованию моделей внимания в задачах медицинской визуализации.

### 1.3 Извлечение семантической информации с использованием метода логит-линзы

Одним из ключевых требований к моделям, применяемым в медицинской визуализации, является их объяснимость. В контексте задач предсказания направленного внимания это означает не только точное воспроизведение траекторий фиксаций, но и возможность обосновать, почему модель выбирает ту или иную область изображения на каждом этапе визуального поиска. С этой целью в архитектуре LogitGaze-Med была реализована интеграция семантических признаков, полученных из визуально-языковой модели с использованием метода логит-линзы. Такая стратегия направлена на достижение двух взаимосвязанных целей: повышение интерпретируемости модели и обеспечение семантической согласованности предсказаний с клинически значимыми областями.

Визуально-языковые модели (англ. Vision–Language Models, сокр. VLM) представляют собой архитектуры, способные одновременно обрабатывать изоб-

ражения и текст, объединяя визуальные и языковые представления в одном латентном пространстве. Наиболее распространённый класс таких моделей построен по принципу «адаптера» (англ. adapter-style architecture), где предварительно обученные компоненты для обработки изображений и текста соединяются через обучаемую промежуточную проекцию.

Рентгеновское изображение грудной клетки  $I \in \mathbb{R}^{H \times W}$  разбивается на фрагменты размером  $P \times P$  пикселей, формируя  $M$  неперекрывающихся фрагментов. Затем фрагменты подаются в визуальный энкодер  $E_{visual}$  (например, CLIP ViT [28]), формируя векторные представления для каждого фрагмента. Результатом является последовательность векторов признаков, соответствующих каждому фрагменту изображения:

$$E_i = E_{visual}(I) \in \mathbb{R}^{M \times d},$$

где  $d$  — размерность признаков. Далее эта последовательность проецируется в пространство входных эмбеддингов  $d_{LM}$  языковой модели  $LM$  с помощью адаптера  $A$  (обычно используется полносвязная нейронная сеть):

$$E_A = A(E_i) \in \mathbb{R}^{M \times d_{LM}}.$$

Полученные векторы  $E_A = \{e_1, \dots, e_M\}$  интерпретируются как визуальные токены — представления визуальной информации, совместимые с входными токенами языковой модели.

Текстовая часть входа  $T$  — например, вопрос или инструкция — токенизируется и преобразуется в эмбеддинги через языковую модель  $LM$ :

$$E_T = L(T) \in \mathbb{R}^{N \times d_{LM}},$$

где  $N$  — количество токенов входного текста.

Полная последовательность входа для языковой модели формируется как конкатенация визуальных и текстовых токенов:

$$X = [E_A; E_T] \in \mathbb{R}^{(M+N) \times d_{LM}}.$$

Визуальные токены представляют собой проецированные признаки фрагментов изображения, поступающие в языковую модель вместе с текстом. Хотя они не соответствуют словам из словаря напрямую, они функционируют как

«мягкие подсказки» (англ. soft prompts), которые позволяют языковой модели обрабатывать визуальный контекст в задачах описания изображений, визуального вопросно-ответного анализа и пр. В настоящем исследовании использовались две визуально-языковых модели: LLaVA [29] и её специализированная медицинская модификация LLaVA-Med [30].

С целью интерпретации того, как визуальная информация обрабатывается внутри визуально-языковых моделей, Клемент и др. [27] применили метод логит-линзы (англ. logit lens) [31], ранее разработанный для анализа языковых моделей. В данной адаптации к визуально-языковым моделям логит-линза позволяет проецировать скрытые состояния визуальных токенов в пространство словаря языковой модели на различных слоях трансформера. Это даёт возможность интерпретировать активации модели как набор текстовых концептов, отражающих внутреннее «понимание» изображения моделью. Например, как показано на рисунке 1.1, фрагменты изображения, соответствующие области средостения, интерпретируются моделью как «сердце», «кардио».

На каждом слое  $l$  и для каждого визуального токена  $i \in \{1, \dots, M\}$  скрытое состояние  $h_i^l$  преобразуется в логит-распределение по словарю  $V$  с помощью весов финального слоя  $W_U \in \mathbb{R}^{d \times |V|}$  (англ. unembedding matrix), после чего извлекается эмбединг токена с максимальным значением логита. Это позволяет декодировать содержимое скрытых слоёв в виде семантически интерпретируемых слов. Таким образом получается набор семантической информации  $E_s \in \mathbb{R}^{M \times d_{lm}}$  для каждого фрагмента рентгеновского изображения.

В исследовании [27] экспериментально показано, что начиная со средних и до поздних слоёв модели визуальные токены всё чаще соответствуют словам, описывающим объекты на исходном изображении. Так, например, токены, соответствующие фрагментам изображения со свитером, декодировались как «pattern», «diam(ond)», «cross»; а в случае изображения с автомобилем — как «wheel», «rub(ber)», «ste(er)ing wheel». Эти результаты подтверждают, что модель постепенно «переводит» визуальные представления в пространство языковых концептов.

Более того, логит-линза позволяет выявить как базовые визуальные категории (например, «dog», «car», «sweater»), так и более тонкие атрибуты, включая счётные категории, цвет, форму, и даже лингвистические особенности (например, декодированные слова на китайском или корейском языках). Это говорит о том, что семантическая информация в VLM развивается по аналогии

с текстовыми трансформерами, несмотря на отсутствие явного предобучения на задаче предсказания следующего токена по визуальному входу.

Описанный механизм логит-линзы [27; 31] позволяет извлекать интерпретируемые концептуальные признаки из скрытых представлений, выявляя потенциально значимые клинические категории (например, «выпот», «консолидация», «кардио») на уровне отдельных фрагментов изображения. Эти векторы далее используются как входные семантические признаки для модели трансформера, направленные на прогнозирование траекторий взгляда в клинически релевантных областях.

## 1.4 Архитектуры нейронных сетей

В основе обеих моделей — LogitGaze (общего назначения) и LogitGaze-Med (медицинская адаптация) — лежит единая архитектура, разработанная по принципу encoder–decoder трансформера. Архитектура реализует парадигму автотегрессивной генерации, аналогично языковым моделям, но применительно к пространственно-временным последовательностям фиксаций взгляда.

Архитектура моделей включает следующие компоненты:

1. визуальный энкодер  $E_{vis}$  — извлекает признаки изображения;
2. семантический энкодер  $E_{sem}$  — извлекает концептуальные признаки каждого патча изображения при помощи логит-линзы. Для LogitGaze используется визуально-языковая модель LLaVA, а для LogitGaze-Med используется LLaVA-Med;
3. текстовый энкодер  $E_{text}$  — кодирует текстовую цель или описание задачи;
4. объединённый embedding-блок — проецирует признаки в общее пространство;
5. transformer encoder — формирует контекстуальные представления;
6. transformer decoder — автотегрессивно предсказывает фиксации взгляда;
7. выходной слой — декодирует координаты  $(\hat{x}_t, \hat{y}_t)$ , длительность  $\hat{\tau}_t$  и токен завершения  $\hat{m}_t$ .

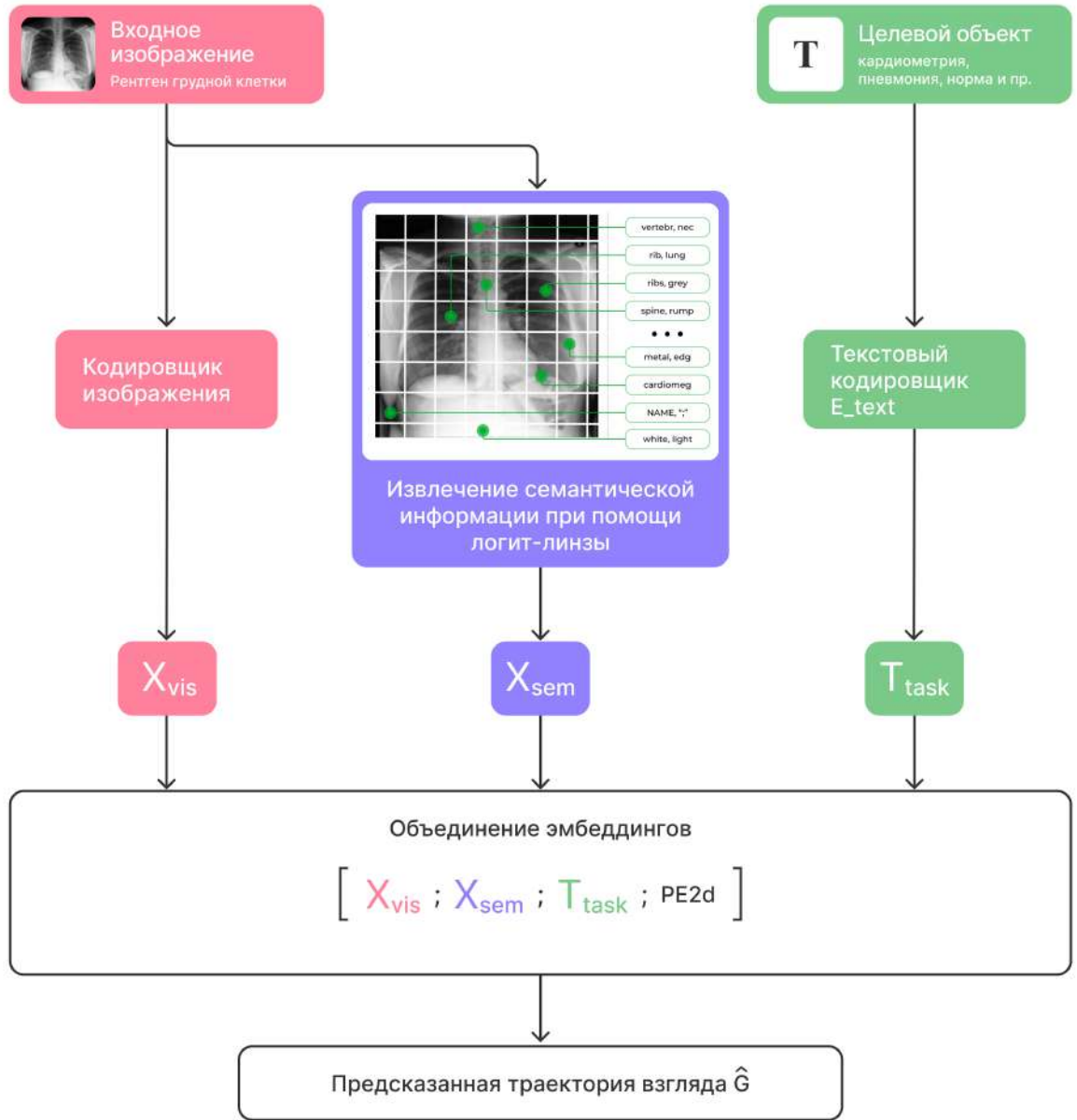


Рисунок 1.2 — Высокоуровневая визуализация архитектуры модели LogitGaze-Med

Высокоуровневая визуализация архитектуры модели LogitGaze-Med показана на рисунке 1.2. Каждое изображение  $I \in \mathbb{R}^{3 \times 768 \times 768}$  пропускается через сверточную сеть  $E_{vis}$  (ResNet50 или CheXNet), формируя тензор признаков  $X_{vis} \in \mathbb{R}^{M \times d}$ , где  $M$  — количество фрагментов изображения,  $d$  — размерность признаков. К этим признакам добавляются семантические эмбединги  $X_{sem} \in \mathbb{R}^{M \times d}$ , извлечённые методом логит-линзы из визуально-языковой модели.

Каждая диагностическая цель  $t \in \mathcal{L}$  кодируется через SentenceTransformer в эмбединг  $e_{\text{task}} \in \mathbb{R}^d$ , после чего реплицируется  $M$  раз и добавляется к каждому визуальному эмбедингу.

Итоговый входной тензор:

$$Z_0 = [X_{\text{vis}}; X_{\text{sem}}; T_{\text{task}}; \text{PE}_{2D}],$$

где  $\text{PE}_{2D}$  – двумерные позиционные эмбединги,  $T_{\text{task}}$  – текстовое описание цели.

В архитектуре Трансформер изначально отсутствует информация о порядке элементов входной последовательности. Чтобы ввести понятие порядка или положения элементов, используются позиционные эмбединги. В задачах компьютерного зрения изображение делится на фрагменты (или патчи). Чтобы сохранить информацию о том, где каждый фрагмент изображения находится, каждому эмбедингу фрагменту добавляется двумерный позиционный вектор (эмбединг):

$$\text{Position}(i, j) = \text{Position}_x(i) + \text{Position}_y(j),$$

где  $i$  и  $j$  – координаты фрагмента изображения по горизонтали и вертикали. Это помогает трансформеру учитывать структуру изображения.

Обучение модели, приведенной выше, осуществляется путём совместной оптимизации трёх компонент:

$$\mathcal{L} = \mathcal{L}_{\text{cls}} + \lambda_1 \mathcal{L}_{\text{spatial}} + \lambda_2 \mathcal{L}_{\text{temporal}},$$

где:

- $\mathcal{L}_{\text{cls}}$  – бинарная кросс-энтропия по токёну завершения;
- $\mathcal{L}_{\text{spatial}}$  – L1-потеря по координатам фиксаций;
- $\mathcal{L}_{\text{temporal}}$  – L2-потеря по длительности фиксации.

Гиперпараметры установлены следующим образом:

- размер эмбединга  $d = 512$ ;
- dropout = 0.4;
- гиперпараметры  $\lambda_1, \lambda_2 \in [0, 1]$  подбираются путем перебора;
- количество слоёв encoder/decoder = 6;
- batch size = 32, обучение 100 эпох;
- optimizer: AdamW.

Модель LogitGaze, ориентированная на общий домен, сохраняет ту же архитектуру, но:

- метод логит-линзы применяется для визуально-языковой модели LLaVa;
- обучается на COCO-Search18;
- использует текстовые цели из бытовой сцены (например, «найди чашку», «ищи газон»);
- не использует доменно-специфичный визуальный энкодер (только ResNet).

## 1.5 Метрики качества предсказания взгляда

Для количественной оценки соответствия предсказанных и реальных траекторий движения взгляда применяются метрики, отражающие различные аспекты зрительного поведения: пространственную точность, временные характеристики, динамику перемещений и структуру траектории в целом. В соответствии с рекомендациями протокола GazeSearch [20] используются четыре основных метрики: **ScanMatch**, **SED**, **STDE** и **MultiMatch**. Ниже представлены их формальные описания и интерпретации.

Метрика ScanMatch [32] реализует выравнивание траекторий взгляда с помощью модифицированной версии алгоритма Нидлмана-Вунша [33], широко применяемого в биоинформатике. Каждая фиксация преобразуется в дискретный символ, отражающий пространственное положение (посредством разбиения изображения на сетку) и порядковый номер во временной последовательности. Вариант метрики с учётом длительности фиксаций (ScanMatch w/ Dur.) включает длительность в вектор признаков; без неё (w/o Dur.) учитывается только пространственно-временной порядок.

Метрика измеряет степень схожести двух последовательностей фиксаций, оценивая количество совпадений, вставок, удалений и замен, необходимых для преобразования одной последовательности в другую. Получаемое значение нормируется по длине последовательностей, обеспечивая сопоставимость между сессиями разной длины.

Метрика Scanpath Edit Distance (SED) [34; 35] является адаптацией расстояния Левенштейна для сравнения траекторий взгляда. Каждая фиксация дискретизируется в зависимости от положения в пространстве (например, по координатной сетке), и сравнение осуществляется на основе операций вставки, удаления и замены. В отличие от ScanMatch, SED игнорирует длительность фиксаций, что делает её менее чувствительной к мелким вариациям, но позволяет сохранять интерпретируемость. Она особенно полезна для оценки последовательностей с разной длиной и при наличии значительных отклонений от эталона.

Метрика STDE [36] предназначена для оценки биологически реалистичных параметров поведения взгляда. В отличие от дискретных подходов, STDE рассматривает траектории как непрерывные кривые в пространстве-времени. Каждая фиксация описывается тройкой  $(x_y, y_t, \tau_t)$ . Сравнение осуществляется через встраивание траектории в трёхмерное пространство и вычисление расстояний между соответствующими точками. Такая оценка учитывает как пространственную точность (координаты фиксации), так и темп зрительного сканирования, отражая биологически правдоподобные закономерности, например, сдвиги, паузы и вариации в длительности фиксаций.

MultiMatch [37] — комплексная метрика, направленная на многомерное сравнение траекторий. Она декомпозирует каждую траекторию на пять непрерывных признаков: форма траектории, угловые отклонения между фиксациями, длина, смещение центра масс и продолжительность фиксаций. Каждое измерение сравнивается отдельно, и результат выражается в виде значения сходства от 0 до 1 (где 1 — полное совпадение). Метрика может быть агрегирована (например, как среднее значение по пяти измерениям) или представлена покомпонентно для более детального анализа расхождений между моделируемым и реальным поведением взгляда.

## 1.6 Эксперименты и сравнительный анализ моделей

Для оценки эффективности предложенной архитектуры LogitGaze-Med и её немедицинского аналога LogitGaze был проведён ряд экспериментов на задачах предсказания траектории визуального внимания в различных доме-

нах: медицинском (рентгенография грудной клетки) и общем (бытовые сцены). В данном разделе представлены количественные и качественные результаты оценки модели LogitGaze-Med на задачах предсказания траекторий направленного взгляда в медицинском домене. Модель сравнивается с существующими подходами, включая GazeFormer [22], HAT [23] и модель ChestSearch [20]. Также включены результаты модели LogitGaze, обученной на общем домене COCO-Search18. Оценка производилась по четырём комплексным метрикам: ScanMatch, SED, STDE и MultiMatch.

Таблица 1 демонстрирует превосходство LogitGaze-Med по всем основным метрикам. Особенно выражено это в метрике ScanMatch (как с учётом, так и без учёта длительности фиксации): модель с энкодером CheXNet достигает 0.419 и 0.330 соответственно, что значительно превышает ближайших конкурентов. Это свидетельствует о том, что предсказанные траектории не только пространственно схожи с реальными, но и воспроизводят правильный временной порядок фиксаций.

Таблица 1 — Значения метрик схожести траекторий взгляда (чем выше, тем лучше для ScanMatch/STDE, чем ниже, тем лучше для SED).

| Метод                       | ScanMatch $\uparrow$                |                                     | SED $\downarrow$                  | STDE $\uparrow$                     |
|-----------------------------|-------------------------------------|-------------------------------------|-----------------------------------|-------------------------------------|
|                             | w/o Dur.                            | w/ Dur.                             |                                   |                                     |
| GazeFormer                  | $0.293 \pm 0.021$                   | $0.201 \pm 0.015$                   | $5.11 \pm 0.08$                   | $0.799 \pm 0.004$                   |
| HAT                         | $0.309 \pm 0.020$                   | —                                   | $5.07 \pm 0.07$                   | $0.800 \pm 0.004$                   |
| GazeSearch                  | $0.332 \pm 0.019$                   | $0.223 \pm 0.014$                   | $4.88 \pm 0.06$                   | $0.809 \pm 0.004$                   |
| LogitGaze                   | $0.328 \pm 0.018$                   | $0.225 \pm 0.015$                   | $5.07 \pm 0.07$                   | $0.810 \pm 0.004$                   |
| LogitGaze-Med (Res)         | $0.416 \pm 0.017$                   | $0.325 \pm 0.012$                   | $4.68 \pm 0.05$                   | $0.852 \pm 0.003$                   |
| <b>LogitGaze-Med (CheX)</b> | <b><math>0.419 \pm 0.016</math></b> | <b><math>0.330 \pm 0.010</math></b> | <b><math>4.68 \pm 0.05</math></b> | <b><math>0.855 \pm 0.003</math></b> |

Метрика SED, измеряющая количество изменений (вставок, удалений и замен) между реальной и предсказанной траекториями, также наименьшая у LogitGaze-Med, что указывает на минимальное расхождение с реальной последовательностью.

STDE демонстрирует, что модель хорошо воспроизводит естественную структуру саккадических движений: более плавные траектории, соответствующие биологически правдоподобным стратегиям визуального поиска, в отличие от резких или неорганизованных предсказаний других моделей.

Таблица 2 содержит покомпонентный анализ по метрике MultiMatch, декомпозирующей траектории на пять измерений. LogitGaze-Med с CheXNet

достигает лучших результатов по четырём из пяти осей: форма траектории, длина, пространственное положение и длительность фиксаций.

Таблица 2 — Сходство MultiMatch (чем выше, тем лучше) по пяти компонентам.

| Метод                       | Vector                              | Direction                           | Length                              | Position                            | Duration                            |
|-----------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| GazeFormer                  | $0.902 \pm 0.008$                   | $0.644 \pm 0.010$                   | $0.899 \pm 0.009$                   | $0.803 \pm 0.007$                   | $0.595 \pm 0.015$                   |
| HAT                         | $0.909 \pm 0.007$                   | $0.649 \pm 0.010$                   | $0.910 \pm 0.008$                   | $0.825 \pm 0.006$                   | –                                   |
| GazeSearch                  | $0.917 \pm 0.006$                   | <b><math>0.679 \pm 0.010</math></b> | $0.917 \pm 0.007$                   | $0.829 \pm 0.006$                   | $0.695 \pm 0.014$                   |
| LogitGaze                   | $0.882 \pm 0.009$                   | $0.643 \pm 0.008$                   | $0.923 \pm 0.005$                   | $0.809 \pm 0.006$                   | $0.625 \pm 0.013$                   |
| LogitGaze-Med (Res)         | $0.935 \pm 0.004$                   | $0.650 \pm 0.008$                   | $0.939 \pm 0.006$                   | $0.823 \pm 0.005$                   | $0.743 \pm 0.010$                   |
| <b>LogitGaze-Med (CheX)</b> | <b><math>0.938 \pm 0.004</math></b> | $0.651 \pm 0.009$                   | <b><math>0.948 \pm 0.005</math></b> | <b><math>0.823 \pm 0.005</math></b> | <b><math>0.740 \pm 0.010</math></b> |

Наиболее интересной является компонент «длительность» (Duration): LogitGaze-Med показывает значение  $0.740 \pm 0.010$ , что значительно выше, чем у GazeFormer ( $0.595 \pm 0.015$ ), и у ChestSearch ( $0.695 \pm 0.014$ ). Это означает, что модель не только предсказывает, куда смотреть, но и насколько долго — важный аспект в клиническом чтении изображений, где длительность фиксации часто коррелирует с диагностической значимостью.

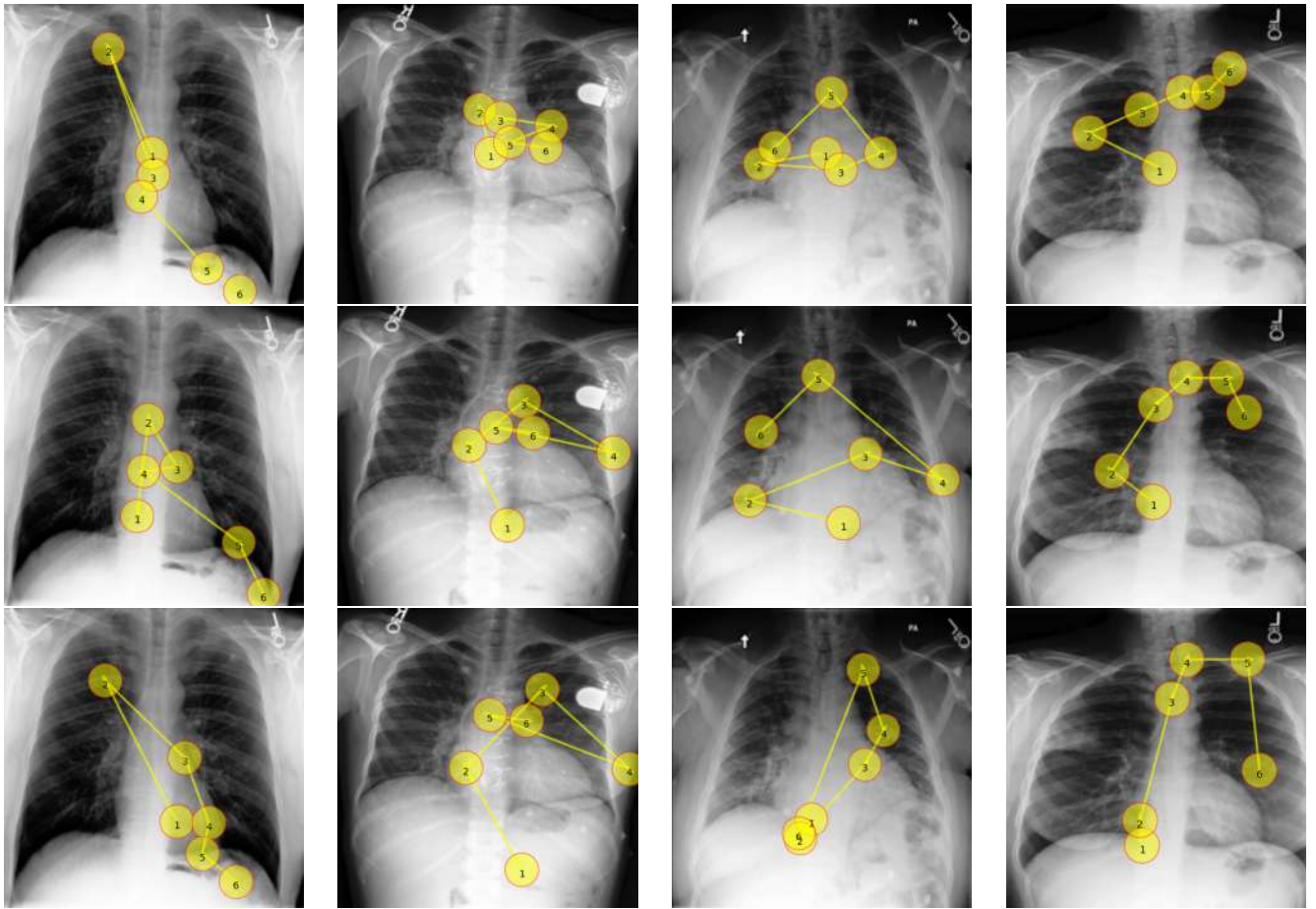
Показатели по «угловому направлению» (Direction), напротив, почти не различаются между моделями. Это говорит о том, что даже простые модели могут примерно воспроизводить ориентацию взгляда, но более сложные — лучше контролируют его длину и логическую структуру.

Рисунок 1.3 иллюстрирует прогнозы LogitGaze-Med и GazeSearch по сравнению с реальными траекториями в четырёх диагностических задачах. Можно видеть, что LogitGaze-Med:

- избегает нерелевантных областей изображения;
- фокусирует внимание на ключевых анатомических структурах;
- демонстрирует согласованное сканирование слева направо и сверху вниз, что характерно для реальных рентгенологов.

Так, при прогнозе взгляда для случая кардиомегалии, модель правильно концентрируется на границах сердца. GazeSearch в этом же примере делает фиксации и в верхних зонах, не имеющих диагностического значения, что снижает интерпретируемость.

Для анализа способности к междоменному обобщению проводилось тестирование моделей на наборе данных COCO18 [38]. Как показано в таблице 3, LogitGaze-Med демонстрирует понижение качества предсказания при переходе на общий домен, что свидетельствует о её специализации на медицинской



а) Без патологии    б) Кардиомегалия    в) Консолидация    г) Пневмония

Рисунок 1.3 — Сравнение путей взгляда человека (верхний ряд), прогнозов LogitGaze-Med (средний ряд) и прогнозов GazeSearch (нижний ряд) по задачам: норма, кардиомегалия, консолидация и пневмония.

визуализации. Наблюдаемый разрыв в производительности подчёркивает критическую важность согласованности между обучающим доменом и областью применения модели. Улучшение не достигается только архитектурными средствами, но требует глубокой интеграции задачи (через текст), семантики (логит-линза) и клинически релевантных визуальных признаков. Результаты указывают на необходимость использования доменно-специфичных данных для обеспечения высокой точности при предсказании траекторий взгляда в медицинских задачах. Модель GazeFormer обучалась на наборе данных COCO18. В случае обучения предложенной модели на наборе данных COCO18 (LogitGaze в таблице), она превосходит все другие модели.

С целью количественной оценки вклада семантической информации в качество предсказания визуального внимания рентенолога было проведено сравнение четырех вариантов моделей:

Таблица 3 — Кросс-доменная оценка на наборе данных COCO18.

| Метод           | ScanMatch<br>w/o Dur | ScanMatch<br>w/ Dur  | MultiMatch<br>Avg ↑  |
|-----------------|----------------------|----------------------|----------------------|
| LogitGaze [39]  | <b>0.527 ± 0.012</b> | <b>0.454 ± 0.013</b> | <b>0.862 ± 0.008</b> |
| GazeFormer [22] | 0.492 ± 0.014        | 0.441 ± 0.015        | 0.816 ± 0.010        |
| LogitGaze-Med   | 0.353 ± 0.018        | 0.289 ± 0.020        | 0.749 ± 0.014        |

Таблица 4 — Влияние семантической информации на предсказание взгляда.

| Метод                               | ScanMatch<br>w/o Dur | ScanMatch<br>w/ Dur  | MultiMatch<br>Avg ↑  |
|-------------------------------------|----------------------|----------------------|----------------------|
| LogitGaze-Med (CheX)                | <b>0.419 ± 0.016</b> | <b>0.330 ± 0.010</b> | <b>0.820 ± 0.009</b> |
| LogitGaze                           | 0.328 ± 0.017        | 0.225 ± 0.020        | 0.776 ± 0.011        |
| Random [25]                         | 0.159 ± 0.014        | 0.148 ± 0.016        | 0.269 ± 0.012        |
| <i>LogitGaze-Med<sub>shuf</sub></i> | 0.178 ± 0.016        | 0.158 ± 0.015        | 0.295 ± 0.013        |

1. **LogitGaze-Med (CheX)** — полнофункциональная модель с использованием визуального кодировщика CheXNet и семантической информацией;
2. **LogitGaze** — базовая версия модели, обученная на COCO-Search18;
3. **Random** — базовое решение, при которой фиксации генерируются случайно;
4. **LogitGaze-Med<sub>shuf</sub>** — модификация основной модели, в которой семантическая информация была случайно перемешана между фрагментами изображениями, тем самым нарушая связь между семантической и визуальной частью.

Результаты представлены в таблице 4. Сравнение LogitGaze-Med с LogitGaze показывает существенный прирост по всем метрикам:

- **ScanMatch (без длительности):** +0.091
- **ScanMatch (с длительностью):** +0.105
- **MultiMatch (среднее):** +0.044

Это подтверждает, что использование семантической информации существенно повышает точность воспроизведения паттернов человеческого взгляда, делая предсказания модели ближе к экспертному поведению.

Ключевое сравнение проводится между полнофункциональной моделью и её версией LogitGaze-Med<sub>shuf</sub>, в которой нарушено соответствие между фрагментами изображения и семантической информацией посредством перемешивания. Результат демонстрирует значительное падение производительности:

- ScanMatch падает более чем в **2.3 раза** ( $0.419 \rightarrow 0.178$ ),
- MultiMatch — почти в **3 раза** ( $0.820 \rightarrow 0.295$ ).

Таким образом, корректно соотнесенная с фрагментами изображения семантическая информация не просто улучшает предсказания, но является необходимым условием для генерации осмысленных траекторий. Нарушение этого приводит к понижению качества до уровня случайного предсказания, даже при сохранении архитектуры и протокола обучения.

Модель с полностью случайными фиксациями (Random) служит нижней оценкой и показывает предсказуемо низкие значения по всем метрикам. Показатели LogitGaze-Med<sub>shuf</sub> лишь незначительно превосходят её, что указывает на отсутствие полезного сигнала в перетасованных семантических признаках. Это подтверждает, что даже небольшие нарушения семантической согласованности критически влияют на способность модели к генерации когерентного взгляда.

Анализ показывает, что:

1. Семантические признаки, извлечённые методом логит-линзы, существенно повышают точность моделирования человеческого внимания;
2. Соотнесение этих признаков с фрагментами изображения играет критическую роль: нарушение этого резко снижает производительность до уровня случайного предсказания.

## 1.7 Обсуждение

Представленная в данной главе модель LogitGaze-Med демонстрирует значительные улучшения в задаче предсказания целенаправленного взгляда рентгенолога при визуальном анализе рентгеновских снимков грудной клетки. Ключевой особенностью подхода является интеграция семантической информации, извлекаемой с помощью метода логит-линзы из визуально-языковой модели LLaVA-Med, в архитектуру трансформера.

Эксперименты показали, что семантические признаки, полученные из LLaVA-Med, вносят существенный вклад в улучшение предсказаний. Перетасовка семантических токенов между патчами приводит к падению качества более чем на 40% по метрике ScanMatch, что указывает на важность структурированного семантического выравнивания между визуальным входом и задачей. Это особенно важно в медицинской визуализации, где семантическая релевантность фрагментов изображения (например, «сердце», «кость», «легкое») определяет диагностическую значимость областей.

Сравнение моделей с различными визуальными энкодерами (ResNet и CheXNet) подтверждает важность использования доменно-специфичных признаков. CheXNet, обученный на медицинских изображениях, позволил достичь более высоких показателей по всем метрикам, включая биологически обоснованную STDE и компоненту длительности фиксации в MultiMatch. Это подчеркивает необходимость использования архитектур, адаптированных к специфике медицинского домена, особенно при моделировании экспертного внимания.

Хотя в рамках данной работы не проводился прямой анализ интерпретируемости, визуализация траекторий (рисунок 1.3) и логит-декодированных концептов (рисунок 1.1) указывают на потенциальную интерпретируемость модели. Предсказанные фиксации совпадают с анатомически и диагностически значимыми областями, а семантические метки, декодированные из скрытых слоев модели, коррелируют с областями внимания рентгенолога. Это открывает перспективу дальнейшей интеграции таких моделей в системы поддержки принятия решений с объяснимыми визуальными подсказками.

Тестирование LogitGaze-Med на наборе данных COCO-Search18 выявило снижение производительности при переходе за пределы медицинского домена. В то же время модель LogitGaze, обученная на общем домене, демонстрирует высокую производительность в условиях визуального поиска на бытовых сценах. Это указывает на ограниченную переносимость моделей, сильно зависящих от семантического контекста, и подчёркивает необходимость калибровки или дообучения при переходе к новым типам изображений.

Важным наблюдением является то, что включение семантической информации увеличивает время инференса модели на видеокарте NVIDIA GeForce RTX 3090 Ti всего на 21% (до 85 мс на изображение), что делает предложенную архитектуру практически применимой в клинических условиях, включая

сценарии интерактивного визуального анализа или симуляции поведения врача в образовательных системах.

Несмотря на высокую точность модели, ряд аспектов требует дальнейшего изучения:

- Необходимо провести количественный анализ, сравнивающий семантические токены с зонами клинического интереса.
- Необходимо учитывать когнитивные особенности поведения рентгенолога, включая влияние патологии, опыта и стиля чтения.
- Метод логит-линзы чувствителен к шуму от фоновых или нерелевантных фрагментов, особенно при отсутствии явных объектов. Введение маскирования или семантической фильтрации может улучшить устойчивость.

Предложенная архитектура LogitGaze-Med задаёт новый уровень качества в задаче предсказания целенаправленного взгляда в медицинских изображениях. Интеграция доменно-специфичных визуальных признаков и семантической информации, извлекаемой через логит-линзу, обеспечивает биологически правдоподобные, клинически интерпретируемые и статистически достоверные траектории визуального поиска. Результаты демонстрируют потенциал использования таких моделей не только в симуляции поведения врача, но и как вспомогательных компонентов в системах принятия решений, обучения и объяснимого ИИ в медицине.

## Глава 2. Внедрение взгляда рентгенолога в глубокую модель классификации патологий легких

### 2.1 Введение

Развитие методов глубокого обучения за последнее десятилетие существенно повлияло на область анализа медицинских изображений, включая автоматическую интерпретацию рентгеновских снимков грудной клетки. Применение сверточных нейронных сетей, трансформеров и мультимодальных моделей продемонстрировало высокий уровень точности в задачах классификации, сегментации и локализации патологий [1; 40]. Особенно значительное расширение практического применения ИИ в медицинской диагностике произошло в период пандемии COVID-19, когда возникла острая потребность в автоматических инструментах поддержки принятия решений врача-рентгенолога. Одним из важнейших шагов в этом направлении стал запуск в Российской Федерации масштабного эксперимента по внедрению технологий искусственного интеллекта в систему здравоохранения [4], охватывающего множество клинических задач, включая интерпретацию лучевых исследований.

Несмотря на впечатляющие количественные результаты, достигаемые современными ИИ-моделями, существует важная проблема, существенно ограничивающая их широкое клиническое применение – это недостаточная интерпретируемость и связанная с ней непрозрачность принятия решений. В области рентгенологии, где диагностические заключения должны быть обоснованы и проверяемы, «чёрный ящик» моделей глубокого обучения вызывает обоснованные сомнения у медицинского сообщества [41]. Неинтерпретируемые ошибки алгоритмов могут привести к утрате доверия со стороны врачей, а также поставить под угрозу безопасность пациентов в случае принятия неверных решений на основе модельных предсказаний.

Более того, исследования показывают, что внедрение ИИ в клиническую практику может сопровождаться когнитивными рисками. В частности, в работе [6] показано, что рентгенологи склонны «следовать» подсказкам системы принятия решений, даже если те ошибочны. Такая избыточная зависимость от рекомендаций ИИ может снижать качество клинической оценки

и усугублять диагностические ошибки, особенно при отсутствии механизмов верификации и объяснения предсказаний. Это подчёркивает необходимость создания ИИ-систем, не только высокоточных, но и поддерживающих контролируемое взаимодействие с пользователем через интерпретируемые механизмы объяснения.

Одним из потенциальных подходов к решению данной проблемы является включение информации о внимании врача, в частности, данных о направлении взгляда рентгенолога при анализе медицинских изображений. Паттерны взгляда представляют собой важный когнитивный сигнал, отражающий область интереса и предполагаемую семантическую значимость регионов изображения [42]. Интеграция такого сигнала в архитектуру модели позволяет не только повысить точность за счёт фокусировки на релевантных областях, но и улучшить объяснимость предсказаний.

Ранее были предложены методы, использующие реальные данные взгляда рентгенолога для локализации областей интереса на КТ-срезах лёгких [43], а также для комбинированного анализа совместно с ИИ-моделью [44]. В рамках одного из исследований [14] был собран и опубликован набор данных, включающий около 1000 рентгеновских изображений грудной клетки из публичной базы данных MIMIC-CXR [45], дополненные информацией о направлении взгляда рентгенолога во время анализа изображений. Авторы исследования использовали глубокую сверточную нейронную сеть для классификации патологий на рентгеновском изображении грудной клетки. Дополнительно к этой нейронной сети был добавлен декодировщик, который предсказывал тепловую карту взгляда рентгенолога на основе скрытого представления изображения, что привело к улучшению качества классификации патологий. Подходы с использованием дополнительной задачи для повышения качества решения целевой задачи активно используются в автоматизированном анализе медицинских изображений. Например, когда при решении задачи сегментации анатомических областей на рентгеновском изображении грудной клетки дополнительно предсказываются контуры этих анатомических областей, позволяя повысить уверенность сети на границах областей регионов, где обычно наибольшая неопределенность в предсказании, приводящая в худшему качеству сегментации [46]. Подобным образом интегрированные данные взгляда в сверточную нейронную сеть позволили повысить качество классификации и улучшить объяснимость предсказаний посредством техник атрибуции.

Однако сбор таких данных в масштабах, необходимых для обучения современных моделей, представляет собой трудоёмкий и дорогостоящий процесс. Помимо привлечения квалифицированных специалистов, он требует технического оснащения и соблюдения условий протоколирования, что делает его непрактичным для большинства исследовательских и клинических учреждений. В связи с этим может оказаться полезным использование синтетически сгенерированных данных взгляда, полученных при помощи моделей прогнозирования внимания.

На основе изложенной информации, в рамках диссертации **выдвинуто предположение** о том, что синтетически сгенерированные данные взгляда рентгенолога, интегрированные в архитектуру сверточной нейронной сети, могут улучшить качество автоматической классификации патологий на рентгеновских изображениях. В главе 2 представлена глубокая модель LogitGaze-Med для предсказания последовательности взгляда на рентгеновском изображении грудной клетки и LogitGaze для предсказания последовательности взгляда на общих изображениях. **Дополнительное предположение** заключается в том, что использование синтетических данных взгляда от специализированной модели LogitGaze-Med будет приводить к лучшему качеству классификации патологий, чем от модели LogitGaze.

## 2.2 Постановка задачи

Рассматривается задача мультиклассовой классификации, в которой каждому рентгеновскому изображению грудной клетки необходимо присвоить один из следующих диагностических классов:

- **норма** – отсутствие выраженной патологии на изображении;
- **застойная сердечная недостаточность (СН)** – признаки сердечной недостаточности, визуализируемые на рентгенограмме;
- **пневмония** – воспалительные изменения лёгочной ткани.

Формально, пусть  $I \in \mathbb{R}^{H \times W}$  – входное рентгеновское изображение размером  $H \times W$  пикселей. Целевой задачей является построение отображения

$$f : \mathbb{R}^{H \times W} \times \mathbb{R}^{m \times 2} \rightarrow \{0, 1, 2\},$$

где  $m$  – количество точек фиксации, а каждое наблюдение из  $\mathbb{R}^{m \times 2}$  представляет собой координаты  $(x_i, y_i)$  фиксации взгляда, и выход  $f(I, \cdot)$  – предсказанный диагностический класс.

Для изучения роли информации о взгляде были реализованы и сопоставлены три различных подхода:

1. **baseline**: классификация осуществляется только по изображению  $I$ , без использования информации о взгляде;
2. **temporal**: используется последовательность тепловых карт, соответствующих фиксациям, дополнительно обрабатываемая через рекуррентную модель;
3. **static**: используется агрегированная тепловая карта по всем фиксациям, которая подаётся как дополнительная входная информация.

Основная гипотеза заключается в том, что синтетически сгенерированные фиксации взгляда рентгенолога позволяют улучшить производительность классификационной модели за счёт дополнительного сигнала, коррелирующего с расположением клинически значимых областей на изображении.

## 2.3 Архитектуры нейронных сетей

**1. Baseline.** В базовой архитектуре изображение  $I$  обрабатывается сверточной нейронной сетью (CNN-энкодером)  $E_I$ , обученной извлекать эмбединг  $X^{img} \in \mathbb{R}^d$  изображения. Далее полученный эмбединг поступает в полносвязный классификационный слой, выдающий вероятности принадлежности к каждому классу:

$$\hat{y} = \text{softmax}(W_c X^{img} + b_c). \quad (2.1)$$

Информация о взгляде не используется. Эта архитектура служит базовым решением для оценки влияния данных взгляда, используемых в других архитектурах, на качество решения задачи классификации.

**2. Temporal.** Архитектура Temporal использует информацию о последовательности фиксаций взгляда. Для каждой фиксации  $k = 1, \dots, m$  создается тепловая карта  $G_k$ , которая передаётся через отдельный CNN-кодировщик  $E_g$ ,

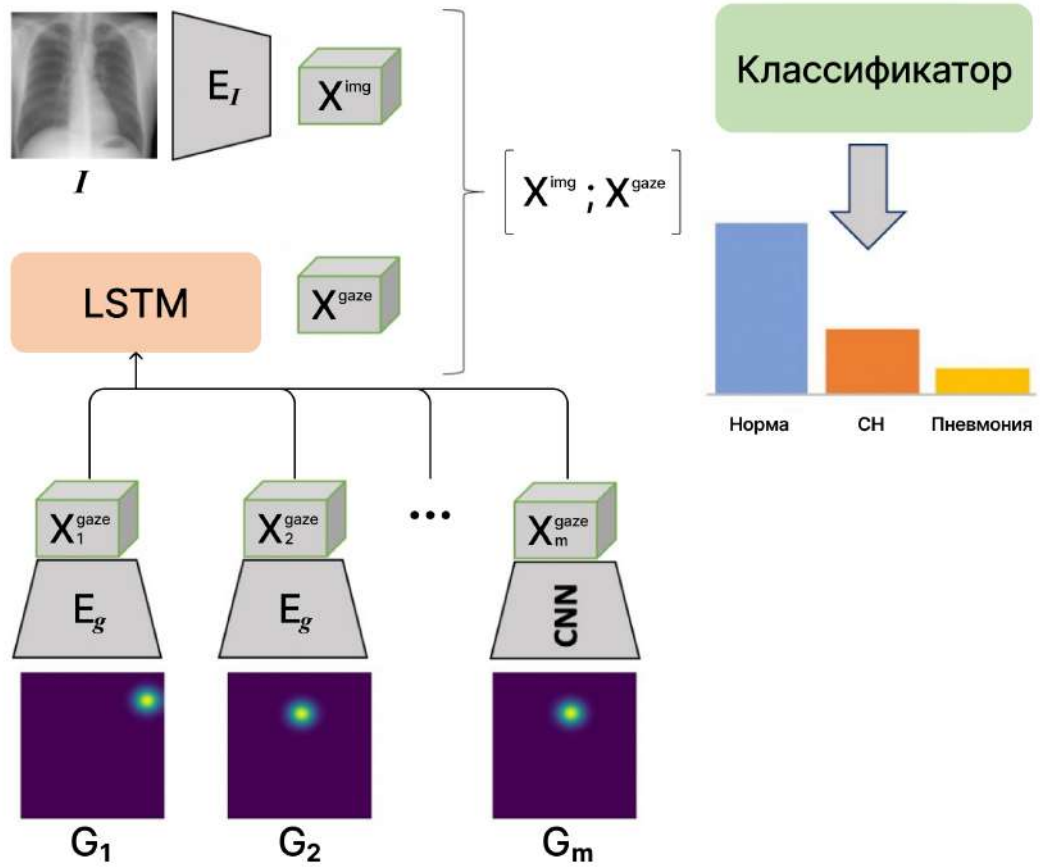


Рисунок 2.1 — Визуализация второго способа (Temporal) для решения задачи классификации патологий

формируя эмбедингг  $X_k^{gaze} \in \mathbb{R}^{d'}$ . Полученная последовательность эмбединггов  $X_1^{gaze}, \dots, X_m^{gaze}$  подаётся в двунаправленную LSTM, извлекающую временные зависимости в паттернах движения взгляда:

$$X^{gaze} = BiLSTM(X_1^{gaze}, \dots, X_m^{gaze}).$$

Далее финальный эмбедингг формируется путём конкатенации выходного вектора LSTM с эмбединггом  $X^{img}$  изображения:

$$Z = [X^{img}; X^{gaze}].$$

Финальный эмбедингг  $Z$  передаётся в классификационный слой. Архитектура визуализирована на рисунке 2.1.

**3. Static.** В этой архитектуре тепловые карты по всем фиксациям усредняются или агрегируются в единую карту  $G_{agg}$ , представляющую обобщённое распределение внимания рентгенолога. Изображение  $I$  обрабатывается энкодером  $E_I$ , как и ранее. После чего эмбедингг  $X^{img}$  изображения подается

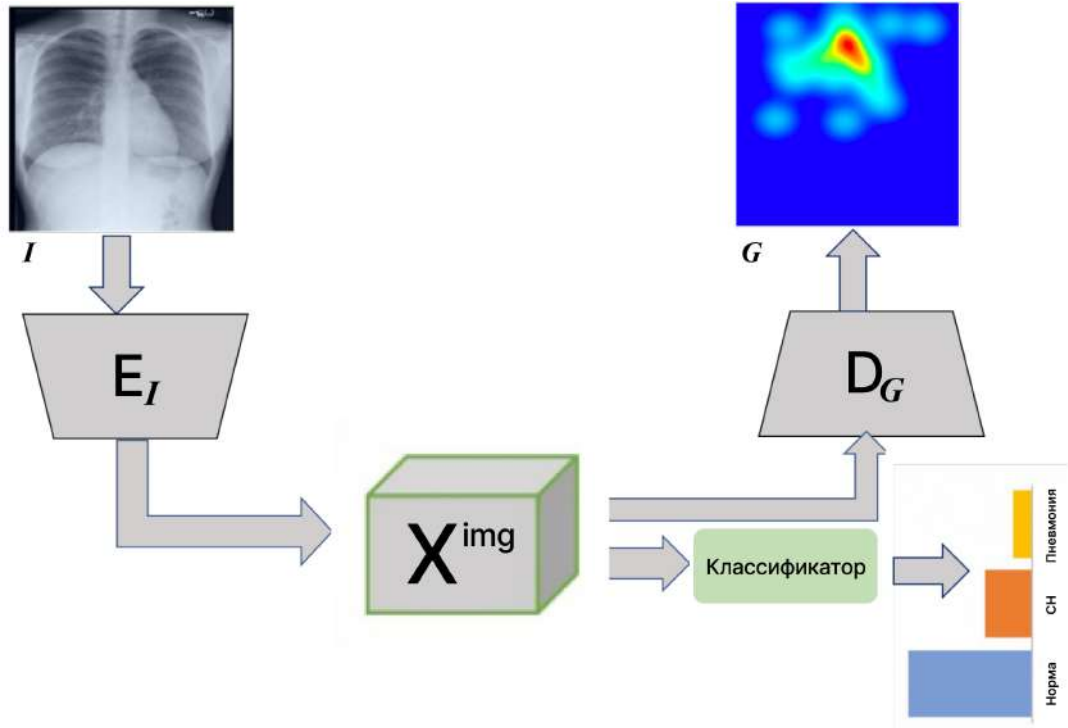


Рисунок 2.2 — Визуализация третьего способа (Static) для решения задачи классификации патологий

параллельно в сверточный декодер  $D_G$  для реконструкции карты взгляда  $G_{agg}$  и в полносвязный слой из выражения 2.1 для предсказания патологии. Такая мультизадачная постановка способствует извлечению признаков, интерпретируемых с точки зрения внимания, и потенциально повышает устойчивость модели к шуму:

$$\mathcal{L} = \mathcal{L}_{cls} + \lambda \mathcal{L}_{heatmap},$$

где  $\mathcal{L}_{cls}$  – кросс-энтропийная потеря классификации,  $\mathcal{L}_{heatmap}$  – ошибка реконструкции карты взгляда,  $\lambda$  – весовой коэффициент. Схема модели приведена на рисунке 2.2.

Для всех трёх архитектур использовался идентичный протокол обучения, соответствующий методологии исходной статьи [14]. В частности:

- Оптимизация производилась с использованием Adam-оптимизатора;
- Размер батча – 32;
- Функция потерь – категориальная кросс-энтропия (для классификации);
- Количество эпох – 50, с ранней остановкой по валидационной метрике;
- Аугментация изображений не применялась;
- Разделение на обучающую и тестовую выборки было фиксированным;

- Все модели обучались на одном и том же наборе изображений, различаясь только по источнику данных взгляда (реальный или синтетический).

Для каждой архитектуры (Temporal и Static) были реализованы версии с использованием синтетических траекторий взгляда, сгенерированных тремя различными моделями: GazeFormer, LogitGaze и LogitGaze-Med. Таким образом, суммарно было обучено шесть моделей, дополняющих базовую конфигурацию без внимания. Это позволяет сопоставить влияние как архитектурного решения, так и качества модели генерации взгляда на итоговую эффективность классификации.

## 2.4 Наборы данных

Для обучения и оценки моделей классификации патологий лёгких использовались два источника данных: (1) масштабный общедоступный набор данных рентгенограмм грудной клетки MIMIC-CXR, используемый для генерации синтетических данных взгляда и (2) специализированный набор представленный Karargyris и др. [14], содержащий реальные записи движений глаз рентгенологов, использованный для тестирования.

Набор данных MIMIC-CXR [45] содержит более 370,000 рентгеновских изображений грудной клетки, сопровождаемых текстовыми заключениями. Для целей настоящего исследования были отобраны все доступные изображения без фильтрации по полу, возрасту, качеству или типу проекции. Такой выбор объясняется стремлением сохранить реалистичное распределение примеров, характерное для реальной клинической практики.

Поскольку сбор эмпирических данных фиксаций взгляда рентгенолога требует специализированного оборудования, программной среды и проведения экспериментов в контролируемых условиях, масштабируемое применение таких данных в медицинских ИИ-системах остаётся затруднительным. Для преодоления этого ограничения в настоящем исследовании использовался подход синтетической генерации траекторий взгляда на основе моделей прогнозирования визуального внимания. Это позволило существенно расширить обучающую выборку без привлечения экспертов и без затрат на отслеживание взгляда.

В эксперименте были применены три модели прогнозирования траекторий визуального внимания, различающиеся по архитектуре и степени адаптации к медицинскому домену:

- **GazeFormer** [22] – трансформерная модель, обученная на общедоменных данных визуального поиска. Она не учитывает специфику медицинских изображений, но широко используется как открытый репрезентативный базовый метод.
- **LogitGaze** – модель, реализующая механизм логит-линзы для интерпретируемого прогнозирования фиксаций. В отличие от GazeFormer, использует внутренние семантические сигналы визуально-языковых моделей (VLM) и задачу регрессии координат фиксаций. Подробное описание этой модели представлено в главе 2.
- **LogitGaze-Med** – модификация LogitGaze, адаптированная к медицинскому контексту. Она обучена на специализированном датасете GazeSearch, содержащем реальные движения глаз рентгенологов в процессе диагностики патологий на рентгенограммах грудной клетки. Подробное описание этой модели представлено в главе 2.

Применение как общедоменных, так и доменно-специфических моделей позволило оценить влияние характера источника взгляда на эффективность классификации патологий.

Каждая модель на вход получает рентгеновское изображение  $I \in \mathbb{R}^{H \times W}$  и генерирует последовательность фиксаций:

$$\{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}, m \in \mathbb{N},$$

где каждая пара координат  $(x_i, y_i)$  соответствует предсказанной точке фиксации взгляда.

Для последующего использования в архитектурах классификации фиксации преобразуются в тепловую карту  $G \in \mathbb{R}^{H \times W}$  с применением двухмерной гауссовой функции, центрированной в каждой точке фиксации. Полученные карты нормализуются по интенсивности в диапазоне  $[0, 1]$ , что обеспечивает согласованный масштаб между различными изображениями и моделями.

Следует отметить, что генерация синтетических траекторий не сопровождается дополнительной калибровкой. В частности:

- не применялась фильтрация фиксаций, выходящих за пределы лёгочных областей или рёберной дуги;

- не проводилась коррекция плотности тепловых карт в зависимости от времени фиксации или приоритетности областей;
- не использовались анатомические маски или сегментации, что могло бы повысить специфичность генерации.

Таким образом, сгенерированные данные могут включать артефакты, нехарактерные для реального поведения врачей. Однако, как показано далее в разделе результатов, даже в таком виде они вносят положительный вклад в задачу классификации, особенно при использовании доменно-специфической модели LogitGaze-Med. Для каждого изображения была сгенерирована последовательность фиксаций, далее преобразованная в тепловую карту. Полученные пары изображение–взгляд образовали синтетическую обучающую выборку для моделей Temporal и Static.

Для финальной оценки модели использовался набор данных, представленный в исследовании Karargyris и др. [14], полученный с использованием оборудования для отслеживания взгляда во время анализа рентгеновских снимков грудной клетки рентгенологами. Набор данных включает:

- около 1000 изображений из набора данных MIMIC-CXR;
- записи движения глаз;
- аудиодиктовку заключений, позволяющую аннотировать временные сегменты, релевантные определённым диагнозам.

В рамках данного эксперимента использовались только изображения и соответствующие фиксации, относящиеся к трём классам: норма, застойная сердечная недостаточность и пневмония. Предварительная фильтрация или балансировка по классам не осуществлялись. Этот набор использовался исключительно как фиксированная тестовая выборка, идентичная для всех сравниваемых моделей, что обеспечило честное сравнение влияния разных источников данных взгляда (синтетических, генерируемых глубокими моделями, и натуральных) на классификацию патологий.

## 2.5 Результаты

В данном разделе представлены результаты количественной оценки моделей классификации патологий лёгких с различными способами интеграции

Таблица 5 — Среднее значение AUROC по всем трем патологиям (норма, застойная сердечная недостаточность и пневмония) для каждой модели.

| Источник взгляда            | Baseline        | Temporal                          | Static                            |
|-----------------------------|-----------------|-----------------------------------|-----------------------------------|
| Натуральные (Eye-Gaze [14]) | $0.77 \pm 0.02$ | $0.82 \pm 0.03$                   | $0.87 \pm 0.02$                   |
| GazeFormer [22]             | $0.78 \pm 0.02$ | $0.84 \pm 0.02$                   | $0.89 \pm 0.01$                   |
| LogitGaze                   | $0.80 \pm 0.01$ | $0.87 \pm 0.02$                   | $0.90 \pm 0.01$                   |
| <b>LogitGaze-Med</b>        | $0.82 \pm 0.01$ | <b><math>0.90 \pm 0.02</math></b> | <b><math>0.91 \pm 0.01</math></b> |

информации о взгляде рентгенолога, включая как реальные фиксации, так и синтетически сгенерированные траектории. Основным критерием оценки выступает среднее значение площади под ROC-кривой (AUROC) по трём классам: норма, застойная сердечная недостаточность (CH) и пневмония.

Результаты представлены в таблице 5. В первую очередь сравниваются три архитектурных варианта: базовая модель без взгляда (Baseline), модель с временной обработкой фиксаций (Temporal) и модель с использованием статической тепловой карты (Static). В качестве источников взгляда рассматривались:

- реальные фиксации рентгенолога (набор данных Eye-Gaze [14]),
- синтетически сгенерированные траектории при помощи моделей GazeFormer, LogitGaze и LogitGaze-Med.

Для всех моделей были вычислены значения AUROC по каждой из трёх патологий, а также средние значения и стандартные отклонения.

Анализ полученных данных позволяет сделать несколько ключевых наблюдений:

1. Синтетические данные взгляда позволяют улучшить качество классификации во всех случаях, в сравнении с подходом [14], в котором используются натуральные данные взгляда.
2. Интеграция взгляда улучшает классификацию. По сравнению с базовой моделью, оба варианта интеграции информации о внимании рентгенолога – как временной (Temporal), так и статической (Static) – приводят к существенному улучшению метрики AUROC. Так, использование реальных данных взгляда позволяет повысить средний AUROC с 0.77 (Baseline) до 0.87 (Static), что свидетельствует о высокой информативности паттернов визуального внимания рентгенолога.

3. Доменно-специфическая генерация предпочтительнее. Сравнение между GazeFormer и LogitGaze-Med подчёркивает важность обучения модели внимания в соответствующем домене. Несмотря на то, что GazeFormer является мощной архитектурой, обученной на широкомасштабных данных общего назначения, её предсказания уступают по качеству тем, что генерируются LogitGaze-Med – моделью, специализированной под рентгеновские изображения грудной клетки и обученной на датасете GazeSearch.
4. Static-модель работает лучше Temporal. Для всех источников данных взглядов вариант Static демонстрирует более высокие значения AUROC, чем Temporal. Это может быть связано с тем, что агрегированная тепловая карта более устойчива к ошибочным или нерелевантным фиксациям, а также проще в обучении. Напротив, Temporal-модель чувствительна к порядку и динамике фиксаций, что требует большей точности синтетического генератора.

Результаты, представленные в данной главе, демонстрируют, что интеграция синтетической информации о направлении взгляда рентгенолога значительно улучшает точность классификации патологий на рентгеновских изображениях грудной клетки. Полученные данные подтверждают выдвинутые гипотезы.

Показанная эффективность синтетических данных взгляда открывает перспективы создания объяснимых и более качественных систем поддержки принятия решений, не требующих дорогостоящего сбора данных взгляда. Такие подходы могут быть особенно полезны в условиях ограниченных ресурсов или в контексте быстрого масштабирования ИИ-сервисов на уровне региональных или федеральных программ здравоохранения.

В дальнейшем исследование может быть расширено в следующих направлениях:

- анализ объяснимости предсказаний с помощью методов атрибуции (Grad-CAM, Integrated Gradients);
- расширение набора задач (например, детекция, сегментация, генерация рентгенологических отчетов) с учётом внимания рентгенолога;

Любой из двух способов интеграции данных взгляда в сверточную нейронную сеть привел к значимому повышению качества классификации. Способ с предсказанием статистической тепловой карты взгляда позволяет добиться

лучшего качества классификации, чем с предсказанием временной тепловой карты. Среди источников взгляда LogitGaze-Med достигает наилучшей производительности во всех случаях.

## Глава 3. Разработка и валидация экспериментальной установки для сбора и анализа данных взгляда в условиях, приближенных к клинической практике

### 3.1 Введение

Настоящая глава посвящена разработке и валидации установки для проведения экспериментов по отслеживанию взгляда рентгенологов во время анализа медицинских изображений. Для проведения валидации установки была выбрана хорошо известная в литературе задача оценки усталости рентгенолога при анализе рентгеновских снимков грудной клетки. Таким образом, задача валидации стала формулироваться как изучение закономерностей взгляда рентгенолога при длительном анализе рентгеновских изображений грудной клетки.

В рамках диссертации разработана экспериментальная установка вместе с программным обеспечением (см. раздел 3.2) для отслеживания взгляда рентгенолога во время анализа рентгеновских снимков грудной клетки. Для участия в эксперименте привлечены 4 практикующих рентгенолога с разным опытом работы. Для того, чтобы компенсировать небольшое количество участников эксперимента и собрать достаточно данных для статистического анализа, был разработан специальный протокол эксперимента (см. раздел 3.3). В рамках эксперимента каждый рентгенолог должен был проанализировать 400 снимков грудной клетки, что заняло более 5 часов для каждого рентгенолога.

Раннее в литературе было показано [47–50], что с увеличением количества просмотренных снимков рентгенологи затрачивают меньше времени на анализ каждого снимка, при этом существуют положительная корреляция между затраченным временем на анализ снимка и качеством диагностики. Например, Берлинг и др. [50] обнаружили, что рентгенологи тратят на 30% меньше времени для анализа последних пяти КТ-снимков брюшной полости, чем первых пяти во время своих рабочих смен.

Причиной такого эффекта может быть усталость рентгенолога, которая является одним из основных источников диагностических ошибок [51]. Проблемы с анализом усталости начинаются с широкого определения усталости как общего чувства утомляемости. Усталость является субъективным чувством и

поэтому ее трудно количественно оценить. Крупински и др. [48] использовали опросник Шведского перечня профессиональной усталости (англ. Swedish Occupational Fatigue Inventory, сокр. SOFI) и подшкалу окуломоторного напряжения из опросника по симуляционной болезни (англ. Simulator Sickness Questionnaire, сокр. SSQ) для оценки усталости рентгенологов. В их исследовании приняли участие 40 рентгенологов, которые просмотрели 60 рентгеновских снимков. Авторы исследования пришли к выводу, что после рабочего дня наблюдалось увеличение симптомов усталости, окуломоторного напряжения и ухудшение качества диагностики. Ханна и др. [52] исследовали влияние ночных смен на качество диагностики рентгенолога. В исследовании приняли участие 12 рентгенологов, каждый из которых просмотрел 20 изображений в течение типичного рабочего дня и утром после ночной смены. После ночных смен наблюдалось ухудшение в опроснике SOFI и качестве диагностики. В этом исследовании использовалось отслеживание взгляда, и было обнаружено, что уставшим рентгенологам требуется больше времени, чтобы найти первую область интереса на рентгеновских снимках. Аналогичное исследование было проведено по интерпретации КТ-изображений рентгенологами с разным опытом, где все рентгенологи демонстрировали худшие баллы SOFI и SSQ, в то время как менее опытные рентгенологи дополнительно демонстрировали значительное снижение диагностической точности [53].

В научном сообществе предполагается, что взгляд рентгенолога содержит информацию о качестве визуального анализа медицинского снимка. Хосп и др. [54] представили подход, основанный на машинном обучении, позволяющий определить опыта хирурга во время артроскопической операции на плече. Тьен и др. [55] ранее пришли к выводу, что менее опытные хирурги имеют большую энтропию зрачка при выполнении когнитивно сложных задач, в то время как эксперты имеют большую частоту фиксаций и тратят больше времени на области интереса. Брунье и др. [56] наблюдали значительные изменения размера зрачка, когда врачи осматривали различные области гистопатологии молочной железы. Аналогичным образом Кастнер и др. [57] зафиксировали увеличенный диаметр зрачка у опытных врачей во время анализа сложных областей, в которых наблюдались изменения зубов по сравнению с менее опытными врачами. Однако небольшие изменения размера зрачка не могут быть точно зафиксированы обычными устройствами отслеживания взгляда и обычно требуют установки специального оборудования. Изменения, связанные с усталостью, также можно

отследить в траекториях взгляда, которые становятся более схематичными и менее сфокусированными [47]. Часто эти изменения не отражаются в числовых показателях, таких как длина траектории взгляда и количество точек фиксации, и поэтому требуют более глубокого анализа траектории взгляда.

На основе изложенной информации в рамках диссертации выдвинуты **следующие предположения** об ожидаемых закономерностях в экспериментальных данных:

- ожидается подтверждение отрицательного коэффициента корреляции Пирсона, обнаруженного ранее исследованиях [47—50], между затраченным временем на анализ снимка и количеством просмотренных снимков;
- ожидается связь между количеством просмотренных снимков и тестами усталости/концентрации, которые ранее были представлены в литературе (см. подробности реализации тестов усталости/концентрации в разделе 3.4;
- ранее было обнаружено, что с увеличением количества просмотренных снимков снижается качество диагностики и взгляд становится менее сфокусированным [47] и требуется больше времени на поиск патологии [52]. Можно предположить, что ухудшение диагностики происходит из-за более поверхностного визуального анализа области потенциального интереса, внутри которой может находиться патология. Поскольку в настоящем исследовании используются рентгеновские снимки грудной клетки, то областью интереса являются легкие. Следовательно, в настоящей диссертации предполагается, что с увеличением количества просмотренных снимков будет снижаться доля покрытия взглядом области интереса (легких) и будет увеличиваться скорость покрытия взглядом области интереса (см. подробности метода в разделе 3.5).

### 3.2 Разработка установки для отслеживания взгляда рентгенолога

Для проведения экспериментов с рентгенологами была разработана установка, имитирующая рабочую станцию рентгенолога для анализа медицинских изображений. Сама установка оснащалась устройством для отслеживания

взгляда. Несколько сеансов калибровки, в ходе которых проверялась общая корректность и надежность установки, показали, что взгляд пользователя слишком часто перемещается на клавиатуру для ввода диагноза и переключения на новый рентгеновский снимок. Это наблюдение побудило переработать установку, добавив в нее запись голоса. Это позволило свести к минимуму использование клавиатуры и мыши. В окончательном варианте установки пользователь диктовал диагнозы, а переход к следующему рентгеновскому снимку осуществлялся нажатием клавиши Enter. Также была реализована возможность изменения яркости и контрастности рентгеновского изображения при помощи компьютерной мыши, что также фиксировалось установкой. Установка записывала и синхронизировала отслеживание движений глаз, голос рентгенолога и команды с контроллеров. Диагностические метки извлекались вручную после эксперимента из голосовых записей. Рабочая станция рентгенолога была установлена в изолированной комнате без постороннего шума. Станция была оборудована 10-битным монитором LG с разрешением  $3840 \times 2160$  пикселей и плотностью пикселей  $\rho = 7.31$  px/mm для отображения рентгеновских изображений. Для отслеживания движений глаз использовался Tobii Eye Tracker 4С с частотой 90 Гц, а для записи голоса – гарнитура Logitech 960. На рабочую станцию была установлена собственная разработка программного обеспечения для проведения экспериментов по анализу рентгеновских изображений с отслеживанием взгляда врача.

На рисунке 3.1 показана разработанная установка. Рентгенолог анализирует медицинское изображение, отображенное на мониторе (см. рисунок 3.1а). При помощи устройства для отслеживания взгляда в режиме реального времени известно куда был направлен взгляд врача на медицинском изображении (см. рисунок 3.1б). Далее взгляд может быть обработан. Например, на рисунке 3.1в показаны области фиксации взгляда в виде кругов, где диаметр круга отображает разброс взгляда внутри области фиксации. Числа на каждом круге обозначают порядковый номер фиксации, а цвет круга обозначает количество точек (чем цвет ближе к красному, тем больше точек взгляда внутри области фиксации). Фиксации соединены саккадами (быстрыми движениями взгляда), которые отображены зелеными линиями.

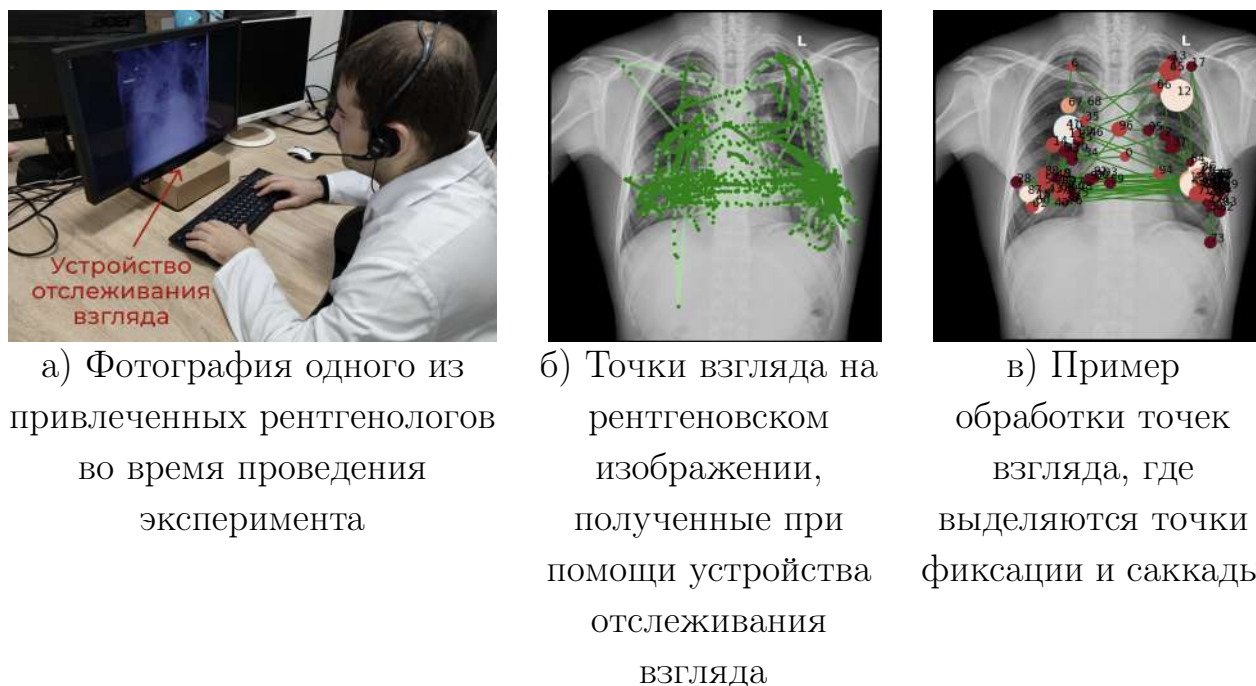


Рисунок 3.1 — Демонстрация установки для отслеживания взгляда рентгенолога и пример обработки данных взгляда

### 3.3 Протокол экспериментов с практикующими рентгенологами

Четыре практикующих рентгенолога были привлечены для участия в эксперименте. Рентгенолог А в своей повседневной клинической практике анализирует как рентгеновские, так и КТ-изображения. Рентгенолог В специализируется на анализе КТ-изображений, в то время как рентгенологи С и D в своей клинической работе анализируют только рентгеновские изображения. Клинический опыт у рентгенологов А, В, С и D составляет 3, 4, 3 и более 30 лет соответственно. Рентгенологи не знали целей эксперимента, чтобы они не пытались изменить свое поведение при визуальном анализе медицинских изображений. Они были проинформированы только о том, что цель эксперимента заключается в анализе их диагностической производительности и записи их движения взгляда. Так же рентгенологи должны были пройти некоторые тесты на оценку концентрации, усталости и реакции во время эксперимента [48; 58; 59]. В течение одного дня каждый рентгенолог проанализировал все 400 рентгеновских снимков, разделенные на четыре партии по 100 снимков, где пропорции патологий и здоровых случаев соответствовали пропорциям во всем наборе данных. Последовательность партий и снимков внутри них были случайным образом перемешаны для каждого рентгенолога.

Каждому рентгенологу было предложено следовать протоколу эксперимента, чтобы стандартизировать процедуру анализа медицинских изображений и минимизировать движения глаз, не связанные с чтением рентгеновских снимков. Начало и завершение анализа каждого рентгеновского снимка контролировались клавишей Enter. Контрастность и яркость рентгеновских снимков можно было регулировать с помощью мыши при необходимости. Чтобы избежать лишнего набора текста, рентгенологи устно аннотировали процесс принятия решения, окончательный диагноз и уровень уверенности (по шкале от одного до десяти) для каждого изображения. Информация была вручную извлечена слушателем из записей голоса врача.

Для каждого рентгенолога эксперимент проводился в выходной день, после полноценного ночного сна. Перед началом эксперимента рентгенолог прошел 10-минутное обучение, чтобы ознакомиться с установкой и программным обеспечением с использованием примерных рентгеновских снимков грудной клетки. Рентгенологу было предложено настроить высоту и наклон кресла для максимального комфорта работы. Технический ассистент обеспечивал, чтобы положение рентгенолога находилось в рекомендуемом диапазоне отслеживания взгляда, который составляет от 50 до 100 сантиметров. Расстояние между пользователем и устройством отслеживания взгляда автоматически отслеживалось программным обеспечением. Специальное звуковое оповещение предупреждало пользователя, если он перемещался к границам рекомендуемого диапазона работы устройства для отслеживания взгляда.

Рентгеновские снимки анализировались партиями. Между партиями проводились измерения усталости, калибровка устройства отслеживания взгляда и короткие перерывы (около 5 минут). После анализа первых двух партий каждый рентгенолог делал 40-минутный обеденный перерыв. Уровень усталости также измерялся в начале и в конце эксперимента. Из-за технических проблем измерения усталости не были правильно записаны для рентгенолога А. Поскольку рентгеновские снимки можно было показать каждому рентгенологу только один раз, не было возможности повторить эксперимент для рентгенолога А. Таким образом, измерения усталости были проанализированы только для рентгенологов В, С и D.

В эксперименте использовалась публичная база данных VinDr-CXR [60], состоящая из 18 000 рентгеновских снимков грудной клетки, собранных в двух больницах Вьетнама. Изображения были получены с помощью различных со-

временных рентгеновских сканеров и имели разные разрешение от  $1624 \times 1775$  до  $3320 \times 3408$  пикселей. Каждое изображение в базе данных было аннотировано тремя рентгенологами из вьетнамских больниц. Всего было выделено 27 различных патологий. Изображение считалось без патологий, если рентгенологи не указали ни одной из них. Из этой базы были отобраны 400 рентгеновских снимков для проведения экспериментов по отслеживанию взгляда рентгенологов. Считалось, что на изображении присутствует определенная патология, если все три рентгенолога, которые размечали снимок в исходной базе данных, её обнаружили. Соотношение здоровых и патологических случаев было обосновано распределением патологий в исходной базе данных [60] и желанием сохранить достаточное количество каждого типа патологии в данных. Состав патологий и демографическая информация, доступная для некоторых снимков, суммированы в таблице 6. Важно подчеркнуть, что один пациент мог иметь несколько патологий или одна патология могла проявляться в нескольких областях изображения.

### 3.4 Методы оценки усталости и концентрации рентгенолога

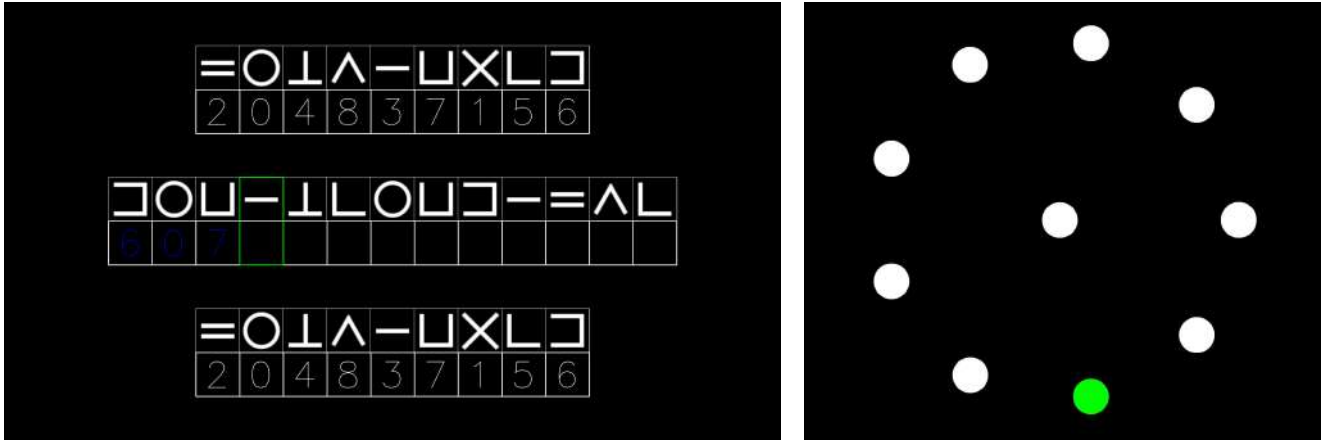
Было реализовано четыре теста для измерения усталости и концентрации. Первый тест был основан на самовыражении усталости с использованием опросника SSQ [48]. Этот тест представляет собой анкету, состоящую из 16 вопросов, сгруппированных по подшкалам офтальмомоторного напряжения, дезориентации и тошноты. Хотя SSQ обычно используется в исследованиях виртуальной реальности, его офтальмомоторная подшкала была использована ранее для оценки усталости в рентгенологии. Например, Крупински и др. [48], Икусима и др. [61] продемонстрировали, что офтальмомоторная подшкала SSQ полезна для оценки усталости рентгенологов и значительно коррелирует с снижением диагностической точности после рабочего дня рентгенолога. Офтальмомоторная подшкала SSQ включает следующие вопросы: общее недомогание, усталость, головная боль, напряжение глаз, трудности с фокусировкой и концентрацией, размытое зрение. Каждый вопрос оценивается участником по четырехбалльной шкале. Таким образом, общий балл теста равен сумме баллов по всем вопросам. Высший балл соответствует большему офтальмомоторному

Таблица 6 — Информация о наборе данных для проведения эксперимента с рентгенологами.

| Характеристика                         | Значение |
|----------------------------------------|----------|
| Количество пациентов                   | 400      |
| Доля пациентов мужского пола           | 61%      |
| Средний возраст (лет)                  | 41       |
| Минимальный возраст (лет)              | 4        |
| Максимальный возраст (лет)             | 90       |
| Доля пациентов женского пола           | 39%      |
| Без обнаруженных патологий             | 168      |
| Кардиомегалия                          | 62       |
| Расширение аорты                       | 53       |
| Утолщение плевры                       | 82       |
| Фиброз                                 | 99       |
| Узелок или массивное новообразование   | 76       |
| Непрозрачность легких                  | 102      |
| Ателектаз                              | 24       |
| Инфильтрация                           | 80       |
| Плевральный выпот                      | 66       |
| Кальцинаты                             | 11       |
| Консолидация                           | 44       |
| Пневмоторакс                           | 49       |
| Прочие заболевания                     | 53       |
| Пациенты с одной патологией            | 52       |
| Пациенты с двумя патологиями           | 35       |
| Пациенты с тремя патологиями           | 51       |
| Пациенты с четырьмя патологиями        | 23       |
| Пациенты с пятью патологиями           | 27       |
| Пациенты с более чем пятью патологиями | 44       |

напряжению. В дальнейшем, при упоминании опросника SSQ будем подразумевать только офтальмомоторную подшкалу SSQ.

Вторая оценка была выполнена с использованием теста на замену цифр символами (англ. Digit Symbol Substitution Test, сокр. DSST) [58]. Участнику предлагалось быстро восстановить последовательность цифр, закодированных графическими символами, используя таблицу соответствия цифр и символов (см. рисунок 3.2а). Метрика производительности в DSST основывалась на отношении точности и затраченного времени.



а) Тест замены цифровых символов (DSST)

б) Тест на времени реакции (RTT)

Рисунок 3.2 — Графические пользовательские интерфейсы для теста замены цифровых символов и теста времени реакции. Эти тесты проводились в определенные временные точки эксперимента по отслеживанию взгляда для оценки уровня концентрации и реакции участвующих рентгенологов

Третья оценка была выполнена с использованием модифицированного теста на покрытие круга взглядом [59]. На экране рабочей станции отображался круг радиусом  $r_{circle}$ . Участнику предлагалось визуально просмотреть контур круга максимально быстро и точно. Метрика оценки определялась как среднее отклонение взгляда от идеальной траектории, нормированное на затраченное время  $T$ :

$$DSST = \frac{\sum_{t=0}^T (|p_t| - r_{circle})^2}{T}, \quad (3.1)$$

где  $p(t)$  — координата взгляда в момент времени  $t$ , нормированная относительно центра круга. Начало координат находилось в центре круга.

Четвертая оценка была выполнена с использованием теста на реакцию (англ. reaction time test, RTT). Десять кругов были расположены в углах и центре

девятиугольника (рисунок 3.26), отображенного на экране рабочей станции. В начале теста центральный круг окрашивался в зеленый цвет, и участнику предлагалось сфокусировать взгляд на этом круге. Каждые 3 секунды текущий зеленый круг становился серым, а новый случайный круг окрашивался в зеленый. Участник должен был перемещать взгляд на зеленые круги как можно быстрее. Тест измерял среднее время, необходимое для перемещения взгляда участника на зеленые круги.

### 3.5 Методы анализа взгляда рентгенолога

Хотя визуальный анализ рентгеновских снимков грудной клетки сильно персонализирован, рентгенолог обычно следует определенным клиническим руководствам. Эти руководства основаны на идее последовательного анализа дыхательных путей, плеврального пространства, кардиомедиастинального контура и костных патологий, что привело к различным правилам, таким как ABCDE и DRSABCDE [62]. Основная идея этих руководств – обеспечить систематическое покрытие всех областей легких, где могут проявляться патологии. В исследовании использовалось глубокое обучение для автоматизации такого анализа покрытия.

Был реализован алгоритм глубокого обучения для автоматической сегментации областей легких на рентгеновских снимках грудной клетки. Для этого использовалась модифицированная архитектура нейронной сети U-Net для сегментации с учетом контуров анатомических областей [63]. Нейронная сеть была обучена предсказывать по входному рентгеновскому снимку грудной клетки две сегментационные маски того же размера, где первая маска соответствовала области легких, а вторая маска – контуру области легких. Пытаясь явно предсказывать как область легкого, так и его контур, ошибки на пикселях вокруг границ легких влияли на функцию потерь U-Net сильнее, чем ошибки на пикселях в центре легочных полей. Поскольку корректная идентификация пикселей на границах более существенна для общего качества сегментации, то кодировщик U-Net был заменен на 50-слойную сеть ResNeXt [64], предварительно обученную на базе данных ImageNet [65]. Функция потерь  $\mathcal{L}$  представляла собой комбинацию бинарной кросс-энтропии

$$BCE(X, Y) = - \sum_{p \in P} [Y_p \ln X_p + (1 - Y_p) \ln(1 - X_p)] \quad (3.2)$$

и коэффициента Дайса

$$DSC(X, Y) = \frac{2 \sum_{p \in P} X_p Y_p}{\sum_{p \in P} X_p^2 + \sum_{p \in P} Y_p^2}, \quad (3.3)$$

где  $X$  – это сегментационная маска, предсказанная глубокой моделью,  $Y$  – соответствующая истинная сегментационная маска, а  $P$  – набор индексов пикселей в масках  $X$  и  $Y$ .

Для обучения глубокой модели использовались публично доступные рентгеновские снимки грудной клетки с ручной сегментацией [66; 67]. Была применена аугментация с помощью поворотов до  $\pm 15^\circ$ , сдвигов до  $\pm 10\%$  и масштабирования до  $\pm 20\%$  как описано в работе [68]. U-Net был обучен с использованием оптимизатора Adam [69] с начальной скоростью обучения 0.001 и размером батча из 16 изображений.

Сегментированные области легких в сочетании с данными взгляда позволяют количественно оценить покрытие легких взглядом при визуальном анализе рентгеновских снимков. Для оценки покрытия сначала нужно определить угла обзора  $\theta$ , позволяющий рассчитать размер области изображения, воспринятого взглядом в каждый момент времени. Хотя универсального значения  $\theta$  нет, некоторые исследования оценивали восприятие информации на медицинских изображениях в зависимости от расстояния между патологией и точкой взгляда [70]. Недавно Wolfe и др. [71] показали, что у рентгенологов есть 33% вероятности переместить взгляд на патологию молочной железы, если она находится внутри угла обзора 2 градуса. Это подтверждается Stransburger и др. [72], которые продемонстрировали, что люди способны воспринимать объекты в угле обзора 1.5-2 градуса от точки фокусировки. Соответственно, в настоящем исследовании было положено  $\theta = 2^\circ$ .

Область  $A_t$  воспринимаемой визуальной информации в точке взгляда  $g(t)$  в момент времени  $t$  можно рассчитать как:=

$$A_t = \{(x, y) : \sqrt{(x - x_t)^2 + (y - y_t)^2} \leq z_t \rho \operatorname{tg}\left(\frac{\theta}{2}\right)\}, \quad (3.4)$$

где  $x_t, y_t \in g_t$  – координаты взгляда на мониторе,  $z_t \in g(t)$  – расстояние в миллиметрах между участником эксперимента и координатами взгляда,  $\theta$  – угол

обзора,  $\rho$  – плотность пикселей монитора, а  $x, y \in I$  – абсолютные координаты изображения  $I$  в пикселях.

Площадь покрытия легких  $C_\tau^{lung}$ , рассчитанная за период времени  $\tau$  от начала визуального анализа рентгеновского снимка, равна

$$C_\tau^{lung} = \frac{|\bigcup_{t=0}^\tau (A_t \cap L)|}{|L|}, \quad (3.5)$$

где  $L$  – множество пикселей, принадлежащих области лёгких, а  $|L|$  обозначает размер множества  $L$ . На рисунке 3.3 представлена информация о покрытии взглядом, которая позволяет количественно оценить область легких, наблюдаемую врачом в каждый момент времени анализа изображения.

### 3.6 Результаты экспериментов

Для каждого просмотренного рентгеновского снимка были рассчитаны следующие характеристики: 1) затраченное время (прошедшее между началом и окончанием анализа); 2) длина траектории взгляда; 3) средняя скорость движения взгляда; 4) доля покрытия области легких; 5) доля покрытия рентгеновского снимка; 6) доля некорректных точек отслеживания взгляда, например, когда взгляд врача направлен не на экран или глаза закрыты; 7) частота моргания; 8) уверенность в диагнозе; 9) средний информационный прирост. Информационный прирост (лёгких) определяется как угловой коэффициент линейной модели между долей покрытой области легких взглядом и затраченным временем для каждого момента  $t$ . Аналогичным образом определяется информационный прирост (рентгена), но вычисляется не доля покрытой области легких, а для всего изображения. Коэффициенты корреляции Пирсона между характеристиками и совокупной работой (порядковым номером рентгеновского снимка в анализе), приведены в таблице 7.

Для каждой партии снимков (то есть 100 рентгеновских снимков грудной клетки) была рассчитана производительность диагностики рентгенологов как среднее значение F-меры, которое представляет собой гармоническое среднее между точностью (англ. precision) и полнотой (англ. recall) и широко использу-

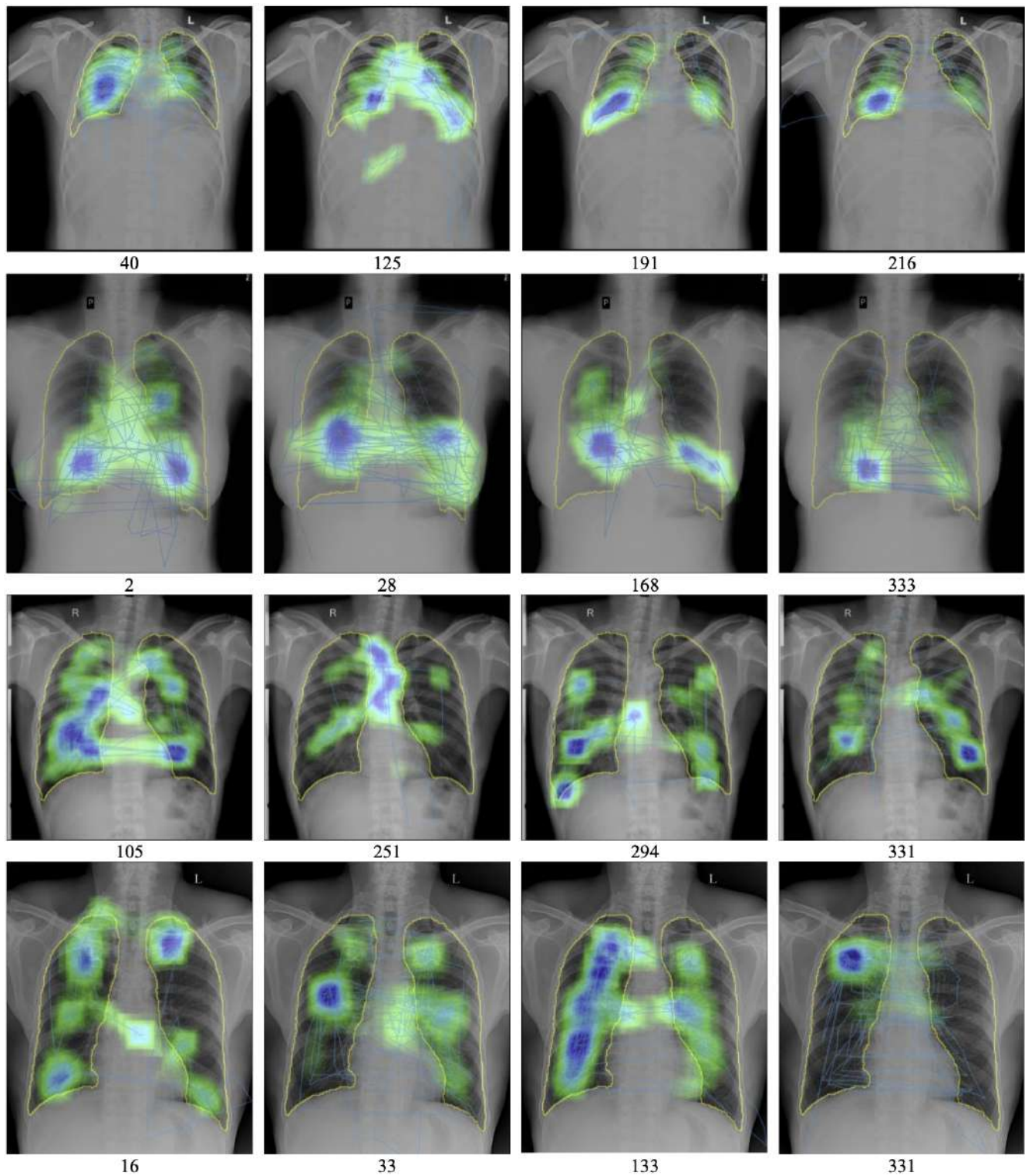


Рисунок 3.3 — Сравнение тепловых карт взгляда четырех рентгенологов, наложенных на четыре изображения грудной клетки. Цифры под каждым рентгеном указывают порядок, в котором этот снимок был показан рентгенологу, чей взгляд наложен на изображение. Каждая строка содержит один снимок, показанный каждому рентгенологу в разные моменты времени эксперимента.

Желтым цветом выделен контур легких

Таблица 7 — Коэффициенты корреляции Пирсона между характеристиками из данных взгляда и порядковым номером снимка во время анализа каждым рентгенологом. Признаки рассчитываются с использованием информации о взгляде рентгенолога и/или сегментации рентгеновского снимка на основе предсказания глубокой модели. Каждая ячейка таблицы содержит коэффициент корреляции человека с доверительным интервалом 95%. Статистически значимые коэффициенты корреляции выделены зеленым цветом.

| Характеристика                            | Рентгенолог |       |       |       |
|-------------------------------------------|-------------|-------|-------|-------|
|                                           | A           | B     | C     | D     |
| Затраченное время                         | -0.16       | -0.52 | -0.42 | 0.23  |
| Количество некорректных точек взгляда     | 0.04        | -0.20 | -0.20 | -0.03 |
| Доля некорректных точек взгляда           | 0.27        | -0.10 | 0.21  | -0.03 |
| Длина траектории взгляда                  | -0.08       | -0.54 | -0.44 | -0.25 |
| Средняя скорость взгляда                  | 0.27        | 0.10  | 0.13  | -0.08 |
| Уверенность в диагнозе                    | -0.01       | -0.05 | -0.10 | 0.15  |
| Доля покрытия рентгена взглядом           | 0.00        | -0.22 | -0.53 | -0.28 |
| Доля покрытия области лёгких взглядом     | -0.20       | -0.12 | -0.52 | -0.28 |
| Средний информационный прирост (рентгена) | -0.04       | -0.26 | -0.53 | -0.35 |
| Средний информационный прирост (лёгких)   | -0.20       | -0.21 | -0.52 | -0.31 |

ется для оценки качества классификации

$$F = \frac{2 * precision * recall}{precision + recall}, \quad (3.6)$$

где  $precision = \frac{TP}{TP+FP}$  отражает долю верно классифицированных объектов (англ. true positive, TP) среди всех объектов выбранного класса (сумма количества верно классифицированных объектов TP и ложноположительных объектов FP, англ. false positive), а  $recall = \frac{TP}{TP+FN}$  отражает отношение верно классифицированных объектов TP выбранного класса к общему числу объектов этого класса (сумма количества верно классифицированных объектов TP и ложноотрицательных объектов FN, англ. false negative).

Диагноз для каждого рентгеновского снимка был вручную извлечен из записей голоса рентгенологов. Если присутствовало несколько патологий, рентгенологам рекомендовали упомянуть все из них при принятии решения, но

просили выбрать одну патологию в качестве окончательного диагноза. Анализ рентгеновского снимка считался правильным, если окончательный диагноз рентгенолога из эксперимента соответствовал одной из эталонных патологий из набора данных или если рентгенолог из эксперимента, так и эталонная группа рентгенологов, считают рентгеновский снимок нормальным. F-мера с 95% доверительными интервалами, рассчитанные для каждого рентгенолога, представлены на рисунке 3.4.

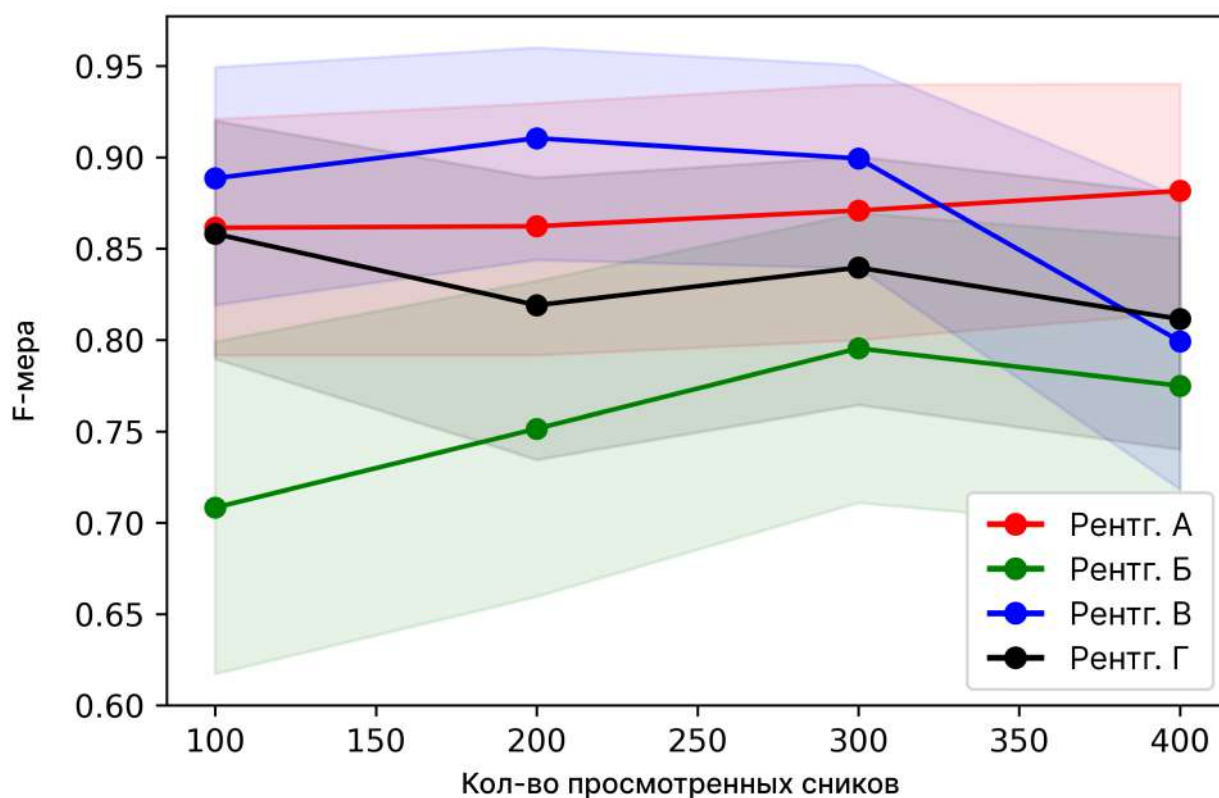
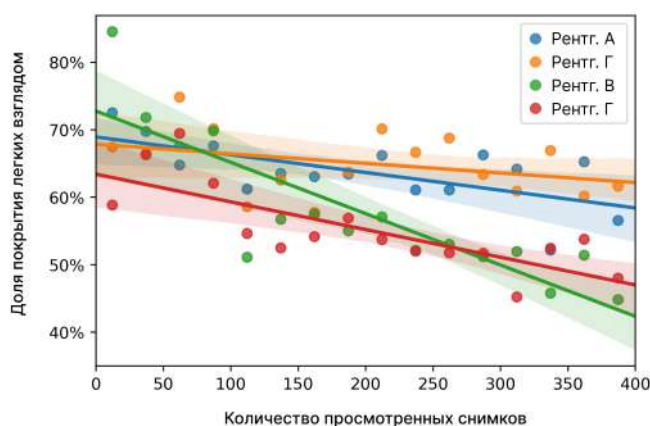


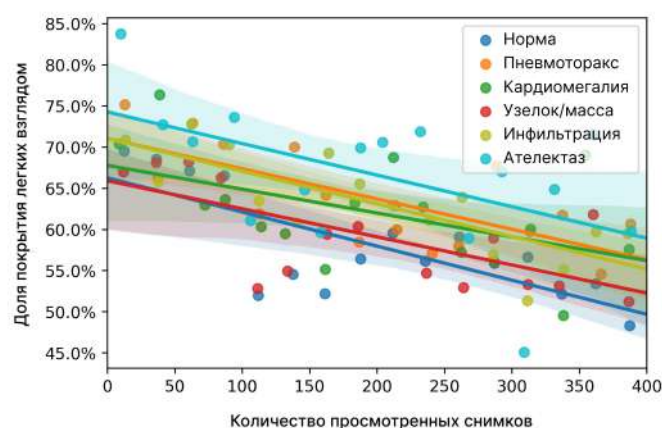
Рисунок 3.4 — Диагностическая точность рентгенологов, рассчитанная для каждой партии из 100 рентгеновских снимков. Производительность трех из четырех рентгенологов улучшилась после обеденного перерыва (200 → 300). Производительность трех из четырех рентгенологов ухудшилась на последних 100 рентгеновских снимках (300 → 400)

Среднее время анализа медицинского изображения составило 24.4 секунды, 35.1 секунды, 21.8 секунды и 25.1 секунды для рентгенологов А, В, С и D соответственно. Время чтения значительно коррелировало ( $p\text{-value} < 0.01$ ) с длиной траектории взгляда, имея коэффициенты корреляции Пирсона 0.93, 0.87, 0.92 и 0.74 для рентгенологов А, В, С и D соответственно.

Среднее покрытие области легких составило 64%, 65%, 58% и 55%, а 32%, 29%, 30% и 30% для всего рентгеновского снимка для рентгенологов А, В, С и D соответственно. Среднее ( $\pm$  стандартное отклонение) покрытие области легких составило  $64 \pm 16\%$ ,  $65 \pm 14\%$ ,  $58 \pm 17\%$  и  $55 \pm 17\%$  для рентгенологов А, В, С и D соответственно. Линейные регрессионные модели, построенные на данных покрытия области легких, имели угловые коэффициенты  $-0.026$ ,  $-0.013$ ,  $-0.076$  и  $-0.04$  для рентгенологов А, В, С и D соответственно. Иными словами, рентгенолог С покрывает на 1% меньше легочных полей после каждого 13-го прочитанного снимка ( $1/0.076 \approx 13$ ). Рисунок 3.5 показывает регрессионные модели с доверительными интервалами. Линейные регрессионные модели также были вычислены для каждого типа патологии (рисунок 3.56). Коэффициенты наклона составили  $-0.041$ ,  $-0.041$ ,  $-0.035$ ,  $-0.034$ ,  $-0.034$  и  $-0.041$  для ателектаза, инфильтрации, пневмоторакса, узелков/массивных образований, кардиомегалии и здоровых рентгеновских снимков грудной клетки соответственно.



а) Модели линейной регрессии, примененные к метрике охвата легких, рассчитанной для каждого из четырех рентгенологов, участвовавших в эксперименте



б) Модели линейной регрессии, примененные к данным охвата взглядом для всех рентгенологов, рассчитанные для каждой патологии

Рисунок 3.5 — Рисунок демонстрирует, как изменяется охват легких взглядом рентгенологов в зависимости от количества изображений, проанализированных рентгенологами. Модели линейной регрессии накладываются на соответствующие точки данных; точки усреднены по 20 выборкам данных для лучшей наглядности

Измерения усталости, рассчитанные для рентгенологов В, С и D, были сравнены с эталонными значениями проделанной работы (порядковым номером

снимка в анализе) и диагностической производительностью. Тест SSQ содержит семь вопросов, каждый из которых может быть оценен от 1 до 4, где более высокие оценки соответствуют большему дискомфорту, потенциально связанному с усталостью. Рентгенологи В, С и D оценили уровень усталости как 7 в начале теста, то есть полностью без усталости, и как 15, 8 и 13 соответственно в конце эксперимента (рисунок 3.6). Производительность DSST была оценена как среднее время, необходимое для замены одного символа, деленное на среднюю точность замены (рисунок 3.6). Нижние значения DSST соответствуют более быстрой и точной замене. Результаты теста на покрытие круга были рассчитаны как среднее расстояние между взглядом рентгенолога и контуром круга (рисунок 3.6). Большие расстояния указывают на большее отклонение от оптимальной траектории. Тест RTT оценивал среднее время, необходимое для перехода к следующему выделенному кругу на экране (рисунок 3.6).

Результаты теста SSQ были записаны для рентгенологов В–D, в то время как для рентгенолога А они не были записаны из-за программной ошибки. Используемая офтальмомоторная подшкала теста SSQ имеет диапазон от 7, указывающий на самый низкий уровень усталости из всех подчастей теста, до 28, указывающий на максимальный уровень усталости. Рентгенологи В–D не сообщили об усталости в начале эксперимента; то есть они оценили все семь офтальмомоторных измерений SSQ наименьшим возможным баллом «1». После анализа 200 рентгеновских снимков средний уровень усталости вырос с 7 до  $11 \pm 1.4$ . Обеденный перерыв вернул уровень усталости к почти исходному уровню  $7.7 \pm 0.5$ . В конце эксперимента средний уровень усталости составил  $12.7 \pm 2.1$ .

Используя средний наклон линейных моделей, построенных на данных отдельных рентгеновских снимков, произведена оценка того, превышает ли покрытие областей легких взглядом для каждого рентгенолога среднюю тенденцию или нет. Использовался порядковый номер рентгеновского снимка для расчета ожидаемого покрытия взгляда и сравнивался его с наблюдаемым покрытием для каждого рентгенолога. Обнаружено, что рентгенолог А превышает тенденцию в 64.5% случаев; то есть он покрывает большую часть легочных полей, чем ожидалось от среднего рентгенолога. Рентгенолог В превышал тенденцию в 66.3% случаев, рентгенолог С – в 42% случаев, а рентгенолог D – в 38.5% случаев.

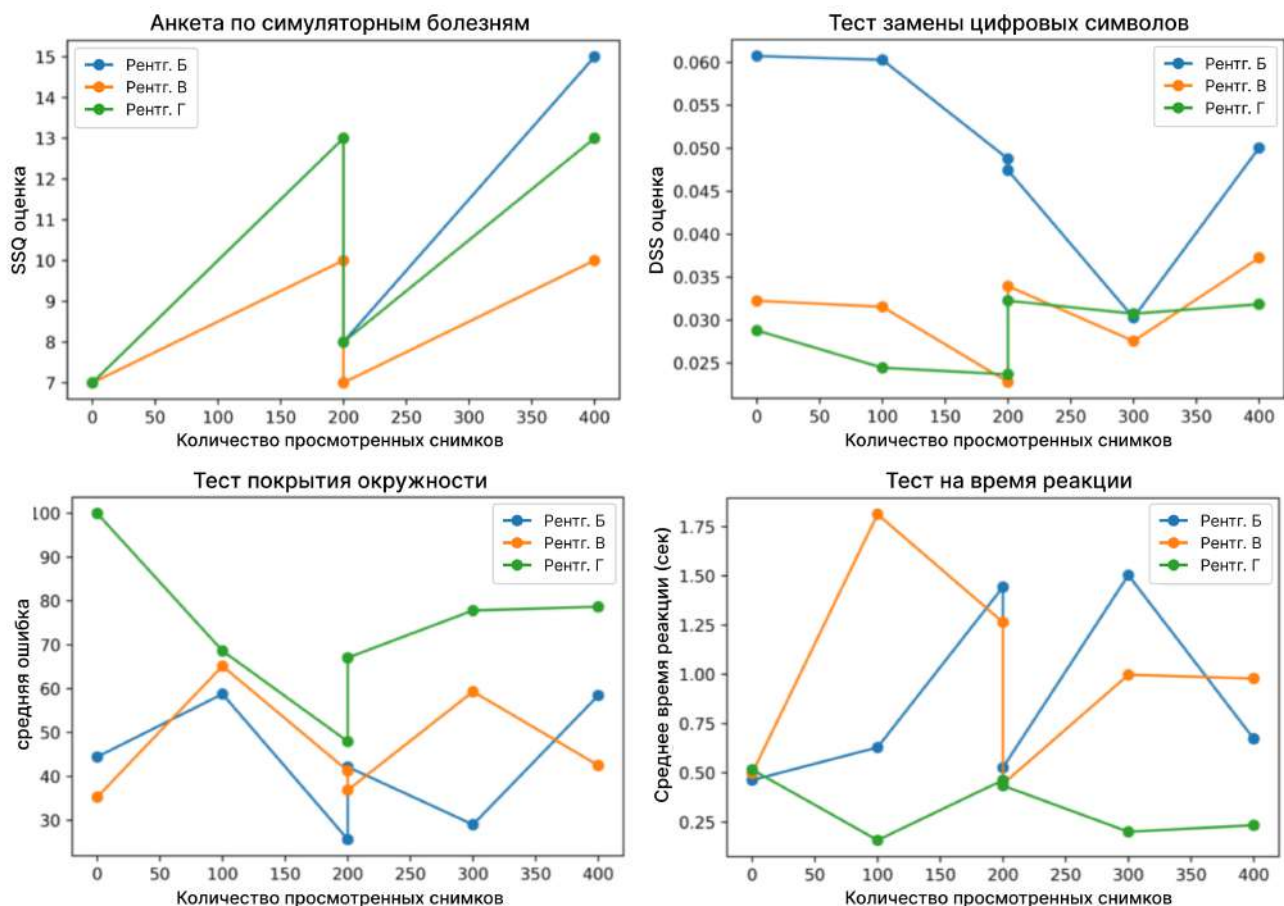


Рисунок 3.6 — Результаты опросника симмуляционной болезни (слева сверху), замены символов цифр (справа сверху), охвата круга (слева внизу) и времени реакции (справа внизу), рассчитанные для рентгенологов В, С и D, участвовавших в эксперименте. Тесты проводились в начале эксперимента и после каждых 100 проанализированных рентгеновских снимков для измерения изменений в утомляемости, реакции и концентрации. Дополнительная серия тестов была проведена после обеденного перерыва после анализа 200 рентгеновских снимков

### 3.7 Обсуждение

Усталость рентгенолога является одним из основных источников диагностических ошибок [51; 52]. Из-за присущей ей субъективности усталость часто оценивается с помощью опросников, заполняемых самостоятельно, таких как SOFI и SSQ. Хотя результаты таких опросников в целом согласуются с ожидаемым ростом усталости после рабочего дня в радиологии [48], у них есть существенные недостатки. Во-первых, анкеты недостаточно детализиро-

ваны, чтобы отразить непрерывность роста усталости. Во-вторых, результаты анкетирования крайне субъективны и не позволяют надежно сравнивать рентгенологов. Вторым недостатком проявился в представленном в настоящей главе исследовании, где рентгенолог С сообщил о минимальном росте баллов SSQ по сравнению с рентгенологами В и D, чья самооценка усталости SSQ удвоилась к концу эксперимента (рисунок 3.6). В отличие от самооценки усталости, диагностическая эффективность рентгенолога С упала к концу эксперимента, тогда как у рентгенолога В она улучшилась. Важно отметить, что рентгенологи получали рентгеновские снимки в другом порядке, поэтому рентгенолог В мог получить «более легкие» рентгеновские снимки в конце.

Поскольку анкеты крайне субъективны, участники могли невольно корректировать свои ответы в соответствии с ожидаемой усталостью среднестатистического человека. В эксперименте использовались три альтернативных теста, измеряющих концентрацию и аккуратность, которые коррелируют с усталостью и использовались для анализа усталости [58; 59; 73; 74]. Результаты теста на покрытие кругом ухудшились для рентгенологов В и С с ростом количества просмотренных снимков. Результаты теста покрытия кругом рентгенолога D также ухудшились, за исключением измерения до эксперимента. DSST и RTT показали непоследовательные результаты. Вопреки ожиданиям, рентгенолог В улучшил свои результаты по замене цифр в DSST, а рентгенолог D – по RTT. Такие наблюдения могут указывать на то, что рентгенологи улучшают результаты теста на концентрацию за счет повышения навыка прохождения этого текста, которая компенсирует ухудшение результатов, связанных с усталостью. Потенциальный эффект повышения навыка подтверждается тем фактом, что все три рентгенолога улучшили свои результаты при второй попытке DSST, а двое из трех рентгенологов улучшили свои результаты при второй попытке RTT. Можно предположить, что при второй попытке рентгенологи уже хорошо ознакомились с целями теста, но еще не устали от анализа рентгеновских снимков.

Значение затраченного времени на анализ снимка имело статистически значимую отрицательную корреляцию с количеством просмотренных снимков для всех рентгенологов в эксперименте. В настоящем исследовании время чтения рентгеновских снимков в среднем сокращалось на 0.02, 0.09, 0.08 и 0.03 сек/рентген для рентгенологов А, В, С и D соответственно. Несмотря на статистически значимую отрицательную корреляцию между временем анали-

зом снимка и количеством просмотренных снимков для всех рентгенологов, время анализа не является достаточным предиктором усталости. Проблема заключается в высокой изменчивости среднего времени анализа снимка между рентгенологами, которое варьируется от 21.8 сек для рентгенолога С до 35.1 сек для рентгенолога В. Следовательно, время чтения для одного или нескольких рентгеновских снимков для произвольного рентгенолога не может предсказать степени усталости без предварительно собранных знаний о том, как быстро этот рентгенолог анализирует снимки, когда он/она свежий и уставший. Расстояние, пройденное взглядом, сильно коррелирует с временем чтения (коэффициенты корреляции составляют 0.87-0.93), поэтому его закономерности и недостатки схожи. Частота недействительных точек взгляда положительно коррелировала с количеством просмотренных снимков, т. е. рентгенологи чаще отвлекают свое внимание от рентгеновского снимка в конце эксперимента, чем в начале. Уверенность в диагнозе незначительно колебалась у разных рентгенологов.

В отличие от пройденного расстояния, доля покрытия взглядом рентгена и легких не является производной от затраченного времени на анализ снимка. Рентгенолог может потратить много времени, рассматривая небольшую область изображения, что не увеличит покрытие изображения взглядом. Доля покрытия легких отрицательно коррелировала с количеством просмотренных снимков для рентгенологов А, С и D, в то время как для рентгенолога В значимой корреляции не наблюдалось. Используя время чтения снимка, пройденное расстояние взглядом, скорость взгляда и долю покрытия легких взглядом, можно реконструировать схему анализа снимка рентгенологами, когда они устают. Рентгенолог А начинал тратить меньше времени, но старался быстро охватывать изображения взглядом. Продолжая охватывать рентгеновские снимки на том же уровне, рентгенолог А охватывал легкие значительно меньше с ростом количества просмотренных снимков. Это наблюдение говорит о том, что его анализ изображений становится менее сосредоточенным со временем. Скорость взгляда для рентгенологов В, С и D слегка менялась в ходе эксперимента, что, учитывая сокращение времени анализа снимка, привело к уменьшению доли покрытия рентгена взглядом с ростом количества просмотренных снимков. Рентгенолог С продемонстрировал наиболее существенное падение доли покрытия рентгена и области легких, что сопровождается самым высоким относительным падением производительности среди всех рентгенологов.

Информационный прирост оказался самым надежным предиктором количества просмотренных снимков в эксперимента. Характеристика, рассчитанная для области легких, показала статистически значимую корреляцию с количеством просмотренных снимков для всех рентгенологов. Судя по коэффициентам корреляции, рентгенологам нужно меньше информации для принятия решения в конце эксперимента, чем в начале. Одним из ключевых преимуществ информационного прироста является то, что он объединяет время анализа снимка, скорость взгляда и покрытие объекта (всего рентгена или области легких) взглядом в одно значение. Такая метрика может смягчить индивидуальные закономерности анализа рентгеновских снимков, наблюдаемые у уставших рентгенологов: более быстрые, но неточные движения взгляда (рентгенолог А), более быстрое считывание с малым (рентгенолог В) или большим (рентгенолог С) количеством недействительных точек времени взгляда и более быстрое считывание с более медленными движениями взгляда (рентгенолог D).

Исследование показало, что сегментация анатомических структур на основе глубокого обучения на медицинских изображениях, в частности сегментация области легких на рентгеновских изображениях грудной клетки, может быть расширена дальше классических приложений компьютерной диагностики. В эксперименте сегментационные маски области легких использовались для количественной оценки покрытия легких взглядом и были сопоставлены с количеством просмотренных снимков каждым рентгенологом. Такая обработка данных взгляда может стать существенным для дальнейшей оценки эффективности рентгенологов. Несмотря на корреляции с количеством просмотренных снимков для некоторых тестов на концентрацию, они не продемонстрировали значительной предсказательной силы. Этот факт не означает, что тесты неприменимы, но необходим более подробный анализ, чтобы исключить потенциальные факторы улучшения производительности из-за повышения навыка прохождения теста.

Представленное исследование имеет несколько ограничений, связанных с контролируемой обстановкой эксперимента. В клинической практике рентгенологи делают нерегулярные перерывы во время анализа снимков, тратят время на клинические заметки, консультируются с коллегами и т.д. Поэтому эксперимент не является по-настоящему клиническим. В то же время для надежного сравнения между участниками эксперимента была необходима стандартизированная среда. Другим ограничением является то, что были выбраны снимки с

патологиями, которые единогласно диагностировали командой рентгенологов, которые создали общедоступную базу данных рентгеновских снимков грудной клетки. Хотя такое единогласие предполагает высокое качество и надежность истинных меток о заболеваниях, оно также может указывать на то, что эти случаи могут быть проще и быстрее диагностированы, чем случаи с неясными диагнозами.

## Глава 4. Интерактивная сегментация медицинских изображений на основе взгляда

### 4.1 Введение

В клинической практике решения на основе глубокого обучения оказывают наиболее существенное влияние в сегментации медицинских изображений [75]. Сегментация медицинских изображений является фундаментальным и важным процессом для идентификации, выделения и визуализации областей интереса на медицинских изображениях [76—78]. Эта задача особенно актуальна для планирования хирургических операций и лучевой терапии, где сегментация играет ключевую роль [79]. При планировании лучевой терапии необходимо оконтурить большое количество анатомических областей пациента на объемном медицинском изображении (КТ или МРТ-снимке) [80; 81]. Внедрение решений на основе глубокого обучения в клиническую практику значительно облегчает задачу оконтуривания, которая может занимать несколько часов работы врача для каждого отдельного случая [82]. Однако неточности в сегментации часто требуют тщательной ручной коррекции. Последние достижения в области интерактивной сегментации улучшили эффективность процесса коррекции в различных областях, включая медицину, позволяя вносить изменения всего несколькими щелчками мыши [83—85].

Методы сегментации медицинских изображений включают ручную, полуавтоматическую и автоматическую сегментации. Ручная сегментация является трудоемкой, подвержена несоответствиям между разными экспертами и имеет низкую надежность, что ограничивает ее применение в клинических условиях [86—88]. В то же время автоматическая сегментация успешно применяется в МРТ для различных медицинских задач [89—92]. Однако эти автоматические методы часто не проходят полную валидацию и в основном ограничены академическими исследованиями из-за возможных неточностей сегментации. Например, архитектура U-Net [63] отлично справляется с сегментацией крупных структур брюшной полости, но сталкивается с трудностями при работе с мелкими органами и сложными структурами, такими как почки, поджелудочная железа и кровеносные сосуды [93; 94]. Аналогично, хотя модель nnU-Net [95]

предоставляет хорошую основу для задач сегментации, она требует дополнительного дообучения для узких клинических приложений, чтобы повысить его эффективность [92]. Полуавтоматическая сегментация показала себя как более надежный вариант для КТ и МРТ. По сравнению с полностью автоматическими методами, полуавтоматическая сегментация дает результаты, которые, как правило, более точны и предпочтительны клиницистами благодаря повышенной прозрачности процесса сегментации [87; 96; 97].

Современные автоматические методы сегментации на основе глубокого обучения, в сочетании с такими методами как отслеживание взгляда, значительно повышают интерактивность, эффективность и отзывчивость инструментов сегментации в медицинской области [98—105].

Интерактивная сегментация относится к подходу, который позволяет пользователям контролировать предсказания глубокой модели для улучшения процесса сегментации. В частности, пользователи предоставляют начальную сегментационную маску при помощи кликов мыши, выделением ограничивающих рамок или текстовыми описаниями области интереса. Затем они могут уточнять предсказанные результаты в несколько итераций [106]. Начиная со времен, когда решения на основе глубокого обучения являлись малораспространенными, интерактивная сегментация развивалась на основе графового представления данных [107] и случайных блужданиях [108], до современных фундаментальных глубоких моделей [109; 110].

Модель Segment Anything Model, или SAM [109] является современной фундаментальной глубокой моделью интерактивной сегментации. SAM демонстрирует впечатляющую производительность в различных задачах, требующих взаимодействия с человеком. Ее архитектура включает кодировщик изображений на основе Vision Transformer [111], декодировщик масок и кодировщик запросов, которые обрабатывают входные данные взаимодействия, такие как координаты кликов мыши, ограничивающие рамки, маски и т.д. Хотя SAM показал многообещающие результаты на естественных изображениях, его производительность в клинических приложениях менее устойчива [112]. Чтобы решить эту проблему, MedSAM [84] был дообучен на большом наборе медицинских данных и специально адаптирован для сегментации медицинских изображений с использованием ограничивающих рамок. Обе модели, SAM и MedSAM, могут быть дообучены для повышения точности сегментации в конкретной клинической задаче. Например, модель Med-SA [113]

использует адаптеры, сохраняя остальные параметры SAM замороженными. Кроме того, различные формы запросов для сегментации медицинских изображений могут привести к существенному повышению качества результата. ScribblePrompt [114] использует ограничивающие рамки, клики и штрихи в качестве запросов для генерации масок сегментации, включая модель nn-Unet [92] для предсказания высококачественных сегментационных масок. Недавно был представлен Medical SAM2 [115], который основан на дообучении SAM2 [116] на медицинских данных и способен предсказывать трехмерные сегментационные маски сразу для всего объемного изображения, в отличие от SAM [109] и MedSAM [84], которые позволяют решать только задачу сегментации двумерных изображений.

В интерактивной сегментации были исследованы различные модальности пользовательских запросов. В отличие от существующих подходов, использующих коррекции пользователя через ограничивающие рамки [84; 117], координат кликов мышью [118] или штрихи [114], данные взгляда предоставляют важные сигналы для обучения, которые необходимы для медицинских изображений, характеризующихся нечеткими границами и низкой контрастностью. Более того, сегментация с использованием взгляда продемонстрировала значительное увеличение эффективности, работая примерно в два раза быстрее, чем традиционные методы, основанные на кликах [119]. Обеспечивая мгновенные коррекции на основе визуального внимания пользователя, сегментация на основе отслеживания взгляда может стать более эффективной альтернативой традиционным методам ввода данных [11].

Технология отслеживания взгляда стала ценным инструментом в области медицины, предоставляя широкий спектр применений в различных дисциплинах [120], особенно в сегментации изображений. Один из заметных методов был предложен Садеги и др. [99]. В этом методе используются данные взгляда для интерактивной пользовательской сегментации. Этот метод продемонстрировал значительное ускорение процесса по сравнению с традиционным управлением мышью.

Насколько известно, существует только одна модель интерактивной медицинской сегментации с использованием взгляда, а именно GazeSAM [119]. Это недавнее исследование использовало оригинальный SAM для генерации сегментационной маски на основе одной точки взгляда. Авторы провели сравнение между сегментацией на основе взгляда и традиционной сегментацией на основе

кликов, показав, что подход с использованием взгляда был примерно в два раза быстрее, хотя и с немного меньшим качеством, чем метод с использованием кликов мыши. Несмотря на эти результаты, исследователи не углублялись в эффекты использования нескольких точек взгляда, оставляя неопределенность в отношении результатов, когда взгляд смещается от области интереса — что вероятно происходит из-за естественных флуктуаций во взгляде.

На основе изложенной информации, в рамках диссертации **выдвинуты следующие предположение:**

1. набор координат взгляда рентгенолога может использоваться в качестве запроса для модели интерактивной сегментации в режиме реального времени;
2. обучение модели возможно на синтетических данных взгляда таким образом, чтобы сократить разницу между естественным и синтетическим источником взгляда во время использования модели.

**В рамках диссертации** представлен способ дообучения Segment Anything Model (SAM) [84; 109] для адаптации к использованию данных взгляда врача в качестве запроса на основе использования синтетических данных взгляда во время дообучения. Этот подход особенно может быть полезен для сложных способов визуализации, включая КТ и МРТ снимки, которые состоят из большого множества срезов, а не из одного изображения [75; 99]. На рисунке 4.1 показан общий процесс коррекции сегментационных масок на КТ-изображении с использованием взгляда на основе разработанного подхода.

## 4.2 Рабочая станция рентгенолога

Представленная в главе 1 рабочая станция рентгенолога была модифицирована для возможности отображения срезов объемных медицинских изображений. В программное обеспечение было внедрено разработанное решение для интерактивной сегментации на основе взгляда и для интерактивной сегментации на основе выделения ограничивающих рамок вокруг области интереса. Программа позволила последовательно отображать срезы объемных медицинских изображений, записывать координаты ограничивающих рамок и точек взгляда с временными метками. У разработанного программного обеспе-

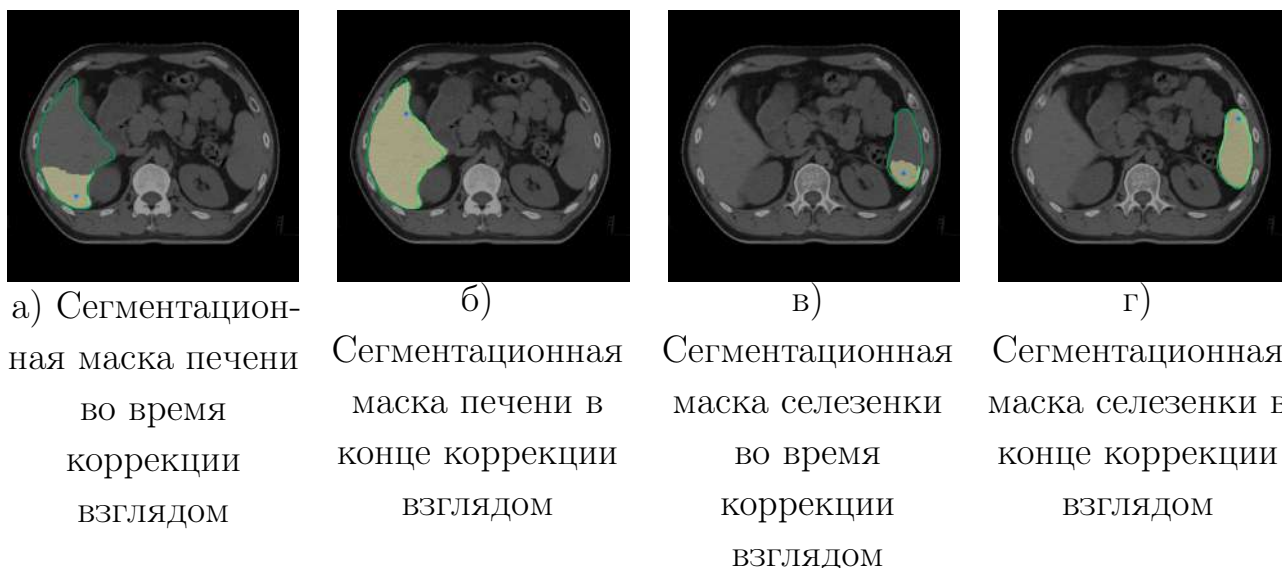


Рисунок 4.1 — Процесс коррекции сегментационных масок для органов брюшной полости, таких как печень и селезенка, с визуализацией контуров эталонной маски сегментации (зеленая), прогнозируемой маски сегментации (желтая) и точек взгляда (синяя) на КТ-изображении

чения графический интерфейс начинался с главного экрана, который знакомит пользователя с основными управляющими кнопками и органом для сегментации. Все операции управляются интуитивными командами:

1. нажатие *Enter* отображает КТ-срез и запускает процесс сегментации. Пользователи сначала сегментируют анатомический орган, выделяя ограничивающие рамки, при необходимости повторяя эту операцию;
2. нажатие *Space* активирует сегментацию с использованием взгляда;
3. пользователи могут включать/выключать визуализацию точки взгляда и предсказанной маски сегментации с помощью клавиш *Shift* и *Alt* соответственно;
4. для улучшения видимости небольших областей, таких как надпочечники или пищевод, можно активировать функцию увеличения нажатием *Space*;
5. частота обновления предсказанной маски может быть изменена через конфигурационный файл, который также позволяет настраивать количество точек взгляда и другие параметры эксперимента;
6. для отмены результата сегментации текущего КТ-среза и повторного запуска процесса с использованием отслеживания взгляда используется клавиша *Backspace*;
7. результат сегментации сохраняется при нажатии клавиши *Enter*.

### 4.3 Постановка задачи

Объемное медицинское изображение  $X \in \mathbb{R}^{H \times W \times D}$  состоит из  $D$  срезов размером  $H \times W$ , полученное с помощью компьютерной томографии. В результате обработки из каждого изображения  $X$  извлекается набор из  $D$  двумерных срезов размером  $H \times W$ . Каждый срез подается в сегментационную модель  $M$  и в результате получается объемное изображение  $\hat{X} \in \mathbb{R}^{H \times W \times C}$ , содержащее сегментационные маски для  $C$  органов.

Сегментационная модель  $M$  основана на интерактивной сегментации, которая использует данные взгляда пользователя в качестве запроса для коррекции сегментационных масок. Данные взгляда состоят из координат  $(x_t, y_t)$ , соответствующих текущему просматриваемому срезу, записываемых в каждый момент времени  $t$ . Цель — улучшить сегментационную маску в реальном времени с помощью последовательности точек взгляда  $G_t = \{(x_i, y_i)\}_{i=0}^t$ ,

$$Q(M(X, G_t), GT) > Q(M(X, G_{t'}), GT),$$

где  $Q$  обозначает метрику качества (более высокие значения указывают на лучшее качество),  $GT$  — референсная сегментационная маска из базы данных WORD,  $t' < t$ .

### 4.4 Метод сегментации с использованием данных взгляда

В работе представлен подход для интерактивной сегментации медицинских изображений с использованием взгляда, который показан на рисунке 4.2. В этом разделе представлены детали дообучения MedSAM — фундаментальной модели сегментации медицинских изображений, а также интеграции данных взгляда в модель.

В работе использована предобученная модель MedSAM для интерактивной сегментации, которая включает кодировщик запросов в виде ограничивающих рамок, как показано на рисунке 4.2. В процессе дообучения параметры кодировщика изображений MedSAM не изменялись, так как он уже эффективно кодирует медицинские изображения и был обучен на большом наборе данных.

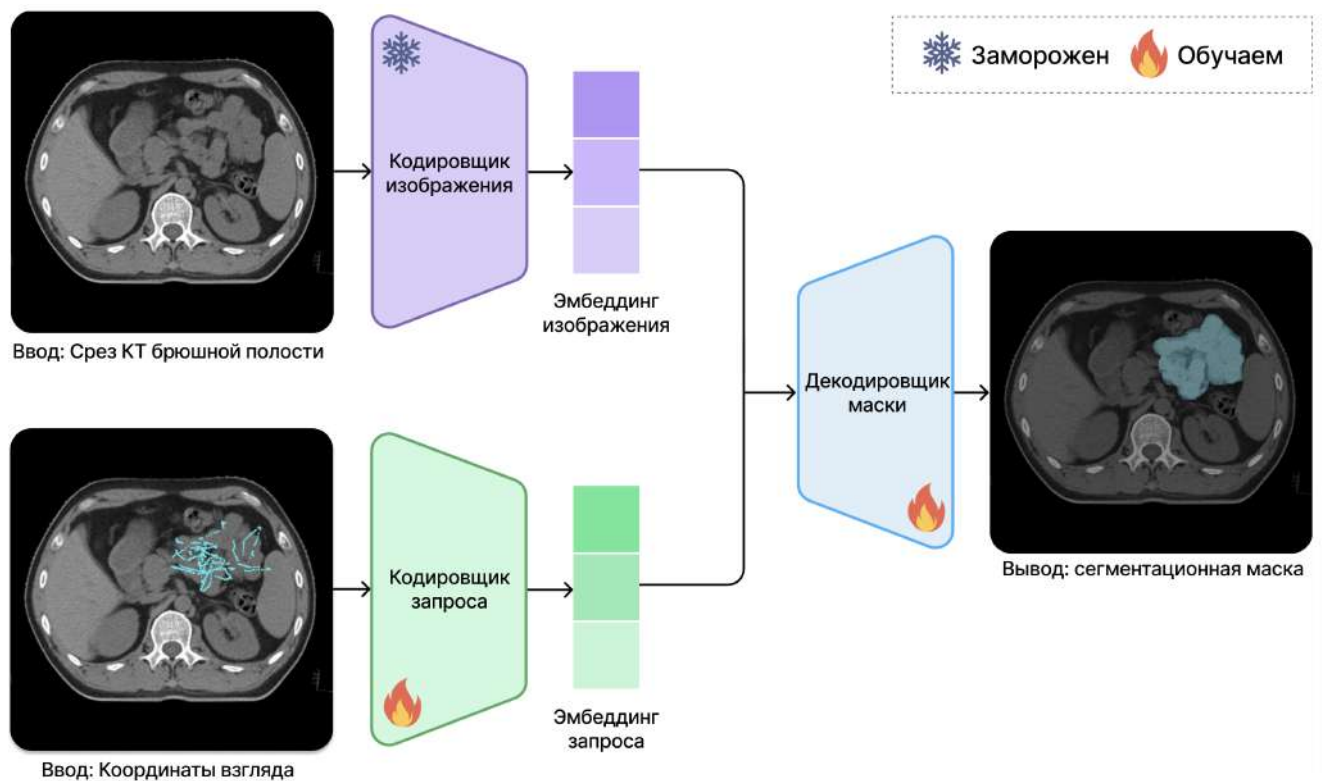


Рисунок 4.2 — Визуализация архитектуры интерактивной медицинской сегментации с использованием взгляда. Система использует данные взгляда для генерации сегментационной маски текущего органа. Это достигается за счет использования набора координат взгляда в качестве запроса для модели MedSAM вместе с входным срезом КТ-изображения. Кодировщик изображений остается замороженным (в процессе обучения его параметры не обновляются). Дообучается кодировщик запросов и декодировщик масок. На рисунке продемонстрирован пример предсказания сегментационной маски кишечника

Чтобы адаптировать MedSAM к запросам в виде данных взгляда, была дообучена часть декодировщика масок и кодировщика запросов с использованием синтетических данных, имитирующих данные взгляда. Эти синтетические данные моделируют входные координаты взгляда различной длины.

Данные взгляда имеют существенные отличия от других источников запросов, таких как ограничивающие рамки, координаты кликов мышью и текстовых описаний, а именно

1. в отличие от кликов мышью, где для каждой координаты клика указывается желание пользователя (добавить область в сегментационную маску или удалить из нее), невозможно указывать желание пользователя для каждой точки взгляда. Это связано с большим количеством

точек взгляда (в работе используется устройство для отслеживания взгляда в частотой 90 Гц);

2. взгляд имеет естественную дисперсию и не полностью контролируется пользователем.

Для решения этих проблем во время дообучения модели использовались синтетические координаты взгляда. Случайным образом выбиралось 80% сгенерированных точек из области референсной сегментационной маски, а оставшиеся 20% точек выбирались вне этой области. Эти пропорции основаны на предположении, что большая часть точек взгляда будет нацелена на область интереса (текущий орган), а оставшиеся 20% точек вне области интереса связаны с флуктуациями взгляда и потенциальными неточностями, связанными с отслеживанием взгляда (например, при значительном отклонении положения головы пользователя от изначального положения).

Модель дообучалась на компьютере с видеокартой NVIDIA GTX 3090 Ti. В процессе дообучения использовался оптимизатор AdamW [121] с начальной скоростью обучения  $5 \times 10^{-5}$ . Модели обучались в течение 200 эпох. Был добавлен дополнительный критерий останова для случая, если на протяжении 10 эпох значение функционала качества на валидационном наборе данных не улучшалось.

В качестве функции потерь использовалась комбинация потери Дайса и кросс-энтропии:

$$\mathcal{L} = \mathcal{L}_{DSC} + \mathcal{L}_{CE},$$

которая показала свою надежность в различных задачах сегментации [84; 92]. Потеря Дайса вычисляется как  $1 - DSC$ , где DSC - коэффициент Дайса, определяемый следующим образом:

$$DSC(G, P) = \frac{2 \cdot |G \cap P|}{|G| + |P|},$$

где  $G$  и  $P$  обозначают референсную и предсказанную сегментационную маски соответственно. Коэффициент Дайса используется для оценки пересечения областей между референсными и предсказанными сегментационными масками, его значения варьируются от 0 до 1.

## 4.5 Набор данных

В исследовании использовался общедоступный набор данных WORD [82], который включает 150 КТ-снимки для 150 пациентов. Набор данных содержит 16 органов брюшной полости: печень, селезенку, левую и правую почки, желудок, желчный пузырь, пищевод, двенадцатиперстную кишку, ободочную кишку, тонкую кишку, надпочечники, прямую кишку, мочевой пузырь, а также левую и правую головки бедренной кости. Каждый КТ-снимок содержит от 159 до 330 срезов размером  $512 \times 512$  пикселей. Размер пикселя составляет  $0.976 \times 0.976$  мм, с расстоянием между срезами от 2.5 до 3.0 мм. Как показано в таблице 8, весь набор данных состоит из 150 трехмерных КТ-сканов, преобразованных в соответствующие срезы. Для оценки качества модели использовался официальный тестовый набор данных из WORD, содержащий КТ-снимки 30 пациентов.

Таблица 8 — Детали и клинические характеристики наборов данных для обучения, валидации и тестирования из базы WORD [82].

| Характеристики                | Обучающий  | Валидационный | Тестовый   |
|-------------------------------|------------|---------------|------------|
| Количество сканов (пациентов) | 100        | 20            | 30         |
| Количество срезов             | 85 921     | 17 652        | 28 110     |
| Возраст (медиана)             | 47 (28-75) | 52 (32-78)    | 49 (26-72) |
| Мужчины                       | 63         | 12            | 13         |
| Женщины                       | 37         | 8             | 17         |
| Рак простаты                  | 28         | 7             | 10         |
| Рак шейки матки               | 29         | 6             | 5          |
| Рак прямой кишки              | 26         | 3             | 8          |
| Метастатические опухоли       | 17         | 4             | 7          |

## 4.6 Эксперименты и результаты

Перед проведением основных экспериментов с практикующими врачами был привлечен рентгенолог-стажёр для проведения предварительного экспе-

римента с целью совершенствования интеграции реальных данных взгляда в разработанную систему.

Взгляд человека постоянно находится в движении. Есть промежутки относительной неподвижности взгляда, которые называются точками фиксации. В этих промежутках происходит восприятие большей части визуальной информации. точки фиксации соединены быстрыми движениями взгляда, которые называются саккадами. Во время саккад визуальная информация почти не воспринимается [122]. Из данных взгляда были извлечены точки фиксации (см. рисунок 3.1в).

В эксперименте со стажёром-рентгенологом были протестированы различные стратегии выбора точек взгляда для использования в модели интерактивной сегментации:

1. *случайные точки взгляда*: из всей последовательности координат выбирается 20 случайных точек взгляда;
2. *фиксации с возвратом*: из набора данных взгляда извлекаются точки фиксации, из которых случайным образом извлекается 20 точек фиксации с возвратом;
3. *фиксации с добавлением*: выбираются 20 последних точек фиксации.

Для выбора оптимальной стратегии рентгенолог-стажер должен быть выполнить интерактивную сегментацию (при помощи взгляда) различных органов на 54 срезах КТ-снимков, используя модель, обученную на 5% набора данных WORD с 20 точками в запросе. В таблице 9 показаны средние значения коэффициента Дайса для трех различных подходов. Результаты показали, что выбор 20 случайных точек из всей истории отслеживания взгляда обеспечивает наилучшую эффективность.

В таблице 10 показаны результаты сравнения качества сегментации предложенного подхода с использованием взгляда (обозначен как GazeSAM) с современными моделями сегментации, которые были протестированы на наборе данных WORD. Предложенный подход GazeSAM был обучен на полном наборе WORD с запросами в виде 20 синтетических точек взгляда, где 80% точек располагались внутри анатомической структуры, а 20% за ее пределами. В сравнение также был включен оригинальный MedSAM, который тестировался с использованием ограничивающих рамок.

Согласно таблице 10 предложенный подход превосходит как оригинальный MedSAM с ограничивающими рамками, так и другие современные подходы

Таблица 9 — Сравнение качества сегментации на основе коэффициента Дайса (в процентах) для трех стратегий выбора точек взгляда. Жирным выделен наилучший результат для каждого органа.

| Орган                    | Случайные точки | Точки фиксации |              |
|--------------------------|-----------------|----------------|--------------|
|                          |                 | Случайные      | С возвратом  |
| Печень                   | <b>96.74</b>    | 62.49          | 94.15        |
| Селезенка                | 91.64           | 92.85          | <b>93.88</b> |
| Левая почка              | 86.05           | 56.56          | <b>96.00</b> |
| Правая почка             | 88.12           | <b>92.68</b>   | 88.69        |
| Желудок                  | <b>96.85</b>    | 90.69          | 95.07        |
| Желчный пузырь           | <b>93.43</b>    | 90.36          | 89.74        |
| Пищевод                  | 88.99           | <b>92.34</b>   | 62.07        |
| Поджелудочная            | <b>93.35</b>    | 81.75          | 91.07        |
| Двенадцатиперстная кишка | <b>88.78</b>    | 79.39          | 82.82        |
| Ободочная кишка          | 85.46           | <b>88.20</b>   | 82.05        |
| Тонкая кишка             | <b>83.94</b>    | 79.33          | 77.50        |
| Надпочечник              | 65.00           | 77.30          | <b>75.45</b> |
| Прямая кишка             | <b>93.87</b>    | 68.98          | 87.84        |
| Мочевой пузырь           | 39.44           | 16.17          | <b>72.83</b> |
| Головка бедра левая      | 89.19           | 87.65          | <b>92.36</b> |
| Головка бедра правая     | <b>92.32</b>    | 85.85          | 82.86        |
| Среднее                  | <b>87.03</b>    | 80.79          | 84.42        |

для сегментации различных органов брюшной полости. Более того, представленный подход показывает значительное улучшение качества сегментации в особенно сложных случаях, таких как двенадцатиперстная кишка и надпочечники. Например, предложенный подход демонстрирует улучшение коэффициента Дайса на 14% по сравнению с nnUnetV2 [92] для надпочечника.

Таблица 10 — Сравнение коэффициента Дайса (в процентах) с современными моделями.

| Орган             | Метод<br>nnUNetV2 [92] | ResUNet [123]                      | MedSAM [84]                        | Предложенный подход                |
|-------------------|------------------------|------------------------------------|------------------------------------|------------------------------------|
| Печень            | $96.19 \pm 2.16$       | <b><math>96.55 \pm 0.89</math></b> | $92.15 \pm 1.22$                   | $96.00 \pm 0.33$                   |
| Селезенка         | $94.33 \pm 7.72$       | $95.26 \pm 2.84$                   | $78.68 \pm 0.00$                   | <b><math>96.18 \pm 0.05</math></b> |
| Почка (Л)         | $91.29 \pm 18.15$      | $95.63 \pm 1.20$                   | $83.40 \pm 0.00$                   | <b><math>96.20 \pm 0.00</math></b> |
| Почка (П)         | $91.20 \pm 17.22$      | $95.84 \pm 1.16$                   | $91.93 \pm 0.00$                   | <b><math>96.35 \pm 0.00</math></b> |
| Желудок           | $91.12 \pm 3.60$       | $91.58 \pm 2.86$                   | $91.61 \pm 0.00$                   | <b><math>95.51 \pm 0.18</math></b> |
| Желчный пузырь    | $83.19 \pm 12.22$      | $82.83 \pm 11.8$                   | $85.43 \pm 0.00$                   | <b><math>92.57 \pm 0.02</math></b> |
| Пищевод           | $77.79 \pm 13.51$      | $77.17 \pm 14.68$                  | $88.94 \pm 5.61$                   | <b><math>90.29 \pm 1.23</math></b> |
| Поджелудочная     | $83.55 \pm 5.87$       | $83.56 \pm 5.60$                   | $79.78 \pm 0.00$                   | <b><math>90.85 \pm 0.23</math></b> |
| 12-ти перстная к. | $64.47 \pm 15.87$      | $66.67 \pm 15.36$                  | $80.07 \pm 0.00$                   | <b><math>89.82 \pm 0.51</math></b> |
| Ободочная кишка   | $83.92 \pm 8.45$       | $83.57 \pm 8.69$                   | $75.96 \pm 0.00$                   | <b><math>91.80 \pm 1.13</math></b> |
| Тонкая кишка      | $86.83 \pm 4.02$       | $86.76 \pm 3.56$                   | $84.00 \pm 0.00$                   | <b><math>90.78 \pm 1.87</math></b> |
| Надпочечник       | $70.0 \pm 11.86$       | $70.9 \pm 10.12$                   | $62.74 \pm 5.28$                   | <b><math>84.77 \pm 0.77</math></b> |
| Прямая кишка      | $81.49 \pm 7.37$       | $82.16 \pm 6.73$                   | <b><math>92.54 \pm 2.33</math></b> | $92.53 \pm 0.27$                   |
| Мочевой пузырь    | $90.15 \pm 16.85$      | $91.0 \pm 13.5$                    | $86.74 \pm 0.00$                   | <b><math>95.26 \pm 0.02</math></b> |
| Головка бедра (Л) | $93.28 \pm 5.12$       | $93.39 \pm 5.11$                   | $70.17 \pm 0.00$                   | <b><math>94.99 \pm 0.30</math></b> |
| Головка бедра (П) | $93.93 \pm 4.29$       | $93.88 \pm 4.30$                   | $63.28 \pm 0.00$                   | <b><math>95.13 \pm 0.29</math></b> |
| Среднее           | $85.80 \pm 5.27$       | $86.67 \pm 4.81$                   | $81.71 \pm 9.39$                   | <b><math>92.53 \pm 3.24</math></b> |

#### 4.7 Эксперименты с рентгенологами

В эксперименте участвовали два практикующих рентгенолога. Рентгенолог А специализируется на анализе рентгеновских и КТ изображений с пятилетним опытом работы. Рентгенолог В имеет восьмилетний опыт работы в классической диагностике, анализе КТ и МРТ снимков с акцентом на онкологию и диагностику злокачественных образований. Оба врача независимо (без совместных обсуждений) использовали предложенный подход для сегментации органов на КТ-срезах. Затраченное время на каждый случай в экспериментах фиксировалось.

Для проведения эксперимента было отобрано 100 КТ-срезов из тестового набора WORD, охватывающих 20 пациентов. Было выбрано по 5 КТ-срезов для каждого пациента, обеспечивая представление разных органов на каждом изображении. Распределение органов в выбранных образцах представлено в таблице 11. Процедура отбора данных для эксперимента основывалась на производительности сегментации оригинального MedSAM с использованием ограничивающих рамок. Было выбрано 60 образцов с низкой производитель-

ностью (коэффициент Дайса менее 0.7) и 40 образцов с высокими значениями коэффициента Дайса (более 0.7). В результате образцы, отобранные для экспериментов, в основном включали сложные случаи для традиционных методов, особенно основанных на ограничивающих рамках. Эти КТ-изображения были представлены рентгенологам А и В во время эксперимента.

Таблица 11 — Образцы срезов, отобранные для экспериментов с врачами.

| Орган                    | Количество образцов |
|--------------------------|---------------------|
| Печень                   | 7                   |
| Селезенка                | 3                   |
| Левая почка              | 7                   |
| Правая почка             | 3                   |
| Желудок                  | 3                   |
| Желчный пузырь           | 7                   |
| Пищевод                  | 7                   |
| Поджелудочная            | 7                   |
| Двенадцатиперстная кишка | 7                   |
| Ободочная кишка          | 7                   |
| Тонкая кишка             | 7                   |
| Надпочечник              | 7                   |
| Прямая кишка             | 7                   |
| Мочевой пузырь           | 7                   |
| Головка бедра (Л)        | 7                   |
| Головка бедра (П)        | 7                   |
| Всего                    | 100                 |

Оба эксперта (рентгенолог А и В) получили задание выполнить сегментацию анатомических областей на основе ограничивающих рамок с использованием оригинального MedSAM. Когда результаты сегментации не соответствовали ожиданиям, рентгенологи уточняли сегментацию с помощью предложенного подхода с использованием взгляда (GazeSAM).

Рентгенологу А потребовалось скорректировать с использованием взгляда 11 масок сегментации от 9 пациентов по 6 органам, включая поджелудочную железу, ободочную кишку, тонкую кишку, надпочечник и обе головки бедренной кости. Подход на основе ограничивающих рамок показал средний коэффициент

Дайса 0.39 при среднем затрачиваемом времени 9.7 секунд. После коррекции с помощью взгляда производительность сегментации значительно улучшилась на 0.24 в среднем, однако среднее время увеличилось примерно до 15.7 секунд, как показано в таблице 12. Остальные 89 органов (не требующие коррекции) были сегментированы с использованием оригинального метода MedSAM с созданными ограничивающими рамками, для них средний коэффициент Дайса составил 0.704.

Таблица 12 — Результаты коррекции масок с использованием взгляда для рентгенолога А.

| Орган             | Подход               |            | Взгляд      |            |
|-------------------|----------------------|------------|-------------|------------|
|                   | Ограничивающие рамки | Взгляд     | Коэф. Дайса | Время, сек |
|                   | Коэф. Дайса          | Время, сек | Коэф. Дайса | Время, сек |
| Поджелудочная     | 0.554                | 3.14       | 0.509       | 5.02       |
| Ободочная кишка   | 0.62                 | 22.15      | 0.897       | 23.64      |
| Тонкая кишка      | 0.497                | 20.69      | 0.627       | 24.54      |
| Надпочечник       | 0.432                | 4.54       | 0.665       | 14.19      |
| Головка бедра (Л) | 0.148                | 9.67       | 0.541       | 7.68       |
| Головка бедра (П) | 0.244                | 6.06       | 0.587       | 19.41      |
| Среднее           | 0.39                 | 9.69       | 0.631       | 15.72      |

Рентгенологу В потребовалось скорректировать с использованием взгляда 14 масок от 9 пациентов по 7 органам. Органы включали печень, пищевод, двенадцатиперстную кишку, ободочную кишку, тонкую кишку, прямую кишку и мочевой пузырь. Согласно таблице 13, подход на основе ограничивающих рамок показал средний коэффициент Дайса 0.617 при среднем времени около 13.5 секунд. После коррекции взглядом средний коэффициент Дайса увеличился до 0.715, в то время как среднее время возросло почти до 42 секунд. Для остальных 86 органов, не требующих коррекции, подход на основе ограничивающих рамок показал средний коэффициент Дайса 0.713.

Представленный подход интерактивной сегментации медицинских изображений с использованием данных взгляда продемонстрировал значительные преимущества по сравнению с традиционными подходами. Ключевые результаты исследования включают:

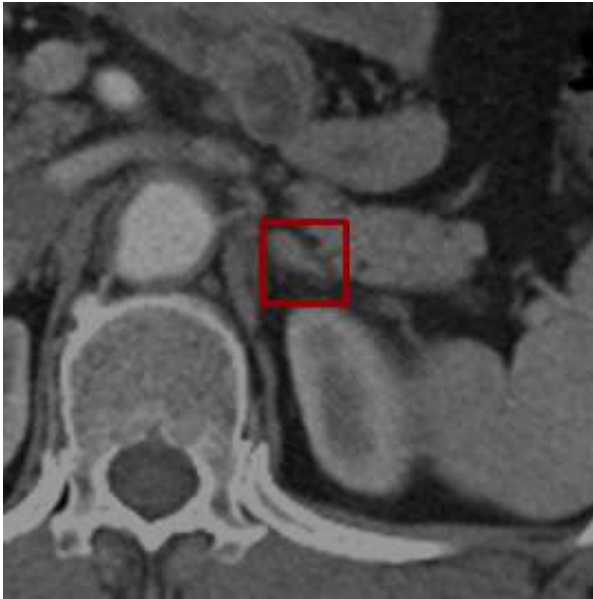
Таблица 13 — Результаты коррекции масок с использованием взгляда для рентгенолога В.

| Орган             | Подход               |            | Взгляд      |            |
|-------------------|----------------------|------------|-------------|------------|
|                   | Ограничивающие рамки | Время, сек | Коэф. Дайса | Время, сек |
|                   | Коэф. Дайса          |            | Коэф. Дайса |            |
| Печень            | 0.577                | 6.67       | 0.609       | 17.67      |
| Пищевод           | 0.685                | 7.75       | 0.752       | 12.55      |
| 12-ти перстная к. | 0.669                | 8.66       | 0.75        | 41.01      |
| Ободочная кишка   | 0.573                | 9.48       | 0.668       | 58.42      |
| Тонкая кишка      | 0.526                | 26.94      | 0.767       | 65.99      |
| Прямая кишка      | 0.853                | 7.27       | 0.666       | 24.03      |
| Мочевой пузырь    | 0.556                | 8.86       | 0.578       | 42.78      |
| Среднее           | 0.617                | 13.58      | 0.715       | 41.78      |

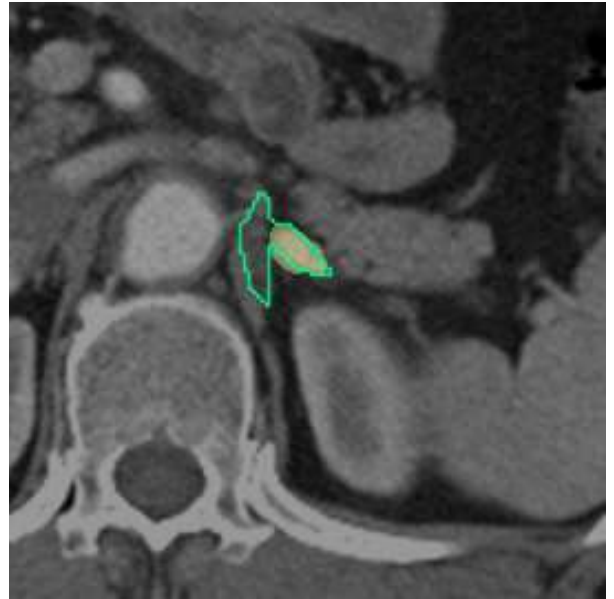
1. подход обеспечивает существенное улучшение точности сегментации (в среднем на 10.82% коэффициента Дайса в сравнении с оригинальным MedSAM) в тестировании на синтетических данных взгляда, особенно для сложных случаев;
2. разработанная система эффективно использует естественные движения глаз врачей для интерактивного уточнения масок сегментации в реальном времени;
3. эксперименты с практикующими рентгенологами подтвердили, что предложенный подход позволяет улучшить качество сегментации в среднем на 24% (DSC) для сложных случаев, требующих коррекции.

Будущие исследования могут быть сконцентрированы на задаче предсказания объемных сегментационных масок, повышения качества генерации синтетических данных взгляда во время обучения (например, за счет внедрения математических моделей взгляда или глубоких моделей для предсказания последовательности взгляда), разработке подходов эффективно сочетающих различные типы запросов (например, взгляд и ограничивающие рамки или текстовое описание).

## 4.8 Обсуждение



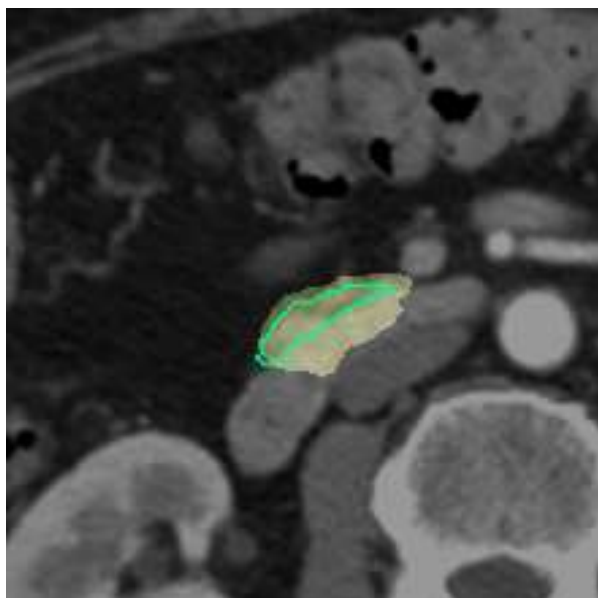
а) Ограничивающий  
прямоугольник, определённый  
радиологом для сегментации  
надпочечника



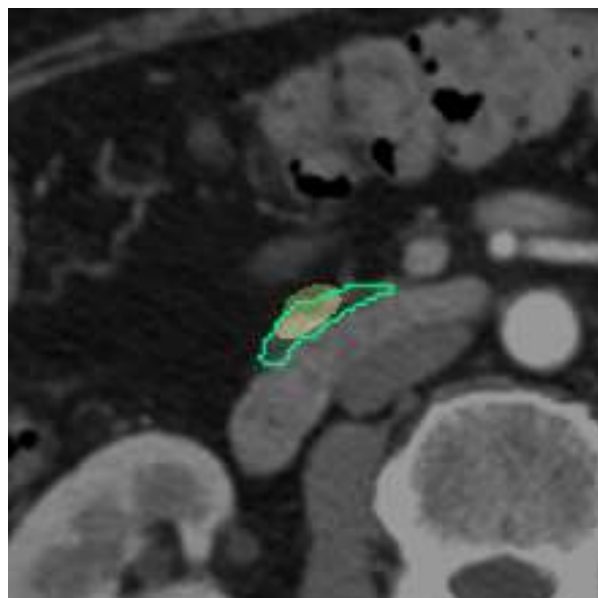
б) Маска сегментации  
надпочечника, полученная на  
основе ограничивающей рамки

Рисунок 4.3 — Результаты сегментации надпочечника, скорректированные медицинским экспертом

В главе представлен метод коррекции сегментации медицинских изображений с использованием информации о взгляде. Эксперименты проводились с участием двух медицинских экспертов, выполнявших сегментацию органов на КТ-изображениях брюшной полости. Рентгенологи изначально выполняли разметку при помощи ограничивающих рамок (англ. *bounding boxes*), после чего при необходимости вносили корректировки на основе данных о направлении взгляда. Проведённые эксперименты показали, что использование коррекции на основе взгляда позволяет существенно повысить точность сегментации. Например, для рентгенолога А среднее значение коэффициента Дайса увеличилось почти в 1.5 раза после применения метода на основе взгляда. Однако, как показали результаты, использование коррекции на основе взгляда требует дополнительного времени, что указывает на компромисс между точностью и эффективностью. Тем не менее, в условиях медицинской диагностики, особенно в онкологической практике, приоритет отдается точности, что подтверждает практическую ценность предложенного подхода. Кроме того, несмотря



а) Маска сегментации поджелудочной железы, полученная с использованием ограничивающей рамки



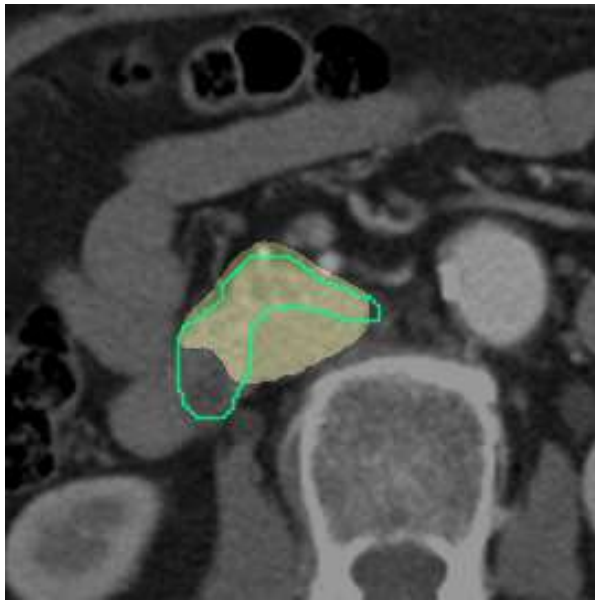
б) Маска поджелудочной железы, полученная на основе данных о взгляде

Рисунок 4.4 — Сравнение методов сегментации поджелудочной железы, проведенных медицинским экспертом

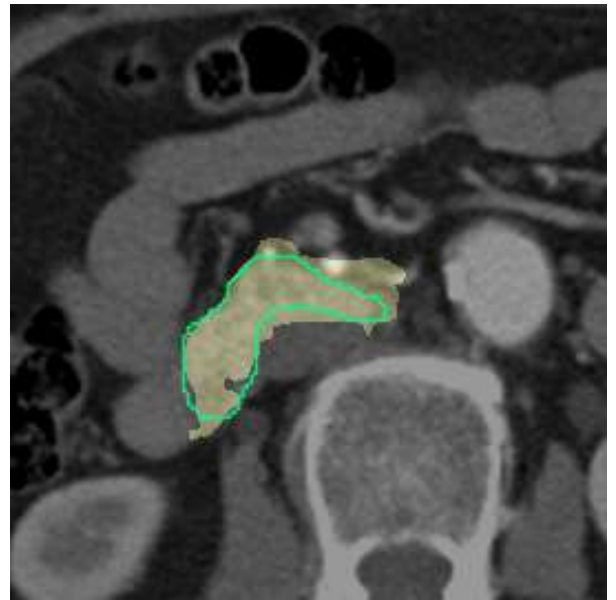
на краткий инструктаж по использованию системы отслеживания взгляда, продемонстрирована её успешная интеграция в рабочий процесс, с потенциалом повышения эффективности при дальнейшей адаптации в клинике.

Несмотря на краткий инструктаж по использованию технологии отслеживания взгляда перед началом экспериментов, оба врача успешно адаптировались к предлагаемому интерфейсу. Следует ожидать, что при более длительном обучении и дальнейшей интеграции подобных интерфейсов в клинический рабочий процесс эффективность использования метода возрастёт.

Дополнительный анализ продемонстрировал, что не для всех анатомических структур требовалась коррекция. Так, для селезёнки, желудка, желчного пузыря, пищевода и почек ограничивающие прямоугольники обеспечивали высокое качество сегментации, не требующее улучшения с точки зрения врача. Например, среднее значение коэффициента Дайса для селезёнки составило 0.882, а для правой почки — 0.869, как показано в таблице 14. Для мочевого пузыря тоже были продемонстрированы хорошая производительность сегментации (коэффициент Дайса 0.778), за исключением одного случая, где у второго врача после коррекции наблюдалось понижение коэффициента Дайса до 0.579,



а) Маска двенадцатиперстной кишки, полученная по bounding box



б) Маска двенадцатиперстной кишки, полученная на основе данных взгляда

Рисунок 4.5 — Результаты сегментации двенадцатиперстной кишки с использованием разных методов

что объясняется небольшой площадью органа и сложностью его выделения, как отмечается в [124].

В отличие от этих примеров, такие органы, как печень и поджелудочная железа, оказались более сложными для автоматического выделения. Несмотря на относительно большие размеры, границы печени нередко неразличимы на фоне окружающих тканей, что согласуется с предыдущими наблюдениями в литературе [92]. Поджелудочная железа, обладая небольшой и нерегулярной формой, также демонстрировала низкое качество сегментации (коэффициент Дайса 0.534), что подтверждает известную трудность в её автоматическом выделении [82; 125; 126]. Использование взгляда позволило компенсировать эти ограничения и существенно повысить точность, особенно в случае органов с малой площадью или сложной геометрией.

На иллюстрациях 4.3–4.5 представлены примеры сегментации надпочечника, поджелудочной железы и двенадцатиперстной кишки. В частности, в случае надпочечника использование ограничивающего прямоугольника привело к значительной ошибке (коэффициент Дайса 0.53), однако врач отказался от коррекции взглядом, посчитав результат удовлетворительным. При повторной оценке медицинский эксперт отметил, что предсказанная маска захватывала

Таблица 14 — Средние значения коэффициента Дайса и времени (в мс) при использовании только ограничивающих рамок (до или без коррекции по взгляду) для рентгенологов А и В.

| Орган                            | DSC     | Время, мс |
|----------------------------------|---------|-----------|
| Печень                           | 0.62700 | 2660      |
| Селезёнка                        | 0.88223 | 2357      |
| Почка (левая)                    | 0.78354 | 1482      |
| Почка (правая)                   | 0.86920 | 1571      |
| Желудок                          | 0.86827 | 24        |
| Жёлчный пузырь                   | 0.73060 | 2145      |
| Пищевод                          | 0.59689 | 4001      |
| Поджелудочная железа             | 0.53449 | 4372      |
| Двенадцатиперстная кишка         | 0.70931 | 5326      |
| Толстая кишка                    | 0.68445 | 7418      |
| Кишечник                         | 0.52838 | 20209     |
| Надпочечник                      | 0.51400 | 3409      |
| Прямая кишка                     | 0.64763 | 5955      |
| Мочевой пузырь                   | 0.77774 | 2818      |
| Головка бедренной кости (левая)  | 0.36910 | 7242      |
| Головка бедренной кости (правая) | 0.36551 | 9513      |
| Среднее                          | 0.64024 | 5495      |

аорту, что снижает точность аннотации. Аналогично, в случае с поджелудочной железой визуально малая ошибка привела к заметному снижению точности, поскольку орган представлен на срезе в виде небольшой области, и маска, полученная на основе взгляда, также оказалась неточной. В случае с двенадцатиперстной кишкой изначальный коэффициент Дайса составил 0.63, но после коррекции по взгляду — увеличился до 0.80, и был признан экспертом точным. Это подтверждает, что в случае несимметричных или вогнутых структур ограничивающие прямоугольники могут охватывать избыточную область, тогда как коррекция взглядом позволяет уточнить форму.

Был проведён корреляционный анализ между различными параметрами. Выявлена положительная связь между временем, затраченным на сегментацию, и размером органа (коэффициент Пирсона  $r = 0.67$ ), что соответствует интуитивным ожиданиям: крупные структуры требуют большего визуального

внимания и, следовательно, увеличенного времени аннотирования. Дополнительно, обнаружено слабое положительное влияние времени на точность сегментации ( $r = 0.13$ ), что согласуется с ранее представленными результатами [82].

По сравнению с другими методами сегментации с интерактивной коррекцией — такими как ScribblePrompt [114], MedLSAM [127] и Med-SA [113] — предложенный подход позволяет повысить качество сегментации за счёт включения данных о взгляде эксперта. В отличие от моделей, нацеленных на генерацию масок, представленная система ориентирована на уточнение границ и устранение ошибок, что особенно важно в задачах клинической точности.

Среди ограничений исследования следует отметить использование лишь одного набора данных и отсутствие трёхмерной аннотации. В перспективе планируется расширение методологии на объёмные медицинские данные с возможностью просмотра в различных плоскостях, а также изучение применения подхода в других модальностях и анатомических зонах.

## Заключение

В диссертационной работе получены следующие основные результаты:

1. Разработана глубокая модель, учитывающая семантическую информацию, полученной при помощи метода логит-линзы, для предсказания взгляда рентгенолога на рентгеновском изображении грудной клетки.
2. Разработан способ повышения качества классификации патологий на рентгеновском изображении грудной клетки глубокой моделью, путем использования синтетических данных взгляда рентгенолога во время обучения модели.
3. Разработана и апробирована экспериментальная установка для анализа визуального поведения рентгенолога при анализе рентгеновских изображений. Проведены валидирующие эксперименты с участием практикующих рентгенологов.
4. Разработан способ и глубокая модель для интерактивной сегментации медицинских изображений на основе взгляда. Проведены эксперименты с отслеживанием взгляда рентгенологов для тестирования глубокой модели;
5. Реализован комплекс программ для проведения экспериментов по анализу медицинских изображений с отслеживанием взгляда в режиме реального времени и оконтуривания анатомических областей на медицинских изображениях на основе взгляда рентгенолога с использованием технологий искусственного интеллекта.

Разработан способ обогащения глубокой модели семантической информацией при помощи метода логит-линзы. Рассмотрено применение результатов метода логит-линзы к визуально-языковой модели LLaVA и адаптированной для медицинского домена LLaVA-Med. При помощи этого способа получена глубокая модель для предсказания взгляда рентгенолога на рентгеновском изображении грудной клетки, которая в сравнении с существующими решениями демонстрирует наилучшее качество предсказания.

Разработан способ повышения качества классификации патологий на рентгеновском изображении грудной клетки глубокой моделью за счет интеграции синтетических данных взгляда рентгенолога во время обучения модели. Способ основан на предсказаниях тепловой карты взгляда рентгенолога как до-

полнительной задачи во время обучения модели. В качестве данных взгляда рентгенолога использовались синтетические данные взгляда рентгенолог, полученные разработанной глубокой моделью в рамках диссертации. Качество классификации полученной модели для обнаружения патологий на рентгеновском изображении грудной клетки с интегрированными синтетическими данными взгляда рентгенолога превосходит качество классификации модели с натуральными данными взгляда рентгенолога.

Разработана экспериментальная установка, схожая с рабочей станцией рентгенолога, дополнительно оснащенная устройством для отслеживания взгляда. Валидация установки проводилась в виде хорошо известно в литературе задачи оценки усталости рентгенолога при длительном анализе рентгеновских снимков грудной клетки. Предложен способ оценки изменений закономерностей взгляда рентгенолога при длительном анализе рентгеновских изображений грудной клетки, который основан на анализе особенностей визуального анализа анатомических областей на рентгене. Для решения задачи сегментации анатомических областей использовалась сверточная нейронная сеть. Проведены эксперименты с привлечением четырех практикующих рентгенологов, каждым из которых было проанализировано 400 рентгеновских изображений. Рассмотрена связь различных характеристик (результаты тестов усталости/концентрации, длина траектории взгляда, скорость взгляда, затраченное время на анализ изображение и пр.) с количеством просмотренных снимков. Установлено, что информационный прирост доли покрытия взглядом легких имел наибольший коэффициент корреляции Пирсона с количеством просмотренных снимков для всех рентгенологов.

Разработан способ для интерактивной сегментации анатомических областей на КТ-срезах брюшной полости по данным взгляда рентгенолога. Способ основан на использовании синтетических данных взгляда в качестве запроса для интерактивной сегментации во время обучения модели. Проведены эксперименты с отслеживанием взгляда рентгенологов для тестирования модели. Показано, что разработанная модель позволяет значительно повысить качество классификации сложных случаев, требующих коррекции результатов автоматической сегментации.

В диссертационной работе рассмотрен полный цикл работы с данными взгляда: от разработки установки и программного обеспечения для сбора данных взгляда и проведения экспериментов, до обработки и анализа полученных

результатов, разработки различных способов и моделей для работы с данными взгляда. Такой всецелый подход позволил провести ряд уникальных экспериментов, получить улучшение качества в ранее известных задачах и предложить новые решения, которые потенциально могут привести к оптимизации рабочих процессов рентгенолога и повышению качества диагностики.

В качестве дальнейшего развития новых подходов применения данных взгляда в клинической практике является переход к задаче трехмерной сегментации на основе взгляда, разработке методов определения расхождения между предсказанием глубокой модели и мнением врача, разработке подходов для построения эффективной системы поддержки принятия решений на основе взгляда рентгенолога, разработке подходов для выравнивания больших языковых визуально-языковых моделей с использованием данных взгляда.

## Словарь терминов

**область интереса:** Участок изображения, на который направлено внимание наблюдателя или модели

**компьютерная томография (КТ):** Метод послойной визуализации внутренних структур организма с помощью рентгеновского излучения

**мрт (МРТ):** Метод визуализации, основанный на ядерном магнитном резонансе, применяемый для получения высококонтрастных изображений мягких тканей

**рентгеновское изображение:** Изображение, полученное в результате прохождения рентгеновских лучей через тело пациента

**точки фиксации взгляда:** Координаты на изображении, в которых наблюдатель останавливает взгляд в течение определённого времени

**саккады:** Быстрые, скачкообразные движения глаз между точками фиксации

**аугментация:** Метод искусственного увеличения обучающей выборки путём внесения трансформаций в исходные данные

**батч:** Набор объектов из обучающей выборки, обрабатываемый моделью за один шаг обновления весов

**трансформер:** Архитектура нейросети, основанная на механизме внимания, широко применяемая в задачах обработки последовательностей

**глубокая модель:** Многослойная нейросетевая архитектура, способная извлекать сложные абстрактные признаки из входных данных

**бейзлайн:** Базовая модель или метод, используемый в качестве точки отсчёта при сравнении новых подходов

**кодировщик:** Компонент модели, преобразующий входные данные в сжатое латентное представление

**инференс:** Процесс применения обученной модели для получения предсказаний на новых данных

**эмбединг:** Векторное представление объекта (текста, изображения и т. п.) в пространстве меньшей размерности, чем исходного объекта

**логит:** Значение на выходе последнего линейного слоя модели до применения функции softmax

**CNN:** Сокращение с английского «convolutional neural network» – сверточная нейронная сеть, применяемая для извлечения признаков из изображений

**Авторегрессионные модели:** Класс моделей машинного обучения, которые автоматически предсказывают следующий элемент последовательности на основе всех предыдущих элементов этой последовательности

## Список рисунков

|     |                                                                                                                                                                                                                                                        |    |
|-----|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1   | Диаграмма решения исследовательской проблемы, отражающая вклад каждой главы диссертации . . . . .                                                                                                                                                      | 10 |
| 1.1 | Визуализация траектории взгляда на рентгеновском снимке грудной клетки, помеченном как «норма». Жёлтые круги обозначают точки фиксации. Фрагменты изображения декодируются в слова при помощи метода логит-линзы и визуально-языковой модели . . . . . | 18 |
| 1.2 | Высокоуровневая визуализация архитектуры модели LogitGaze-Med                                                                                                                                                                                          | 24 |
| 1.3 | Сравнение путей взгляда человека (верхний ряд), прогнозов LogitGaze-Med (средний ряд) и прогнозов GazeSearch (нижний ряд) по задачам: норма, кардиомегалия, консолидация и пневмония. . . .                                                            | 30 |
| 2.1 | Визуализация второго способа (Temporal) для решения задачи классификации патологий . . . . .                                                                                                                                                           | 39 |
| 2.2 | Визуализация третьего способа (Static) для решения задачи классификации патологий . . . . .                                                                                                                                                            | 40 |
| 3.1 | Демонстрация установки для отслеживания взгляда рентгенолога и пример обработки данных взгляда . . . . .                                                                                                                                               | 51 |
| 3.2 | Графический пользовательский интерфейс для теста замены цифровых символов и теста времени реакции . . . . .                                                                                                                                            | 55 |
| 3.3 | Сравнение тепловых карт взгляда четырех рентгенологов . . . . .                                                                                                                                                                                        | 59 |
| 3.4 | Сравнение диагностической точности рентгенологов . . . . .                                                                                                                                                                                             | 61 |
| 3.5 | Изменения покрытия легких для каждого рентгенолога и для каждой патологии в зависимости от количества просмотренных снимков во время эксперимента . . . . .                                                                                            | 62 |
| 3.6 | Результаты тестов усталости у рентгенологов . . . . .                                                                                                                                                                                                  | 64 |
| 4.1 | Процесс коррекции сегментационных масок на основе взгляда рентгенолога . . . . .                                                                                                                                                                       | 73 |
| 4.2 | Архитектура интерактивной медицинской сегментации с использованием взгляда . . . . .                                                                                                                                                                   | 75 |
| 4.3 | Результаты сегментации надпочечника, скорректированные медицинским экспертом . . . . .                                                                                                                                                                 | 84 |

|     |                                                                                                    |    |
|-----|----------------------------------------------------------------------------------------------------|----|
| 4.4 | Сравнение методов сегментации поджелудочной железы,<br>проверенных медицинским экспертом . . . . . | 85 |
| 4.5 | Результаты сегментации двенадцатиперстной кишки с<br>использованием разных методов . . . . .       | 86 |

## Список таблиц

|    |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                       |    |
|----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 1  | Значения метрик схожести траекторий взгляда (чем выше, тем лучше для ScanMatch/STDE, чем ниже, тем лучше для SED). . . . .                                                                                                                                                                                                                                                                                                                                                            | 28 |
| 2  | Сходство MultiMatch (чем выше, тем лучше) по пяти компонентам. .                                                                                                                                                                                                                                                                                                                                                                                                                      | 29 |
| 3  | Кросс-доменная оценка на наборе данных COCO18. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                                | 31 |
| 4  | Влияние семантической информации на предсказание взгляда. . . .                                                                                                                                                                                                                                                                                                                                                                                                                       | 31 |
| 5  | Среднее значение AUROC по всем трем патологиям (норма, застойная сердечная недостаточность и пневмония) для каждой модели. . . . .                                                                                                                                                                                                                                                                                                                                                    | 44 |
| 6  | Информация о наборе данных для проведения эксперимента с рентгенологами. . . . .                                                                                                                                                                                                                                                                                                                                                                                                      | 54 |
| 7  | Коэффициенты корреляции Пирсона между характеристиками из данных взгляда и порядковым номером снимка во время анализа каждым рентгенологом. Признаки рассчитываются с использованием информации о взгляде рентгенолога и/или сегментации рентгеновского снимка на основе предсказания глубокой модели. Каждая ячейка таблицы содержит коэффициент корреляции человека с доверительным интервалом 95%. Статистически значимые коэффициенты корреляции выделены зеленым цветом. . . . . | 60 |
| 8  | Детали и клинические характеристики наборов данных для обучения, валидации и тестирования из базы WORD [82]. . . . .                                                                                                                                                                                                                                                                                                                                                                  | 77 |
| 9  | Сравнение качества сегментации на основе коэффициента Дайса (в процентах) для трех стратегий выбора точек взгляда. Жирным выделен наилучший результат для каждого органа. . . . .                                                                                                                                                                                                                                                                                                     | 79 |
| 10 | Сравнение коэффициента Дайса (в процентах) с современными моделями. . . . .                                                                                                                                                                                                                                                                                                                                                                                                           | 80 |
| 11 | Образцы срезов, отобранные для экспериментов с врачами. . . . .                                                                                                                                                                                                                                                                                                                                                                                                                       | 81 |
| 12 | Результаты коррекции масок с использованием взгляда для рентгенолога А. . . . .                                                                                                                                                                                                                                                                                                                                                                                                       | 82 |
| 13 | Результаты коррекции масок с использованием взгляда для рентгенолога В. . . . .                                                                                                                                                                                                                                                                                                                                                                                                       | 83 |

- 14 Средние значения коэффициента Дайса и времени (в мс) при использовании только ограничивающих рамок (до или без коррекции по взгляду) для рентгенологов А и В. . . . . 87

## Список литературы

1. Role of Artificial Intelligence in Medical Image Analysis: A Review of Current Trends and Future Directions / X. Li [и др.] // Journal of Medical and Biological Engineering. — 2024. — Апр. — Т. 44, № 2. — С. 231—243. — URL: <http://dx.doi.org/10.1007/s40846-024-00863-x>.
2. Revolutionizing healthcare: the role of artificial intelligence in clinical practice / S. A. Alowais [и др.] // BMC Medical Education. — 2023. — Сент. — Т. 23, № 1. — URL: <http://dx.doi.org/10.1186/s12909-023-04698-z>.
3. A Review of the Role of Artificial Intelligence in Healthcare / A. Al Kuwaiti [и др.] // Journal of Personalized Medicine. — 2023. — Июнь. — Т. 13, № 6. — С. 951. — URL: <http://dx.doi.org/10.3390/jpm13060951>.
4. FEASIBILITY OF USING ARTIFICIAL INTELLIGENCE IN RADIOLOGY (FIRST YEAR OF MOSCOW EXPERIMENT ON COMPUTER VISION) / S. Morozov [и др.] // Vrach i informacionnye tehnologii. — 2022. — № 1. — С. 12—29. — URL: [http://dx.doi.org/10.25881/18110193\\_2022\\_1\\_12](http://dx.doi.org/10.25881/18110193_2022_1_12).
5. *Chustecki M.* Benefits and Risks of AI in Health Care: Narrative Review // Interactive Journal of Medical Research. — 2024. — Ноябрь. — Т. 13. — e53616. — URL: <http://dx.doi.org/10.2196/53616>.
6. Do as AI say: susceptibility in deployment of clinical decision-aids / S. Gaube [и др.] // npj Digital Medicine. — 2021. — Февр. — Т. 4, № 1. — URL: <http://dx.doi.org/10.1038/s41746-021-00385-9>.
7. *Park S. H., Langlotz C. P.* Crucial Role of Understanding in Human-Artificial Intelligence Interaction for Successful Clinical Adoption // Korean Journal of Radiology. — 2025. — Т. 26, № 4. — С. 287. — URL: <http://dx.doi.org/10.3348/kjr.2025.0071>.
8. AI-based analysis of radiologist's eye movements for fatigue estimation: a pilot study on chest X-rays / I. Pershin [и др.] // Medical Imaging 2022: Image Perception, Observer Performance, and Technology Assessment / под ред. C. R. Mello-Thoms, S. Taylor-Phillips. — SPIE, 04.2022. — С. 39. — URL: <http://doi.org/10.1117/12.2612760>.

9. Artificial Intelligence for the Analysis of Workload-Related Changes in Radiologists' Gaze Patterns / I. Pershin [и др.] // IEEE Journal of Biomedical and Health Informatics. — 2022. — Сент. — Т. 26, № 9. — С. 4541—4550. — URL: <http://doi.org/10.1109/JBHI.2022.3183299>.
10. Changes in Radiologists' Gaze Patterns Against Lung X-rays with Different Abnormalities: a Randomized Experiment / I. Pershin [и др.] // Journal of Digital Imaging. — 2023. — Янв. — Т. 36, № 3. — С. 767—775. — URL: <http://doi.org/10.1007/s10278-022-00760-2>.
11. Gaze-Assisted Medical Image Segmentation / L. Khaertdinova [и др.] // Advances in Neural Information Processing Systems 37. — Vancouver, BC, Canada, 2024. — URL: <https://neurips.cc/virtual/2024/103920>.
12. Gaze Assistance for Efficient Segmentation Correction of Medical Images / L. Khaertdinova [и др.] // IEEE Access. — 2025. — Т. 13. — С. 14199—14213. — URL: <http://doi.org/10.1109/access.2025.3530701>.
13. Shedding light on ai in radiology: A systematic review and taxonomy of eye gaze-driven interpretability in deep learning / J. Neves [и др.] // European Journal of Radiology. — 2024. — Март. — Т. 172. — С. 111341. — URL: <http://dx.doi.org/10.1016/j.ejrad.2024.111341>.
14. Creation and Validation of a Chest X-Ray Dataset with Eye-tracking and Report Dictation for AI Tool Development / A. Karargyris [и др.]. — 2020.
15. REFLACX, a dataset of reports and eye-tracking data for localization of abnormalities in chest x-rays / R. Bigolin Lanfredi [и др.] // Scientific Data. — 2022. — Июнь. — Т. 9, № 1. — URL: <http://dx.doi.org/10.1038/s41597-022-01441-z>.
16. FG-CXR: A Radiologist-Aligned Gaze Dataset for Enhancing Interpretability in Chest X-Ray Report Generation / Т. Т. Pham [и др.] // Proceedings of the Asian Conference on Computer Vision (ACCV). — 12.2024. — С. 941—958.
17. *Halverson T., Hornof A. J.* A computational model of “active vision” for visual search in human–computer interaction // Human–Computer Interaction. — 2011. — Т. 26, № 4. — С. 285—314.
18. Modelling eye movements in a categorical search task / G. J. Zelinsky [и др.] // Philosophical Transactions of the Royal Society B: Biological Sciences. — 2013. — Т. 368, № 1628. — С. 20130058.

19. Modeling radiologists' cognitive processes using a digital gaze twin to enhance radiology training / A. Awasthi [и др.] // Scientific Reports. — 2025. — Т. 15, № 1. — С. 13685.
20. Gazesearch: radiology findings search benchmark / Т. Т. Pham [и др.] // 2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). — IEEE. 2025. — С. 96—106.
21. *Nguyen M. T., Siritanawan P., Kotani K.* Saliency Map Extraction in Human Crowd RGB Data // 2019 58th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE). — IEEE. 2019. — С. 941—946.
22. Gazeformer: Scalable, effective and fast prediction of goal-directed human attention / S. Mondal [и др.] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2023. — С. 1441—1450.
23. Unifying top-down and bottom-up scanpath prediction using transformers / Z. Yang [и др.] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2024. — С. 1683—1693.
24. *Chen X., Jiang M., Zhao Q.* Gazexplain: Learning to predict natural language explanations of visual scanpaths // European Conference on Computer Vision. — Springer. 2024. — С. 314—333.
25. Look Hear: Gaze Prediction for Speech-directed Human Attention / S. Mondal [и др.] // European Conference on Computer Vision. — Springer. 2024. — С. 236—255.
26. LLaMA: Open and Efficient Foundation Language Models / H. Touvron [и др.]. — 2023. — URL: <https://arxiv.org/abs/2302.13971>.
27. Towards interpreting visual information processing in vision-language models / C. Neo [и др.] // arXiv preprint arXiv:2410.07149. — 2024.
28. Learning Transferable Visual Models From Natural Language Supervision / A. Radford [и др.]. — 2021. — URL: <https://arxiv.org/abs/2103.00020>.
29. Visual instruction tuning / H. Liu [и др.] // Advances in neural information processing systems. — 2023. — Т. 36. — С. 34892—34916.

30. Enhancing Cognition and Explainability of Multimodal Foundation Models with Self-Synthesized Data / Y. Shi [и др.] // arXiv preprint arXiv:2502.14044. — 2025.
31. *Nostalgebraist*. Interpreting gpt: the logit lens. — 2020. — URL: <https://www.lesswrong.com/posts/AcKRB8wDpdaN6v6ru/interpreting-gpt-the-logit-lens>.
32. ScanMatch: A novel method for comparing fixation sequences / F. Cristino [и др.] // Behavior research methods. — 2010. — Т. 42. — С. 692—700.
33. *Needleman S. B., Wunsch C. D.* A general method applicable to the search for similarities in the amino acid sequence of two proteins // Journal of molecular biology. — 1970. — Т. 48, № 3. — С. 443—453.
34. *Brandt S. A., Stark L. W.* Spontaneous eye movements during visual imagery reflect the content of the visual scene // Journal of cognitive neuroscience. — 1997. — Т. 9, № 1. — С. 27—38.
35. *Yujian L., Bo L.* A normalized Levenshtein distance metric // IEEE transactions on pattern analysis and machine intelligence. — 2007. — Т. 29, № 6. — С. 1091—1095.
36. Simulating human saccadic scanpaths on natural images / W. Wang [и др.] // CVPR 2011. — IEEE. 2011. — С. 441—448.
37. It depends on how you look at it: Scanpath comparison in multiple dimensions with MultiMatch, a vector-based approach / R. Dewhurst [и др.] // Behavior research methods. — 2012. — Т. 44. — С. 1079—1100.
38. Coco-search18 fixation dataset for predicting goal-directed attention control / Y. Chen [и др.] // Scientific reports. — 2021. — Т. 11, № 1. — С. 8776.
39. *Lvov D., Pershin I.* LogitGaze: Predicting Human Attention Using Semantic Information from Vision-Language Models // Workshop on Reasoning and Planning for Large Language Models. — 2025.
40. A survey on deep learning in medical image analysis / G. Litjens [и др.] // Medical Image Analysis. — 2017. — Т. 42. — С. 60—88.

41. Explainable medical imaging AI needs human-centered design: guidelines and evidence from a systematic review / H. Chen [и др.] // npj Digital Medicine. — 2022. — Окт. — Т. 5, № 1. — URL: <http://dx.doi.org/10.1038/s41746-022-00699-2>.
42. *Harezlak K., Kasprowski P.* Application of eye tracking in medicine: A survey, research issues and challenges // Computerized Medical Imaging and Graphics. — 2018. — Апр. — Т. 65. — С. 176—190. — URL: <http://dx.doi.org/10.1016/j.compmedimag.2017.04.006>.
43. A collaborative computer aided diagnosis (C-CAD) system with eye-tracking, sparse attentional model, and deep learning / N. Khosravan [и др.] // Medical Image Analysis. — 2019. — Янв. — Т. 51. — С. 101—115. — URL: <http://dx.doi.org/10.1016/j.media.2018.10.010>.
44. Automatic Lung Nodule Detection Combined With Gaze Information Improves Radiologists' Screening Performance / G. Aresta [и др.] // IEEE Journal of Biomedical and Health Informatics. — 2020. — Окт. — Т. 24, № 10. — С. 2894—2901. — URL: <http://dx.doi.org/10.1109/JBHI.2020.2976150>.
45. Mimic-cxr database / A. Johnson [и др.] // PhysioNet10. — 2024. — Т. 13026. — C2JT1Q.
46. Contour-aware multi-label chest X-ray organ segmentation / M. Kholiavchenko [и др.] // International Journal of Computer Assisted Radiology and Surgery. — 2020. — Февр. — Т. 15, № 3. — С. 425—436. — URL: <http://dx.doi.org/10.1007/s11548-019-02115-9>.
47. How visual search relates to visual diagnostic performance: a narrative systematic review of eye-tracking research in radiology / A. van der Gijp [и др.] // Advances in Health Sciences Education. — 2016. — Июль. — Т. 22, № 3. — С. 765—787. — URL: <http://dx.doi.org/10.1007/s10459-016-9698-1>.
48. Long Radiology Workdays Reduce Detection and Accommodation Accuracy / E. A. Krupinski [и др.] // Journal of the American College of Radiology. — 2010. — Сент. — Т. 7, № 9. — С. 698—704. — URL: <http://dx.doi.org/10.1016/j.jacr.2010.03.004>.

49. The Effect of Faster Reporting Speed for Imaging Studies on the Number of Misses and Interpretation Errors: A Pilot Study / E. Sokolovskaya [и др.] // Journal of the American College of Radiology. — 2015. — Июль. — Т. 12, № 7. — С. 683—688. — URL: <http://dx.doi.org/10.1016/j.jacr.2015.03.040>.
50. CT colonography interpretation times: effect of reader experience, fatigue, and scan findings in a multi-centre setting / D. Burling [и др.] // European Radiology. — 2006. — Апр. — Т. 16, № 8. — С. 1745—1749. — URL: <http://dx.doi.org/10.1007/s00330-006-0190-9>.
51. *Bruno M. A., Walker E. A., Abujudeh H. H.* Understanding and Confronting Our Mistakes: The Epidemiology of Error in Radiology and Strategies for Error Reduction // RadioGraphics. — 2015. — Окт. — Т. 35, № 6. — С. 1668—1676. — URL: <http://dx.doi.org/10.1148/rg.2015150023>.
52. The Effects of Fatigue From Overnight Shifts on Radiology Search Patterns and Diagnostic Performance / T. N. Hanna [и др.] // Journal of the American College of Radiology. — 2018. — Дек. — Т. 15, № 12. — С. 1709—1716. — URL: <http://dx.doi.org/10.1016/j.jacr.2017.12.019>.
53. The Impact of Fatigue on Complex CT Case Interpretation by Radiology Residents / H. Zhan [и др.] // Academic Radiology. — 2021. — Март. — Т. 28, № 3. — С. 424—432. — URL: <http://dx.doi.org/10.1016/j.acra.2020.06.005>.
54. Differentiating Surgeons' Expertise solely by Eye Movement Features. — ACM, 10.2021. — С. 371—375. — (ICMI '21). — URL: <http://dx.doi.org/10.1145/3461615.3485437>.
55. Differences in gaze behaviour of expert and junior surgeons performing open inguinal hernia repair / T. Tien [и др.] // Surgical Endoscopy. — 2014. — Авг. — Т. 29, № 2. — С. 405—413. — URL: <http://dx.doi.org/10.1007/s00464-014-3683-7>.
56. Pupil diameter changes reflect difficulty and diagnostic accuracy during medical image interpretation / T. T. Brunye [и др.] // BMC Medical Informatics and Decision Making. — 2016. — Июль. — Т. 16, № 1. — URL: <http://dx.doi.org/10.1186/s12911-016-0322-3>.

57. Pupil diameter differentiates expertise in dental radiography visual search / N. Castner [и др.] // PLOS ONE / под ред. S. Martinez-Conde. — 2020. — Май. — Т. 15, № 5. — e0223941. — URL: <http://dx.doi.org/10.1371/journal.pone.0223941>.
58. *Jaeger J.* Digit Symbol Substitution Test: The Case for Sensitivity Over Specificity in Neuropsychological Testing // Journal of Clinical Psychopharmacology. — 2018. — Окт. — Т. 38, № 5. — С. 513—519. — URL: <http://dx.doi.org/10.1097/JCP.0000000000000941>.
59. Digital biomarker of mental fatigue / V. W.-S. Tseng [и др.] // npj Digital Medicine. — 2021. — Март. — Т. 4, № 1. — URL: <http://dx.doi.org/10.1038/s41746-021-00415-6>.
60. VinDr-CXR: An open dataset of chest X-rays with radiologist's annotations / H. Q. Nguyen [и др.] // Scientific Data. — 2022. — Июль. — Т. 9, № 1. — URL: <http://dx.doi.org/10.1038/s41597-022-01498-w>.
61. SU-E-I-62: Investigation of Dominant Factors Affecting Fatigue in Image Reading of Radiologists / Y. Ikushima [и др.] // Medical Physics. — 2012. — ИЮНЬ. — Т. 39, 6Part5. — С. 3639—3639. — URL: <http://dx.doi.org/10.1118/1.4734778>.
62. *Klein J. S., Rosado-de-Christenson M. L.* A Systematic Approach to Chest Radiographic Analysis // Diseases of the Chest, Breast, Heart and Vessels 2019-2022. — Springer International Publishing, 2019. — С. 1—16. — URL: [http://dx.doi.org/10.1007/978-3-030-11149-6\\_1](http://dx.doi.org/10.1007/978-3-030-11149-6_1).
63. *Ronneberger O., Fischer P., Brox T.* U-Net: Convolutional Networks for Biomedical Image Segmentation // Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. — Springer International Publishing, 2015. — С. 234—241. — URL: [http://dx.doi.org/10.1007/978-3-319-24574-4\\_28](http://dx.doi.org/10.1007/978-3-319-24574-4_28).
64. Artifact Removal using Improved GoogLeNet for Sparse-view CT Reconstruction / S. Xie [и др.] // Scientific Reports. — 2018. — Апп. — Т. 8, № 1. — URL: <http://dx.doi.org/10.1038/s41598-018-25153-w>.
65. ImageNet: A large-scale hierarchical image database / J. Deng [и др.] // 2009 IEEE Conference on Computer Vision and Pattern Recognition. — IEEE, 06.2009. — URL: <http://dx.doi.org/10.1109/CVPR.2009.5206848>.

66. *Ginneken B. van, Stegmann M. B., Loog M.* Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database // *Medical Image Analysis*. — 2006. — Февр. — Т. 10, № 1. — С. 19—40. — URL: <http://dx.doi.org/10.1016/j.media.2005.02.002>.
67. Development of a Digital Image Database for Chest Radiographs With and Without a Lung Nodule: Receiver Operating Characteristic Analysis of Radiologists' Detection of Pulmonary Nodules / J. Shiraishi [и др.] // *American Journal of Roentgenology*. — 2000. — ЯНВ. — Т. 174, № 1. — С. 71—74. — URL: <http://dx.doi.org/10.2214/ajr.174.1.1740071>.
68. Data Augmentation for Chest Pathologies Classification / I. Sirazitdinov [и др.] // 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). — IEEE, 04.2019. — С. 1216—1219. — URL: <http://dx.doi.org/10.1109/ISBI.2019.8759573>.
69. *Kingma D. P., Ba J.* Adam: A Method for Stochastic Optimization. — 2014. — URL: <https://arxiv.org/abs/1412.6980>.
70. Searching for Lung Nodules A Comparison of Human Performance with Random and Systematic Scanning Models / H. L. KUNDEL [и др.] // *Investigative Radiology*. — 1987. — Май. — Т. 22, № 5. — С. 417—422. — URL: <http://dx.doi.org/10.1097/00004424-198705000-00010>.
71. What do experts look at and what do experts find when reading mammograms? / J. M. Wolfe [и др.] // *Journal of Medical Imaging*. — 2021. — ИЮЛЬ. — Т. 8, № 04. — URL: <http://dx.doi.org/10.1117/1.JMI.8.4.045501>.
72. *Strasburger H., Rentschler I., Jüttner M.* Peripheral vision and pattern recognition: A review // *Journal of Vision*. — 2011. — Дек. — Т. 11, № 5. — С. 13. — URL: <http://dx.doi.org/10.1167/11.5.13>.
73. The Impairing Effect of Mental Fatigue on Visual Sustained Attention under Monotonous Multi-Object Visual Attention Task in Long Durations: An Event-Related Potential Based Study / Z. Guo [и др.] // *PLOS ONE* / под ред. F. Di Russo. — 2016. — СЕНТ. — Т. 11, № 9. — e0163360. — URL: <http://dx.doi.org/10.1371/journal.pone.0163360>.

74. *Faber L. G., Maurits N. M., Lorist M. M.* Mental Fatigue Affects Visual Selective Attention // PLoS ONE / под ред. F. P. de Lange. — 2012. — Окт. — Т. 7, № 10. — e48073. — URL: <http://dx.doi.org/10.1371/journal.pone.0048073>.
75. Deep Learning in Radiology / M. P. McBee [и др.] // Academic Radiology. — 2018. — Т. 25, № 11. — С. 1472—1480. — URL: <https://www.sciencedirect.com/science/article/pii/S1076633218301041>.
76. A review of semantic segmentation using deep neural networks / Y. Guo [и др.] // International journal of multimedia information retrieval. — 2018. — Т. 7. — С. 87—93.
77. Image segmentation using deep learning: A survey / S. Minaee [и др.] // IEEE transactions on pattern analysis and machine intelligence. — 2021. — Т. 44, № 7. — С. 3523—3542.
78. Deep learning for medical image segmentation: State-of-the-art advancements and challenges / M. E. Rayed [и др.] // Informatics in Medicine Unlocked. — 2024. — С. 101504.
79. *Li G., Wu X., Ma X.* Artificial intelligence in radiotherapy // Seminars in Cancer Biology. — 2022. — Нояб. — Т. 86. — С. 160—171. — URL: <https://doi.org/10.1016/j.semancer.2022.08.005>.
80. *Sharma N., Aggarwal L. M.* Automated medical image segmentation techniques // Journal of medical physics. — 2010. — Т. 35, № 1. — С. 3—14.
81. A deep learning-based auto-segmentation system for organs-at-risk on whole-body computed tomography images for radiation therapy / X. Chen [и др.] // Radiotherapy and Oncology. — 2021. — Т. 160. — С. 175—184.
82. WORD: A large scale dataset, benchmark and clinical applicable study for abdominal organ segmentation from CT image / X. Luo [и др.] // arXiv preprint arXiv:2111.02403. — 2021.
83. A Comprehensive Survey on Segment Anything Model for Vision and Beyond / C. Zhang [и др.]. — 2023. — URL: <https://arxiv.org/abs/2305.08196>.
84. Segment anything in medical images / J. Ma [и др.] // Nature Communications. — 2024. — Т. 15, № 1. — С. 654.

85. Segment anything model (sam) for radiation oncology / L. Zhang [и др.] // arXiv preprint arXiv:2306.11730. — 2023.
86. Variations in the contouring of organs at risk: test case from a patient with oropharyngeal cancer / B. E. Nelms [и др.] // International Journal of Radiation Oncology\* Biology\* Physics. — 2012. — Т. 82, № 1. — С. 368—378.
87. Manual segmentation versus semi-automated segmentation for quantifying vestibular schwannoma volume on MRI / H. McGrath [и др.] // International journal of computer assisted radiology and surgery. — 2020. — Т. 15. — С. 1445—1455.
88. A deep-learning approach for segmentation of liver tumors in magnetic resonance imaging using UNet++ / J. Wang [и др.] // BMC cancer. — 2023. — Т. 23, № 1. — С. 1060.
89. *Summers R. M.* Progress in fully automated abdominal CT interpretation // American Journal of Roentgenology. — 2016. — Т. 207, № 1. — С. 67—79.
90. Automatic breast and fibroglandular tissue segmentation in breast MRI using deep learning by a fully-convolutional residual neural network U-net / Y. Zhang [и др.] // Academic radiology. — 2019. — Т. 26, № 11. — С. 1526—1535.
91. DENSE-INception U-net for medical image segmentation / Z. Zhang [и др.] // Computer methods and programs in biomedicine. — 2020. — Т. 192. — С. 105395.
92. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation / F. Isensee [и др.] // Nature methods. — 2021. — Т. 18, № 2. — С. 203—211.
93. *Tekchandani H., Verma S., Londhe N. D.* Improving the detection of abdominal and mediastinal lymph nodes in CT images using attention U-Net based deep learning model // Artificial Intelligence Applications for Health Care. — CRC Press, 2022. — С. 181—202.
94. Computer aided diagnosis system for cervical lymph nodes in CT images using deep learning / H. Tekchandani [и др.] // Biomedical Signal Processing and Control. — 2022. — Т. 71. — С. 103158.

95. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation / F. Isensee [и др.] // Nature Methods. — 2020. — Дек. — Т. 18, № 2. — С. 203—211. — URL: <http://dx.doi.org/10.1038/s41592-020-01008-z>.
96. Semi-automated segmentation of pre-operative low grade gliomas in magnetic resonance imaging / Z. Akkus [и др.] // Cancer Imaging. — 2015. — Т. 15. — С. 1—10.
97. Improving abdominal image segmentation with overcomplete shape priors / A. Sadikine [и др.] // Computerized Medical Imaging and Graphics. — 2024. — Т. 113. — С. 102356.
98. Gaze2Segment: a pilot study for integrating eye-tracking technology into medical image segmentation / N. Khosravan [и др.] // Medical Computer Vision and Bayesian and Graphical Models for Biomedical Imaging: MICCAI 2016 International Workshops, MCV and BAMBI, Athens, Greece, October 21, 2016, Revised Selected Papers 8. — Springer. 2017. — С. 94—104.
99. Hands-free interactive image segmentation using eyegaze // Medical Imaging 2009: Computer-Aided Diagnosis. Т. 7260. — SPIE. 2009. — С. 441—450.
100. Integrating eye tracking and speech recognition accurately annotates MR brain images for deep learning: proof of principle / J. N. Stember [и др.] // Radiology: Artificial Intelligence. — 2020. — Т. 3, № 1. — e200047.
101. *Budd S., Robinson E. C., Kainz B.* A survey on active learning and human-in-the-loop deep learning for medical image analysis // Medical image analysis. — 2021. — Т. 71. — С. 102062.
102. Gaze-probe joint guidance with multi-task learning in obstetric ultrasound scanning / Q. Men [и др.] // Medical image analysis. — 2023. — Т. 90. — С. 102981.
103. Mammo-net: Integrating gaze supervision and interactive information in multi-view mammogram classification / C. Ji [и др.] // International Conference on Medical Image Computing and Computer-Assisted Intervention. — Springer. 2023. — С. 68—78.
104. Supporting Mitosis Detection AI Training with Inter-Observer Eye-Gaze Consistencies / H. Gu [и др.] // 2024 IEEE 12th International Conference on Healthcare Informatics (ICHI). — IEEE. 2024. — С. 40—45.

105. *Ibragimov B., Mello-Thoms C.* The Use of Machine Learning in Eye Tracking Studies in Medical Imaging: A Review // IEEE Journal of Biomedical and Health Informatics. — 2024.
106. *Sofiuk K., Petrov I. A., Konushin A.* Reviving iterative training with mask guidance for interactive segmentation // 2022 IEEE International Conference on Image Processing (ICIP). — IEEE. 2022. — С. 3141—3145.
107. *Boykov Y., Jolly M.-P.* Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images // Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001. T. 1. — 2001. — 105—112 vol.1.
108. *Grady L.* Random Walks for Image Segmentation // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2006. — Т. 28, № 11. — С. 1768—1783.
109. Segment anything / A. Kirillov [и др.] // Proceedings of the IEEE/CVF International Conference on Computer Vision. — 2023. — С. 4015—4026.
110. Segment Everything Everywhere All at Once / X. Zou [и др.]. — 2023. — URL: <https://arxiv.org/abs/2304.06718>.
111. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale / A. Dosovitskiy [и др.] // ICLR. — 2021.
112. Segment Anything Model for Medical Image Analysis: an Experimental Study / M. A. Mazurowski [и др.] // Medical Image Analysis. — 2023. — URL: <https://arxiv.org/abs/2304.10517>.
113. Medical sam adapter: Adapting segment anything model for medical image segmentation / J. Wu [и др.] // arXiv preprint arXiv:2304.12620. — 2023.
114. Scribbleprompt: Fast and flexible interactive segmentation for any medical image / H. E. Wong [и др.] // arXiv preprint arXiv:2312.07381. — 2023.
115. *Zhu J., Qi Y., Wu J.* Medical sam 2: Segment medical images as video via segment anything model 2 // arXiv preprint arXiv:2408.00874. — 2024.
116. Sam 2: Segment anything in images and videos / N. Ravi [и др.] // arXiv preprint arXiv:2408.00714. — 2024.

117. Boxinst: High-performance instance segmentation with box annotations / Z. Tian [и др.] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2021. — С. 5443—5452.
118. *Cheng B., Parkhi O., Kirillov A.* Pointly-supervised instance segmentation // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2022. — С. 2617—2626.
119. GazeSAM: Interactive Image Segmentation with Eye Gaze and Segment Anything Model / B. Wang [и др.] // NeuRIPS 2023 Workshop on Gaze Meets ML. — 2023. — URL: <https://openreview.net/forum?id=hJ5DREWdjs>.
120. Eye tracking technology in medical practice: a perspective on its diverse applications / M. Tahri Sqalli [и др.] // Frontiers in Medical Technology. — 2023. — Т. 5. — С. 1253001.
121. *Loshchilov I.* Decoupled weight decay regularization // arXiv preprint arXiv:1711.05101. — 2017.
122. *Blignaut P., Beelders T.* The effect of fixational eye movements on fixation identification with a dispersion-based fixation detection algorithm // Journal of eye movement research. — 2008. — Т. 2, № 5.
123. ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data / F. I. Diakogiannis [и др.] // ISPRS Journal of Photogrammetry and Remote Sensing. — 2020. — Т. 162. — С. 94—114.
124. CT-ORG, a new dataset for multiple organ segmentation in computed tomography / B. Rister [и др.] // Scientific Data. — 2020. — Т. 7, № 1. — С. 381.
125. Attention u-net: Learning where to look for the pancreas / O. Oktay [и др.] // arXiv preprint arXiv:1804.03999. — 2018.
126. Amos: A large-scale abdominal multi-organ benchmark for versatile medical image segmentation / Y. Ji [и др.] // Advances in neural information processing systems. — 2022. — Т. 35. — С. 36722—36732.
127. Medlsam: Localize and segment anything model for 3d medical images / W. Lei [и др.] // arXiv preprint arXiv:2306.14752. — 2023.