



Применение детекторов RF-DETR и Yolo26 в задаче локализации опухолей головного мозга по МРТ-данным

Виталий Петрович Фраленко

Институт программных систем им. А. К. Айламазяна РАН, Веськово, Россия

Аннотация. Исследование посвящено сравнительному анализу современных детекторов объектов для локализации опухолей головного мозга по данным магнитно-резонансной томографии.

В первой части проведен аналитический обзор современных подходов к неригидной регистрации и сегментации опухолей, а также архитектур трансформерных и гибридных детекторов объектов с деформируемым вниманием, многоканальной агрегацией признаков и специализированными функциями потерь, ориентированных на обработку деформируемых структур с размытыми границами, а также анализ программных реализаций и лицензионных ограничений соответствующих фреймворков.

Во второй части выполнено экспериментальное сравнение нейронных сетей RF-DETR и Yolo26 на размеченном датасете опухолей головного мозга в задаче детекции ограничивающих прямоугольников. Отметим, что данная задача представляет собой стандартную процедуру локализации области интереса и не требует специального учета деформируемости структур или неопределенности границ, в отличие от задач сегментации или неригидной регистрации. Оценка качества проведена по метрикам mAP0.50, mAP0.50–0.95 и F1-мерам для IoU 0.50 и 0.50–0.95 при переборе порогов уверенности.

Полученные результаты показали, что RF-DETR обеспечивает более высокие интегральные показатели качества и лучшую скорость по сравнению с Yolo26, что подтверждает, что трансформерные детекторы являются перспективным инструментом автоматизированного анализа МРТ-изображений головного мозга. На основе аналитического и экспериментального исследований сформулированы рекомендации по развитию RF-DETR для задач медицинской визуализации, направленные на повышение обобщающей способности модели, устойчивости к сдвигу распределения данных и клинической интерпретируемости результатов.

Ключевые слова и фразы: опухоль головного мозга, нейронная сеть, трансформер, аналитический обзор, детекция объектов, RF-DETR, Yolo26, медицинская визуализация

Благодарности: Работа выполнена в рамках госбюджетной темы ИПС им. А.К. Айламазяна РАН № 125021302067-9 (срок выполнения: 2025–2027 гг.)

Для цитирования: Фраленко В.П. *Применение детекторов RF-DETR и Yolo26 в задаче локализации опухолей головного мозга по МРТ-данным* // Программные системы: теория и приложения. 2026. Т. 17. № 2(71). С. 263–293. https://psta.psiras.ru/read/psta2026_2_263–293.pdf

Введение

Обнаружение опухолей головного мозга на изображениях, полученных с помощью магнитно-резонансной томографии (МРТ), является одной из ключевых задач современной медицинской визуализации. Опухолевые образования относятся к категории неригидных структур, характеризующихся значительными вариациями формы, границ и текстуры, обусловленными неоднородным ростом, отеком окружающих тканей, постоперационными изменениями, артефактами сканирования и инфильтративным характером распространения. Эти особенности принципиально отличают опухоли от ригидных анатомических структур (например, костей черепа или тел позвонков) и формируют каскад вызовов для традиционных методов компьютерного зрения, ориентированных преимущественно на фиксированные объекты.

В условиях роста объемов нейровизуализационных данных и дефицита квалифицированных радиологов возрастает потребность в автоматизированных системах анализа и поддержки принятия решений, способных устойчиво работать с деформируемыми структурами и данными различных модальностей. Настоящая работа направлена на систематический анализ указанных проблем и оценку применимости современных детекторов объектов к задаче автоматизированной локализации опухолей головного мозга на МРТ-изображениях.

Далее будет представлен обзор подходов к неригидной регистрации и сегментации опухолей мозга – от байесовских и физико-математических моделей до нейросетевых архитектур, ориентированных на точное моделирование деформаций и изменчивости опухолевых структур. Затем в работе рассматриваются архитектурные особенности трансформерных и гибридных детекторов, в том числе успешно применяемых к неригидным объектам с неустойчивыми границами (таким как огонь и дым), включая современные решения с многоуровневой агрегацией признаков и оптимизацией к вычислениям в режиме реального времени, обсуждается возможность переноса этих решений в медицинский домен. Особое внимание в обзорной части уделяется программным аспектам рассмотренных архитектур: инженерной доступности исходного кода, лицензионным ограничениям и возможности воспроизводимого сравнительного анализа на единой вычислительной платформе, что критически важно для выбора базовой модели в условиях клинического внедрения.

В настоящем исследовании датасет *Labeled MRI brain Tumor*¹ (лицензия CC BY 4.0), содержащий 2175 изображений (550 здоровых образцов,

¹*Labeled MRI brain Tumor dataset. Kaggle, 2023*^{URL}

455 с глиомой, 550 с менингиомой, 620 с опухолями гипофиза), был стратифицирован по четырем классам и случайным образом разделен на семь блоков одинакового размера с сохранением пропорций классов в каждом блоке. Пять блоков (1557 изображений) объединены в обучающую выборку, один блок (309 изображений) использован для валидации в процессе обучения, один блок (309 изображений) – для итогового тестирования обученных нейросетей. Кросс-валидация не проводилась, поскольку целью являлось сравнение моделей на фиксированном тестовом наборе по воспроизводимому протоколу.

Отметим, что описание, данное на странице датасета, некорректное, в настоящей работе приведена точная информация. Примеры опухолей мозга из датасета приведены на рисунке 1.

Отметим, что в дальнейшем в настоящем исследовании используется ряд общепринятых терминов. «Голова» нейронной сети (от англ. head) – финальные слои, которые получают признаки как от бэкбона (от англ. backbone), так и зачастую от «шеи» (от англ. neck), и выполняют целевую задачу: классификацию, регрессию или предсказание ключевых точек в задачах вроде детекции объектов. Бэкбон отвечает за извлечение многоуровневых карт признаков (от англ. feature maps) из входных данных, таких как изображения, и, как правило, реализуется в виде предобученной сверточной или трансформерной нейросети. «Шея» (от англ. neck) служит для агрегации и перераспределения признаков с разных уровней бэкбона, обеспечивая их согласование по масштабу и семантике.

Следует подчеркнуть, что настоящее исследование носит обзорно-экспериментальный характер и построено по двухуровневой схеме.

На первом уровне проводится аналитический обзор широкого спектра методов искусственного интеллекта в медицинской визуализации: от байесовских и физико-математических моделей неригидной регистрации до гибридных трансформерно-сверточных архитектур сегментации и детекции. Этот обзор систематизирует архитектурные решения, применяемые к деформируемым структурам с размытыми границами, и задает предметный контекст для оценки детекторов объектов.

На втором уровне выполняется сравнительный эксперимент наиболее перспективных представителей двух доминирующих парадигм детекции: трансформерной (RF-DETR [1]) и сверточной (Yolo26 [2]). Выбор моделей обусловлен их статусом флагманов своих классов и наличием в RF-DETR механизма деформируемого внимания, теоретически способствующего точной локализации объектов сложной формы. Таким образом, обзорная часть обеспечивает систематизацию методов, тогда как экспериментальная

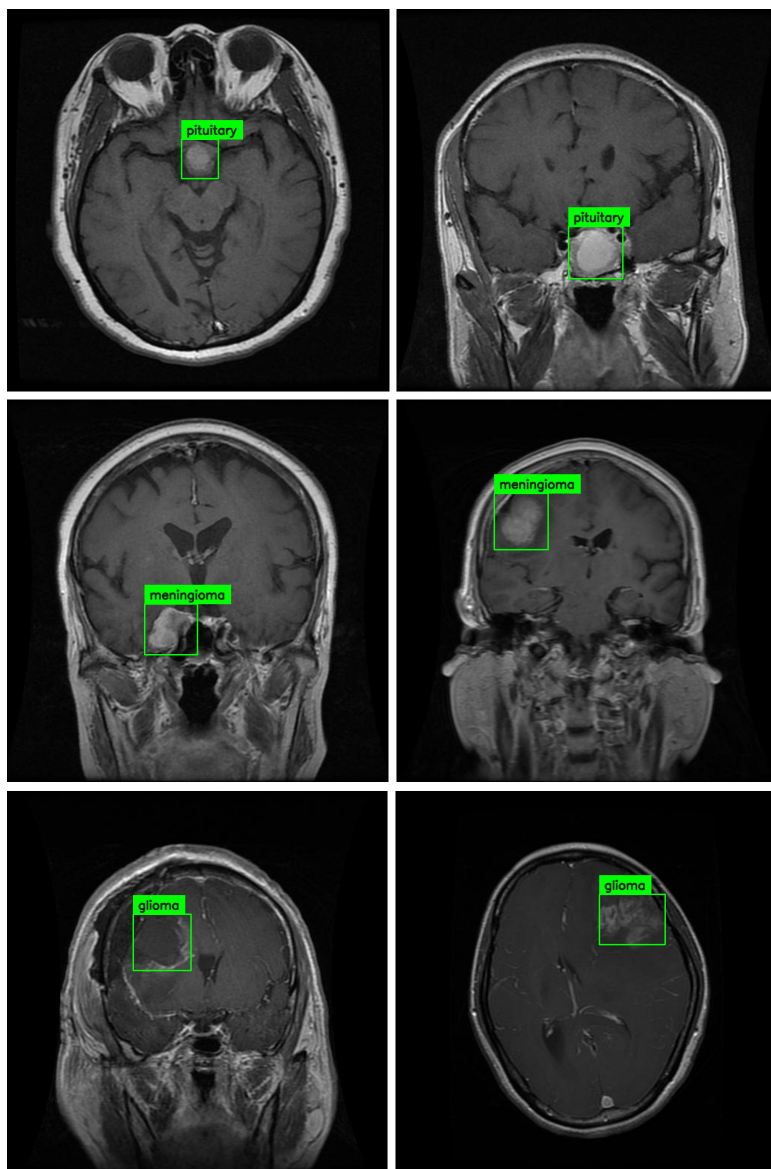


РИСУНОК 1. Примеры опухолей гипофиза, менингиом и глиом

дает количественную оценку их представителей в стандартной постановке детекции ограничивающих прямоугольников, которая не требует специаль-

ного учета деформируемости структур и одинаково применима к объектам любой природы.

1. Неригидная регистрация и сегментация опухолей мозга, архитектурные инновации трансформерных и гибридных детекторов

В работе [3] опухоли рассматриваются как неригидные объекты, подверженные деформациям во времени под влиянием лечения. Авторы предлагают новаторский метод совместной оценки карты меток (от англ. label map) и неригидной регистрации для выравнивания лонгитюдных МРТ-изображений пациентов с опухолями мозга, учитывающий изменяющиеся соответствия особенностей (от англ. feature correspondences), то есть непостоянные во времени взаимные соответствия между анатомическими или структурными признаками на последовательных снимках, вызванные ростом, смещением и деформацией опухолей. Предложен фреймворк в байесовской парадигме с моделями схожести изображений, априорными распределениями интенсивности, сглаживанием карты меток и трансформационным приором (то есть заранее заданным допущением о свойствах искомой деформации, например, ее гладкости или степени гибкости), который поощряет более свободные деформации в области опухоли. Тестирование на синтетических данных показало снижение ошибки целевой регистрации (Target Registration Error, TRE) на 29%, а на реальных лонгитюдных МРТ-сериях пациентов с глиобластомами – рост коэффициента Дайса (Dice) для опухолевой зоны на 12–18%.

В статье [4] представлена адаптивная неригидная регистрация пре- и интраоперационных МРТ-изображений, ориентированная на процесс удаления опухоли. Метод основан на гетерогенной биомеханической модели конечных элементов, которая в реальном времени адаптируется к типу ткани и стадии резекции. Модель воспроизводит процесс хирургического удаления опухоли и динамически обновляется для компенсации деформаций окружающих тканей, что позволяет точнее соответствовать наблюдаемым на МРТ-изображениях смещениям и сжатиям. В 14 клинических случаях показано существенное снижение показателя TRE: до семи раз относительно ригидной регистрации и до пяти раз по сравнению с неадаптивной физически обоснованной неригидной регистрацией (Physics-Based Non-Rigid Registration, PBNRR) из библиотеки ИТК², особенно в области опухоли после резекции.

² ИТК | Insight Toolkit, 2026[©]

Фреймворк SIBIA, предложенный в работе [5], – интеграция биофизической реакционно-диффузионной модели роста глиобластомы с методом Large Deformation Diffeomorphic Metric Mapping (LDDMM). Метод основывается на эйлеровом представлении полей скоростей, он обеспечивает диффеоморфную (гладкую и обратимую) регистрацию медицинских изображений, параметризуя большие деформации через поток эйлеровых полей скоростей и минимизируя функционал расстояния в метрике Соболева для поддержания гладкости и топологической согласованности преобразований. В фреймворке для регистрации используется нормальный атлас мозга и набор МРТ-изображений пациента (T1, T1ce, T2, FLAIR) с аннотированными областями опухоли, инвазии, некроза и отека. Это позволяет выполнять согласованное совмещение данных, корректно моделируя топологические изменения, вызванные ростом опухоли, без разрывов и искажений анатомических структур.

Экспериментальная валидация SIBIA на синтетических данных показала относительную ошибку реконструкции концентрации опухолевых клеток в пределах 1–5% и восстановление параметров реакционно-диффузионной модели с ошибкой менее 10%. На реальных МРТ-данных пациентов с глиобластомой совместная биофизико-диффеоморфная регистрация обеспечила прирост коэффициента Дайса на 0.05–0.10 для анатомических структур по сравнению с чисто диффеоморфной регистрацией без учета модели роста опухоли.

В работе [6] представлена концепция Non-Enhancing Tumor (NET) – деформируемой структуры с высокой морфологической вариабельностью. Авторы предлагают модификацию архитектуры U-Net, направленную на повышение пространственного разрешения сегментационных масок и более точное выделение неконтрастных границ опухолей. Особое внимание уделено неригидности опухолевых контуров, рассматриваемой как ключевой вызов, требующий использования деформируемых методов обработки изображений. Экспериментальные результаты, полученные на наборах данных BraTS 2018 и BraTS 2021, демонстрируют улучшение метрик качества: +12–15% для коэффициента Дайса и +10% для коэффициента Жаккара (Intersection over Union, IoU) по сравнению с базовой моделью U-Net, особенно в случаях мультифокальных и фрагментированных опухолевых очагов.

Следует отметить, что архитектурные тенденции в области нейронных сетей в последнее время смещаются в сторону гибридных трансформерно-сверточных моделей, в которых глобальное контекстное моделирование, обеспечиваемое механизмами внимания трансформера, сочетается с локальной инвариантностью и вычислительной эффективностью сверточных

операций. Такой подход позволяет объединить преимущества обеих парадигм – способность трансформеров улавливать дальнodelействующие зависимости и структурную избирательность сверточных слоев.

Классическая архитектура нейронной сети-трансформера, предложенная в работе [7], является моделью, полностью основанной на механизмах внимания. Она состоит из энкодера и декодера, каждый из которых содержит несколько одинаковых по структуре блоков, включающих многоголовое самовнимание (от англ. multi-head self-attention), позиционные кодировки (от англ. positional embeddings), а также полносвязные слои с остаточными связями (от англ. residual connections) и нормализацией (от англ. layer normalization). Благодаря этой архитектуре трансформеры способны эффективно моделировать долгосрочные зависимости в последовательностях, не требуя поэтапной обработки данных. Это открывает путь к развитию современных крупных языковых и мультимодальных моделей.

Использование деформируемого внимания (от англ. deformable attention) представляет собой типичный пример трансформер-сверточного подхода, реализующего интеграцию преимуществ механизмов внимания и сверточных нейронных сетей (СНС). В архитектурах нейронных сетей, подобных Deformable DETR [8], деформируемое внимание применяется поверх или совместно с СНС-бэкбонами, выступающими в роли базовых экстракторов признаков. Такой подход обеспечивает эффективное моделирование контекстной информации при одновременном снижении вычислительной сложности по сравнению с классическим глобальным вниманием.

В отличие от стандартных трансформеров, где декодер анализирует все точки признаковой карты, деформируемое внимание ограничивает область поиска набором многомасштабных ключевых точек (от англ. multi-scale key points). Вокруг каждой из них вычисляется внимание: для небольшой группы заранее предсказанных позиций (от англ. sampling locations) формируется взвешенное суммирование признаков, где коэффициенты весов определяются на основе степени релевантности каждой позиции текущему запросу. Благодаря этому модель концентрируется на информативных областях изображения, минимизируя избыточные вычисления.

Ключевая особенность деформируемого внимания заключается в его динамическом характере: смещения, масштабы и веса для каждой «головы» внимания определяются непосредственно на основе входных признаков. Такой механизм позволяет адаптировать форму эффективного рецептивного поля к локальным геометрическим деформациям объектов, повышая устойчивость модели к вариациям формы, положения и масштаба.

Подобные методы доказали высокую эффективность при решении задач детектирования объектов, особенно в области дистанционного зондирования, где наблюдаются значительные геометрические и текстурные искажения (см., например, [9]).

Отдельное направление современных исследований связано с разработкой гибридных модулей многократного внимания (от англ. hybrid multi-attention), которые объединяют несколько типов механизмов внимания – пространственное, канальное и временное – в единой трансформерной архитектуре. Такие модули могут использоваться как в блоках самовнимания (от англ. self-attention), так и в блоках взаимного внимания (от англ. cross-attention).

Механизм самовнимания обеспечивает вычисление весов внимания между элементами одной и той же последовательности (например, токенами в тексте или пикселями в изображении), что позволяет модели эффективно захватывать зависимости внутри входных данных. Типичный пример – энкодер в архитектуре трансформера.

Взаимное внимание, напротив, отражает взаимодействие между различными последовательностями: запросы (от англ. queries) из одной последовательности сопоставляются с ключами и значениями (от англ. keys и values) из другой. Такой подход традиционно применяется, например, в задачах машинного перевода, где декодер концентрируется на наиболее релевантных участках контекста, сформированного энкодером.

В гибридных архитектурах сверточные слои, как правило, используются для извлечения низкоуровневых локальных признаков, тогда как трансформерные блоки с гибридным вниманием повторно интерпретируют и обобщают эти признаки с учетом временной динамики. Это позволяет снижать уровень шума и устранять неоднозначности, характерные для объектов с неригидной структурой на последовательностях кадров. Так, в исследовании архитектуры НуМАТОД [10] показано, что оптимизированная обработка признаков и более избирательное внимание к фону позволяют достичь высоких значений показателя mean Average Precision (mAP) при одновременном снижении вычислительных затрат. Базовый модуль НуМАТ, используемый в данной архитектуре, усиливает релевантные корреляции и подавляет ошибочную информацию за счет согласования (консенсуса) между векторами корреляций внимания из разных «голов» и типов механизмов внимания. Благодаря универсальности конструкции, модуль может быть интегрирован как в блоки самовнимания, так и во взаимные блоки внимания, обеспечивая дополнительную устойчивость модели к изменению контекста и фона.

Эффективные алгоритмы обнаружения и исследования динамики опухолей могут быть получены путем перенесения и адаптации отработанных приемов и подходов, применяемых при распознавании объектов совершенно другой природы, – с резко выраженной неригидностью. Так в работе [11] представлена модификация архитектуры YOLOv8 под названием YOLOGX, специально адаптированная для детекции лесных пожаров. Основным направлением улучшения модели стала оптимизация механизма извлечения и слияния признаков.

В модифицированной части «шеи» нейросетевой архитектуры реализован механизм Gather-Distribution (GD), включающий две параллельные ветви, Low-GD и High-GD, отличающиеся назначением: Low-GD ориентирована на локальные, мелкомасштабные детали (например, текстуры пламени), в то время как High-GD акцентирует глобальный контекст и крупномасштабные структуры (например, обширные дымовые облака). В каждой из этих ветвей последовательно применяются три функциональных модуля: Feature Alignment Module (FAM) для начального выравнивания пространственных и семантических несоответствий признаков из разных уровней пирамиды признаков; Information Fusion Module (IFM) для эффективного слияния выровненных локальных и глобальных представлений через взаимное внимание и агрегацию; и Information Injection Module (IIM) для точечной инъекции ключевой информации обратно в основной поток для усиления релевантных информативных признаков без потери контекста. Совместное действие этих модулей обеспечивает более эффективное объединение локальных и глобальных признаков, что повышает точность детекции объектов различных масштабов без существенного увеличения вычислительной сложности.

В блок «головы» модели разработчики интегрировали усовершенствованный модуль SE-ResNeXt, использующий групповые свертки (32 группы) и механизм Squeeze-and-Excitation. Это позволило одновременно сократить количество параметров и усилить межканальные зависимости признаков, улучшая способность сети различать тонкие визуальные различия между огнем, дымом и фоном. Кроме того, в YOLOGX используется новая функция потерь Focal-SIoU (Focal Scaled Intersection over Union), учитывающая не только степень пересечения рамок (IoU), но также угловое отклонение и расстояние между центрами предсказанной и истинной рамок. Такой подход способствует более стабильной сходимости и повышает точность локализации целевых объектов. Эффективность YOLOGX проверена на специализированном датасете D-Fire, включающем 21527 изображений с четырьмя категориями: огонь, дым, комбинированные сцены и фон.

В результате модель достигла показателей $mAP_{0.50} = 0.8092$ при скорости 115 кадров в секунду, превзойдя оригинальную YOLOv8 ($mAP_{0.50} = 0.7946$), а также специализированные решения вроде Fire-YOLO и LFNet. Таким образом, YOLOGX демонстрирует, что целевая оптимизация архитектуры под специфику задачи и введение многоуровневых механизмов слияния признаков способны заметно повысить эффективность и устойчивость моделей детекции.

В работе [12] представлена облегченная версия алгоритма детекции пламени, предназначенная для применения на вычислительно ограниченных устройствах (edge-системах) в режиме реального времени. Архитектура Light-YOLOv8-Flame основана на модифицированной модели YOLOv8 и предполагает замену стандартных C2f-модулей в бэкбоне на блок FasterNet, использующий комбинацию решений Partial Convolution (PConv 3×3) и Pointwise Convolution (PWConv 1×1). Partial Convolution в данном контексте означает «частичную» свертку, при которой обычная свертка 3×3 выполняется не по всем каналам входного тензора, а лишь по их подмножеству (например, по каждому четвертому каналу). Это позволяет уменьшить число операций, поскольку большая часть каналов не проходит через вычислительно затратную 3×3 -свертку; оставшиеся каналы затем объединяются с «обработанными». Pointwise Convolution (1×1 -свертка) используется для смешивания информации между каналами без изменения пространственного разрешения и практически не увеличивает вычислительную нагрузку.

В составе FasterNet блок PConv отвечает за извлечение пространственных признаков на ограниченном подмножестве каналов, а PWConv — за последующую агрегацию и перераспределение этих признаков по всем каналам, что обеспечивает эффективное выделение пространственных особенностей при минимальных вычислительных затратах. Для обучения и оценки модели авторами исследования был сформирован собственный датасет, включающий 7431 изображение и разделенный в соотношении 10:1:1 на обучающую, тестовую и валидационную выборки.

По результатам экспериментов Light-YOLOv8-Flame демонстрирует повышение показателя mAP на 0.78% и рост полноты (от англ. recall) на 2.05% при одновременном уменьшении числа параметров модели на 25.34% и снижении точности (от англ. precision) всего на 0.82%. При этом достигается скорость обработки около 80 кадров в секунду на видеокарте Nvidia RTX 3070. В сравнении с базовыми версиями YOLOv8n/s/m предложенная модель обеспечивает скорость, сопоставимую с YOLOv8n, при уровне точности, близком к YOLOv8s. Использование функции потерь CloU-loss (Complete Intersection over Union loss) повышает устойчивость

к сложным фонам и обеспечивает уверенное распознавание небольших или удаленных очагов огня.

RF-DETR [1] – инновационная архитектура для детекции объектов в реальном времени, основывающаяся на трансформерной модели DETR. В RF-DETR используется стохастическое варьирование архитектурных гиперпараметров во время обучения, что позволяет достичь оптимального баланса между точностью и вычислительной эффективностью, адаптируя конфигурацию сети к конкретным требованиям задачи. В основе этой нейронной сети лежит модифицированная структура LW-DETR, в которой бэкбон CAEv2 заменен на более мощную сеть DINOv2, предварительно обученную на датасете LVD-142M. Эта замена обеспечивает выделение сильных визуальных приоров и повышает способность модели к обобщению на данных, отличающихся от тех, что были в обучающем датасете (так называемых данных out-of-distribution).

Приоры – не знание конкретных классов, а универсальные визуальные представления (например, контуры, формы, текстуры), которые обладают инвариантностью к фону, освещению и другим доменным факторам. Такие представления хорошо переносятся между задачами и доменами.

Обработка изображения начинается с разбиения его на равномерные фрагменты (патчи), которые интерпретируются как последовательность токенов и подаются в бэкбон на основе Vision Transformer (ViT). ViT адаптирует трансформерную архитектуру, изначально разработанную для обработки текста, к задачам компьютерного зрения, применяя механизм самовнимания для извлечения пространственных зависимостей между токенами. Полученные признаки далее поступают в слои энкодера с оконным вниманием и в группы декодеров, где формируются итоговые предсказания.

Для повышения эффективности моделирования контекста в RF-DETR реализовано деформируемое внимание как в энкодере, так и в декодере. Этот механизм позволяет модели уделять вычислительные ресурсы наиболее информативным областям изображения, что особенно важно при наличии окклюзий и сложных многосоставных сцен. В отличие от традиционных архитектур, структура RF-DETR не задается заранее. В процессе нейронного поиска архитектуры динамически исследуется пространство возможных конфигураций, оптимизируя ряд параметров.

Размер патча варьируется с использованием механизма FlexiViT, который интерполирует масштаб изображений и подбирает оптимальное соотношение между скоростью и точностью. Глубина декодера изменяется с применением регрессионной функции потерь на каждом уровне, что дает

возможность при инференсе сокращать число слоев без существенной потери качества. Одновременно оптимизируется количество запросных токенов, определяющее число объектов, которые модель способна анализировать параллельно; этот показатель балансируется между точностью и латентностью, а во время вывода система может автоматически исключать токены с низкой уверенностью.

Кроме того, при обучении подбирается разрешение входного изображения с учетом вычислительных ограничений, корректируя позиционные эмбединги через интерполяцию для сохранения пространственной согласованности. Наконец, размер окна внимания регулируется для адаптации архитектуры к сценам различной плотности и сложности. Обучение модели сопровождается стохастическим выбором значений указанных параметров на каждой итерации, что создает эффект архитектурной аугментации.

Этот механизм выполняет роль регуляризатора, повышая устойчивость сети к изменению статистики данных и улучшая ее способность к обобщению. Таким образом, RF-DETR представляет собой гибкую и адаптивную архитектуру, способную эффективно решать задачи детекции объектов в реальном времени при сохранении высокого уровня точности и оптимальной производительности.

Yolo26 [2] – новейшая итерация семейства моделей YOLO, разработанная компанией Ultralytics для задач детекции объектов в реальном времени, с упором на использование в вычислительных системах с ограниченными ресурсами и низким энергопотреблением. В архитектуре Yolo26 произведено радикальное упрощение модели за счет полного устранения модуля Distribution Focal Loss (DFL), что позволило снизить вычислительную сложность регрессии ограничивающих прямоугольников и повысить совместимость модели с форматами экспорта, включая TensorFlow Lite, CoreML, OpenVINO, TensorRT и ONNX. Это решение также устранило проблемы, возникающие при работе с динамическими механизмами внимания вне среды PyTorch.

Ключевым архитектурным нововведением Yolo26 является нативная сквозная (от англ. end-to-end) структура выходной «головы», функционирующая без этапа подавления немаксимумов (Non-Maximum Suppression, NMS). Такая схема способствует снижению задержки инференса на центральном процессоре до 43% по сравнению с предыдущими версиями (11 и 12, соответственно), а также повышению пропускной способности при работе на встроенных устройствах, IoT-платформах, беспилотниках и робототехнических системах. Дополнительно в Yolo26 реализован новый оптимизатор MuSGD, ускоряющий достижение целевого состояния (или

целевых метрик, таких как mAP) при обучении; сохранена поддержка квантизации весов в форматах INT8 и FP16, что обеспечивает оптимальный баланс между производительностью и точностью при сжатии модели для развертывания на edge-устройствах.

Аналитический обзор показывает, что для неригидных объектов с размытыми границами (опухоли головного мозга, очаги инфильтрации, деформируемые структуры) ключевыми являются три архитектурных свойства:

- (1) способность к глобальному контекстному моделированию для удержания анатомической когерентности;
- (2) адаптация рецептивных полей к локальным геометрическим деформациям;
- (3) устойчивость к сдвигу распределения данных и клиническим артефактам.

Трансформерные архитектуры с деформируемым вниманием (Deformable DETR, RF-DETR) теоретически обладают преимуществом в первых двух свойствах благодаря механизмам самовнимания и динамическим смещениям опорных точек. Сверточные архитектуры последнего поколения (Yolo26) достигают конкурентоспособной точности за счет глубокой оптимизации вычислительного графа и end-to-end-структуры без NMS, что критично для скоростной обработки потоков МРТ-данных.

Отдельно стоит отметить, что при написании настоящей работы выявлена принципиальная гетерогенность программных платформ, на которых реализованы подходы к анализу неригидных структур с размытыми границами, причем эта гетерогенность определяет не столько математические возможности моделей, сколько их инженерную доступность и пригодность для воспроизводимого сравнительного анализа. Физико-математические и байесовские модели неригидной регистрации, ориентированные на совместную оценку деформаций и карт меток, традиционно конструируются в рамках библиотеки ИТК, реализованной на языке C++ и распространяемой под лицензией Apache 2.0³. Программная реализация таких моделей предполагает оперирование абстракциями `itk::Image` и `itk::Transform`, построение пользовательских конвейеров фильтрации и ручную настройку параметров регуляризации, что обеспечивает строгий контроль над биомеханическими ограничениями, но одновременно создает высокий порог входа для исследователя, не владеющего системным программированием.

³ ИТК: *The Insight Toolkit*. Github, 2026^{URL}

Адаптивные физически обоснованные методы регистрации, учитывающие динамику резекции опухоли, требуют интеграции конечно-элементного моделирования с конвейером обработки медицинских изображений, что технически реализуется как надстройка над ядром ИТК с привлечением внешних вычислительных библиотек для решения систем линейных уравнений. Фреймворки, сочетающие биофизическое моделирование роста опухоли с диффеоморфной регистрацией, программно выглядят как гибридные конвейеры, где часть компонентов отвечает за численное интегрирование реакционно-диффузионных уравнений, а часть – за параметризацию полей скоростей в метрике Соболева; при этом полноценная программная реализация, позволяющая воспроизвести результаты без участия авторов, фактически отсутствует в открытом доступе, что ограничивает возможность прямого сравнения с нейросетевыми подходами на единой вычислительной платформе. Таким образом, физико-математическая парадигма демонстрирует высокую теоретическую точность и физическую интерпретируемость, но низкую программную мобильность и сложность интеграции в конвейеры автоматизированной детекции, ориентированные на обработку потоковых данных в реальном времени.

Сверточные архитектуры сегментации обычно разрабатываются в рамках стандартных экосистем глубокого обучения, преимущественно PyTorch и TensorFlow. Программно эти модели реализуются как последовательности стандартных слоев Conv2d, BatchNorm и Upsample. При этом конкретная модификация архитектуры нейросети в работе, описывающей концепцию Non-Enhancing Tumor, в открытом доступе не представлена, что характерно для ряда академических публикаций в данной области.

Воспроизведение подобных модификаций возможно на базе общедоступных реализаций U-Net, например, в репозитории Pytorch-UNet⁴, где исследователь может модифицировать архитектуру декодера, менять функции активации и внедрять механизмы повышения пространственного разрешения без необходимости разработки низкоуровневых операторов. Однако программная экосистема сегментационных сетей ориентирована на работу с полносвязными масками, тогда как детекция объектов с помощью прямоугольных рамок требует иной постановки функции потерь и иной структуры выходного слоя, что делает прямое заимствование программного кода затруднительным.

Трансформерные детекторы представляют собой иную программную парадигму, базирующуюся на механизмах внимания и требующую управления тензорами высокой размерности. Архитектура, реализующая

⁴ U-Net: Semantic segmentation with PyTorch. Github, 2024 

деформируемое внимание, программно встраивается в стандартный трансформерный энкодер-декодер через пользовательские слои, вычисляющие смещения опорных точек относительно регулярной сетки, что требует либо компиляции специализированных CUDA-расширений, либо использования готовых операторов, например, из фреймворков Detectron2 или MMDetection; отметим, однако, что оригинальная реализация операторов деформируемого внимания доступна в репозитории проекта Deformable DETR⁵. Программная специфика таких решений заключается в необходимости согласования размерностей тензоров запросов, ключей и значений при многомасштабной обработке признаков, а также в корректной интерполяции позиционных кодировок при изменении разрешения входных изображений.

Гибридные архитектуры с механизмами многократного внимания, объединяющие пространственное, канальное и временное внимание, программно реализуются как модификации стандартных блоков MultiheadAttention⁶, с добавлением консенсусных модулей, согласующих векторы корреляций между различными головами; такие модули функционируют как в блоках самовнимания, так и в блоках взаимного внимания. Отметим, что прямая реализация подобных гибридных модулей в открытом доступе отсутствует; переход в медицинский домен требует замены бэкбона, перерасчета позиционных эмбеддингов, а особенности реализации таких переходов фактически не документированы в официальных инструкциях.

Архитектурные модификации детекторов на базе архитектуры YOLO обычно предполагают перестройку «шей» с внедрением параллельных ветвей сбора и распределения признаков, и программно реализуются как переопределение классов стандартных слоев в директории модулей нейронной сети⁷, где механизмы слияния локальных и глобальных признаков встраиваются между бэкбоном и «головой» детектора через дополнительные модули выравнивания и инъекции информации.

Программное внедрение механизмов групповых сверток с последующим перекалибровочным взвешиванием каналов по типу Squeeze-and-Excitation требует модификации файлов, отвечающих за построение бэкбона и «головы». Функции потерь, учитывающие геометрические отношения между предсказанными и истинными рамками, внедряются на уровне python-кода обучающего конвейера путем переопределения класса, отвечающего за вычисление регрессионной составляющей ошибки. Важной особенностью

⁵ Deformable DETR.Github, 2020^{URL}

⁶ MultiheadAttention – PyTorch: Tensors and Dynamic neural networks in Python with strong GPU acceleration.Github, 2026^{URL}

⁷ Ultralytics nn modules – Ultralytics YOLO.Github, 2026^{URL}

программной экосистемы Ultralytics является наличие готовых скриптов обучения, валидации и экспорта, что обеспечивает быстрое развертывание на целевых устройствах, но одновременно ограничивает исследователя жестко заданной структурой API и лицензионными рамками.

Сравнение программных реализаций RF-DETR и Yolo26, выбранных в качестве ключевых объектов собственного экспериментального исследования, выявляет фундаментальное различие в философии открытого кода и инженерной культуре сопровождения. Программный пакет RF-DETR, развиваемый компанией Roboflow, распространяется под лицензией Apache 2.0 через публичный репозиторий на GitHub⁸ и построен на стеке PyTorch. Архитектура данного детектора предусматривает стохастическое варьирование гиперпараметров на уровне обучающего скрипта, динамическое изменение глубины декодера и интерполяцию позиционных эмбедингов при варьировании разрешения входных изображений, что реализовано python-средствами и не требует компиляции пользовательских CUDA-операторов, но предъявляет повышенные требования к стабильности процедур оптимизации и контролю за динамикой градиентов.

В отличие от этого, программные коды Yolo26, разрабатываемые компанией Ultralytics, распространяются под лицензией AGPL-3.0⁹, что накладывает обязательства по открытию производного кода при коммерческом использовании модификаций. Таким образом, RF-DETR и Yolo26 различаются не только архитектурно, но и экологически: первая модель предлагает большую свободу архитектурной модификации и перестройки «шей» при менее жестких лицензионных ограничениях, тогда как вторая обеспечивает глубокую интеграцию с промышленными форматами развертывания на встраиваемых системах, но в рамках более строгой лицензионной политики и фиксированной программной иерархии.

Анализ программных паттернов, лежащих в основе рассмотренных архитектур, позволяет выделить три уровня технологической зрелости, имеющие непосредственное отношение к задаче детекции опухолей головного мозга.

На первом уровне находятся физико-математические модели регистрации, функционирующие как специализированные научные библиотеки, требующие значительных усилий по интеграции и часто не имеющие единого централизованного репозитория с открытым исходным кодом; их применимость к задаче быстрой детекции

⁸ RF-DETR: Real-Time SOTA Detection and Segmentation. Github, 2026^(url)

⁹ Ultralytics YOLO. Github, 2026^(url)

ограничивается принципиальной ориентацией на пиксельное выравнивание и сегментацию, а не на локализацию областей интереса прямоугольными рамками.

На втором уровне располагаются гибридные трансформерные архитектуры, развивающиеся в рамках академических репозиториях с использованием PyTorch и обеспечивающие воспроизводимость базовых экспериментов, но нуждающиеся в адаптации позиционных эмбеддингов и механизмов внимания под медицинские изображения с их специфической интенсивностью и анатомическим контекстом.

На третьем уровне концентрируются оптимизированные сверточные детекторы и современные флагманские модели, предлагающие готовые конвейеры обучения и стандартизированные конфигурационные файлы, но ограничивающие исследователя фиксированной структурой программного интерфейса и требованиями лицензионного соглашения.

Учитывая данную программную дифференциацию, переход от обзорных архитектурных решений к собственному экспериментальному сравнению требует осознанного выбора программной парадигмы.

Модели RF-DETR и Yolo26 выбраны в качестве альтернативных полюсов современной детекции: первая модель представляет парадигму внимания с предобученными визуальными приорами (DINOv2) и лояльной лицензией Apache 2.0, вторая – парадигму эффективных сверточных детекторов, ориентированных на edge-вычисления, но распространяемую в рамках строгой лицензии AGPL-3.0. Сравнение этих моделей на едином медицинском датасете позволит дать оценку применимости двух фундаментально разных классов методов искусственного интеллекта к клинической задаче детекции неригидных опухолей, одновременно учитывая инженерные и лицензионные ограничения их программных реализаций.

2. Экспериментальное исследование и его результаты

В проведенных экспериментах решалась задача детекции прямоугольных окон, содержащих целевые опухолевые области, без последующей сегментации. Эксперименты проводились на рабочей станции, оснащенной процессором Intel Core i5 12600k, 32 ГБ памяти DDR-5 с частотой 6400 МГц и видеокартой Nvidia RTX 5070 Ti с 16 ГБ видеопамяти. Базовые системные библиотеки: python v3.13.11 и pytorch v2.11.0+cu130.

Важно отметить специфику постановки задачи, выбранной в настоящем исследовании. В отличие от сегментации, где требуется пиксельная

точность выделения границ и где вариативность межэкспертной разметки критически влияет на метрики Dice или IoU, задача детекции ограничивающих прямоугольников (от англ. bounding box detection) принципиально менее чувствительна к точному определению контура опухоли. Цель детекции – надежная локализация области интереса, содержащей патологическое образование, что соответствует клиническому сценарию первичного скрининга и маршрутизации снимков на дальнейшую экспертизу. Используемый датасет *Labeled MRI brain Tumor* содержит разметку, выполненную в соответствии со стандартизированным протоколом аннотирования МРТ-снимков, при этом представленные в нем классы (глиомы, менингиомы, опухоли гипофиза) отражают естественную морфологическую вариативность исследуемых структур, включая мультифокальные очаги и инфильтративные границы. В этом смысле детекция прямоугольных областей служит обоснованным первым этапом: она позволяет оценить способность архитектуры устойчиво обнаруживать присутствие опухоли независимо от субъективных вариаций точного контура. Перспективным направлением, выходящим за рамки настоящей работы, является переход к сегментации с использованием размытых меток (от англ. soft labels), учитывающих неопределенность границ при аннотировании опухолевых образований.

Всего проведено семь экспериментов, в которых исследовались

- RF-DETR версии 1.3.0;
- RF-DETR версии 1.4.2;
- RF-DETR версии 1.6.5 (с аугментацией данных);
- Yolo26 с выключенной аугментацией по яркости и оптимизатором AdamW;
- Yolo26 с выключенной аугментацией по яркости и оптимизатором MuSGD;
- Yolo26 с включенной аугментацией по яркости и оптимизатором AdamW;
- Yolo26 с включенной аугментацией по яркости и оптимизатором MuSGD.

Основные параметры, использованные в экспериментах с RF-DETR¹⁰:

- базовая модель: RFDETRLarge;
- максимальное число эпох обучения (epochs): 100;
- эпоха, на которой скорость обучения снизится до 0.1 от базовой скорости (lr_drop): 83;

¹⁰ *Training Parameters – RF-DETR, 2026* 

- число стартовых эпох обучения, на которых скорость обучения растёт от минимального до целевого (warmup_epochs): 3;
- число изображений в батче (batch_size): 9;
- количество шагов накопления градиентов в процессе обучения нейросети (grad_accum_steps): 4;
- скорость обучения (lr): 3e-5;
- скорость обучения энкодерной части модели (lr_encoder): 1.5e-5;
- высота и ширина окна модели (resolution): 704;
- остановка обучения при плато mAP0.50–0.95 (early_stopping): True;
- число эпох без улучшения mAP0.50–0.95, после которого будет ранний останов (early_stopping_patience): 10;
- минимальная дельта mAP0.50–0.95 на валидационной выборке при сравнении с ранее достигнутым лучшим значением, чтобы сбросить счетчик эпох без улучшения (early_stopping_min_delta): 0.001;
- использование метрики mAP от EMA-модели (Exponential Moving Average) для принятия решений об остановке обучения нейросети (early_stopping_use_ema): True;
- конфиг аугментации (только для RF-DETR 1.6.5) (aug_config):
 {«HorizontalFlip»: {«p»: 0.25}, «VerticalFlip»: {«p»: 0.1},
 "RandomRotate90": {«p»: 0.1}, «Affine»: {«scale»: {«x»: (0.85, 1.15),
 «y»: (0.85, 1.15)}, «rotate»: (-7.5, 7.5), «shear»: (-7.5, 7.5), «p»: 0.25}}.

Основные параметры, используемые в экспериментах с Yolo26¹¹:

- базовая модель: Yolo26 Large (yolo26l.pt);
- оптимизатор обучения (optimizer): AdamW или MuSGD;
- косинусное изменение learning rate (cos_lr): True;
- стартовая скорость обучения (lr): 0.001;
- конечный множитель для скорости обучения (lrf): 0.001;
- число изображений в батче (batch): 9;
- максимальное число эпох обучения (epochs): 1000;
- высота и ширина окна модели (imgsz): 768;
- прямоугольный режим (rect): False;
- использование данных в формате FP16 (half): True;
- вариативность значения V в цветовой схеме HSV (hsv_v): 0 или 0.4;
- повороты обучающих изображений в градусах (degrees): 7.5;
- сдвиги обучающих изображений в радианах (translate): 0.0125;

¹¹ *Ultralytics YOLO Руководство по настройке гиперпараметров – Ultralytics YOLO Docs, 2026*^[9]

- масштабирование обучающих изображений (scale): 0.1;
- искажение обучающих изображений в радианах (shear): 0.0125.

Для RF-DETR и Yolo26 выполнялся перебор порогов уверенности (от 0.001 до 0.951 с шагом 0.0025), отсекающих обнаруженные целевые объекты, и выбиралась конфигурация нейросети с лучшим значением F1-меры с IoU 0.50–0.95. На этой же нейросети затем считалась F1-мера с IoU 0.50, но уже с полученным порогом уверенности. Критерии mAP0.50 и mAP0.50–0.95 считались на лучшей нейросети с порогом уверенности 0.001, что соответствует принятому в COCO протоколу оценки. Результаты экспериментов приведены в таблице 1.

ТАБЛИЦА 1. Результаты экспериментального исследования RF-DETR и Yolo26

	RF-DETR			Yolo26 в Ultralytics v8.4.21			
	v1.3.0	v1.4.2	v1.6.5	без аугм. яркости		с аугм. яркости	
				AdamW	MuSGD	AdamW	MuSGD
F1 IoU 0.50	0.944	0.923	0.927	0.852	0.919	0.862	0.892
F1 IoU 0.50–0.95	0.608	0.603	0.612	0.536	0.571	0.527	0.569
mAP 0.50	0.944	0.927	0.936	0.905	0.801	0.887	0.845
mAP 0.50–0.95	0.562	0.554	0.561	0.513	0.512	0.484	0.531
Порог уверенности	0.5185	0.3760	0.5660	0.3435	0.1085	0.0460	0.2210

Важно отметить, что в экспериментах с RF-DETR использовался тюнинг весовых коэффициентов предобученной на COCO-датасете Large-конфигурации нейросети, а не обучение заново, поэтому и resolution использовался равным 704, как в базовой RFDETRLarge. Обучение с нуля требует использования более сложных стратегий обучения и стабилизации градиентов. Повышенная скорость обучения приводит к NaN в тензорах из-за градиентного взрыва и численных неустойчивостей; как показали эксперименты, косинусоидальный планировщик не предотвращает локальные проблемы уже в процессе обновления весов где-нибудь в середине обучения, поэтому разработчикам RF-DETR следует оптимизировать процедуру обучения нейросети, например, через градиентный клиппинг (от англ. clipping), мониторинг появления NaN с откатом к последнему валидному состоянию и снижением скорости обучения. Возможность корректной работы с предобученной моделью и настраиваемым resolution реализовали только в RF-DETR v1.6.5, соответственно¹².

¹² RF-DETR 1.6.5: fixing EMA & PE.Github, 2026^{url}

F1-мера с IoU 0.50 – F1-мера на оптимальном пороге уверенности (найденном путем перебора), при котором из всех сырых предсказаний (с порогом уверенности $\text{score} \geq 0.001$) отбираются только уверенные ($\text{score} \geq$ оптимального порога), а затем каждый предсказанный прямоугольник сравнивается с GT-прямоугольниками (истинными ограничивающими рамками) по фиксированному $\text{IoU} = 0.50$: истинно положительными (TP) засчитываются прямоугольники с перекрытием $\geq 50\%$, ложно положительными (FP) – лишние предсказания, ложно отрицательными (FN) – пропущенные объекты. F1-мера с IoU 0.50–0.95 – аналогичная метрика, усредненная по десяти порогам IoU от 0.50 до 0.95; она оценивает способность модели выдавать более точно локализованные ограничивающие рамки. Предложенные F1-меры более информативны, чем $\text{mAP}_{0.50}$ и $\text{mAP}_{0.50-0.95}$, потому что они считаются с оптимизацией порога уверенности по реальной задаче, используя F1-меру с IoU 0.50–0.95 как фитнес-функцию, а не полагаются на стандартный COCO-подход с кривыми точность-полнота, что позволяет с их помощью показать качество работы модели в клиническом сценарии. Эксперименты показали, что, используя порог уверенности, можно добиться высоких значений F1-меры даже при относительно низком $\text{mAP}_{0.50}$.

Полученные результаты (см. таблицу 1) позволяют провести интерпретацию преимуществ RF-DETR через призму архитектурных особенностей, рассмотренных в обзорной части. Превосходство RF-DETR v1.3.0 по метрике $\text{mAP}_{0.50-0.95}$ (0.562 против 0.531 у лучшей конфигурации Yolo26) свидетельствует о более точной локализации ограничивающих рамок. Данный результат коррелирует с наличием в архитектуре RF-DETR механизма деформируемого внимания, который адаптирует рецептивное поле к локальной геометрии объекта.

Использование предобученного бэббона DINOv2 в RF-DETR обеспечивает перенос визуальных представлений с крупномасштабного датасета LVD-142M, что, как показывают полученные метрики, способствует лучшему обобщению на медицинские изображения при сдвиге распределения данных. В свою очередь, Yolo26 демонстрирует компромисс между вычислительной эффективностью и точностью локализации: отказ от этапа подавления немаксимумов и упрощение регрессионной головы снижают задержку инференса, однако ограничивают способность модели к точной локализации при малых размерах объектов или низком контрасте.

Стохастическое варьирование гиперпараметров в процессе обучения RF-DETR выполняет роль архитектурной регуляризации, повышая обобщаемость на редкие морфологические варианты опухолей. Напротив, Yolo26, достигая высокой скорости за счет радикального упрощения

головы детектора и отказа от NMS, демонстрирует компромисс: снижение вычислительной сложности регрессии ограничивает его способность к точной локализации сложных объектов, особенно при малых размерах или низком контрасте.

Таким образом, проведенное тестирование подтверждает возможность использования обученной нейросети как вспомогательного инструмента при диагностике опухолей головного мозга по данным МРТ. Лучшие результаты, в целом, получены с использованием RF-DETR v1.3.0. Отметим, что RF-DETR v1.6.5 чуть более точно выделяет зоны с целевыми объектами, но эта разница в пределах погрешности. RF-DETR в среднем обрабатывает 149 файлов МРТ в секунду, а Yolo26 лишь 112. Тесты скорости проводились на батчах по 48 файлов, время чтения файлов с твердотельного диска в данном случае было исключено.

Процесс обучения RF-DETR версии 1.3.0 демонстрирует более стабильную сходимость по сравнению с версией 1.6.5, обучаемой с дополнительной аугментацией данных, что особенно заметно на рисунках 2–3, где визуализи-

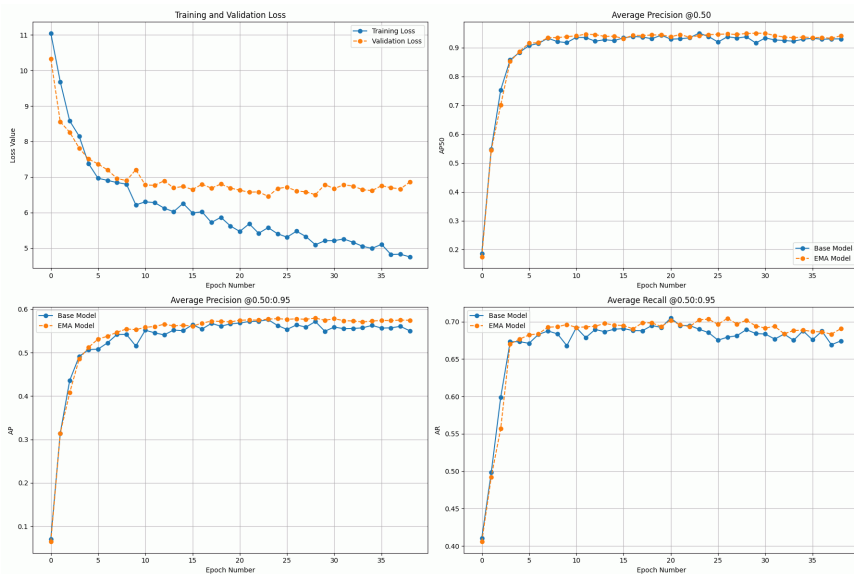


Рисунок 2. Графическое описание процесса обучения RF-DETR версии 1.3.0

зированы динамика функции потерь на обучающей и валидационной выборках (Training Loss и Validation Loss), а также метрик качества

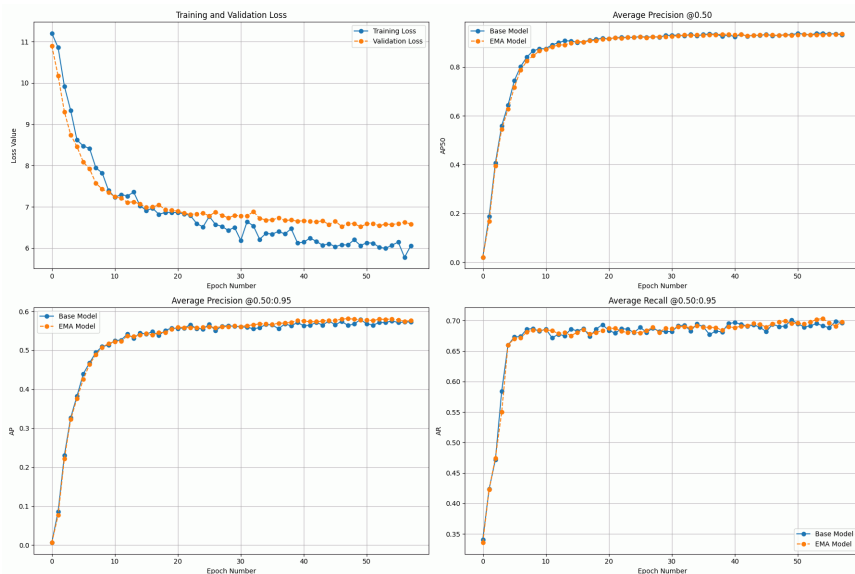


РИСУНОК 3. Графическое описание процесса обучения RF-DETR версии 1.6.5

в зависимости от эпохи обучения (Average Precision и Average Recall, соответственно, но только для валидационной выборки).

Отметим, что при анализе статистических данных, приведенных на рисунках, необходимо учитывать различия в масштабах осей. При анализе метрик качества предлагается оценить ЕМА-модели (см. оранжевые графики). ЕМА-модели (от англ. Exponential Moving Average models) – сглаженные копии основной обучаемой модели, параметры которой обновляются не по стандартному градиентному спуску, а через экспоненциальное скользящее среднее с определенным коэффициентом импульса.

Функция потерь для версии 1.3.0 быстро снижается с первых эпох и стабилизируется на низком уровне без значительных колебаний после 10-й эпохи, при этом разрыв между кривыми Training Loss и Validation Loss остается минимальным, указывая на отсутствие признаков выраженного переобучения и хорошую приспособленность базовой архитектуры к вариативности МРТ-изображений опухолей мозга с их размытыми границами. В версии 1.6.5 функция потерь на обучающей выборке также падает резко, но на валидационной остается заметно выше и показывает небольшие осцилляции к концу обучения (около эпохи 40) с большим

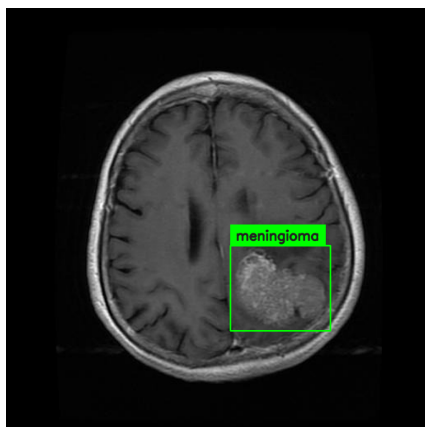
расхождением между кривыми train/val, что может свидетельствовать о несколько худшей обобщающей способности модели под влиянием расширенной аугментации, имитирующей клинические артефакты и сдвиги интенсивности. Average Precision @0.50 для версии 1.3.0 растет плавно и достигает высокого плато ($\sim 0.7\text{--}0.8$) уже к середине обучения, подтверждая быструю конвергенцию модели на датасете с неравномерным распределением опухолевых объектов. Версия 1.6.5 показывает более крутой начальный подъем Average Precision @0.50:0.95, но затем замедляется и заканчивает на схожем или чуть более низком уровне с заметными флуктуациями.

Полная метрика Average Precision @0.50:0.95 для версии 1.3.0 демонстрирует устойчивый рост до ~ 0.5 с минимальными отклонениями от оптимального поведения, что говорит о балансе модели в задачах детекции деформируемых структур; в 1.6.5 эта метрика растет медленнее после пика и имеет больший разрыв относительно ожидаемых значений (~ 0.1). Average Recall в версии 1.3.0 показывает резкий скачок уже на третьей эпохе с последующим монотонным и стабильным нарастанием до высоких значений ($\sim 0.75\text{--}0.85$) без значительных просадок и признаков стагнации, что критически важно для медицинских задач скрининга опухолей, где минимизация пропусков истинно положительных случаев (от англ. false negatives) имеет первостепенное значение и напрямую влияет на своевременность диагностики; такая динамика отражает высокую надежность базовой модели в захвате даже малоконтрастных или частично инфильтративных образований на фоне сложного анатомического контекста МРТ-срезов.

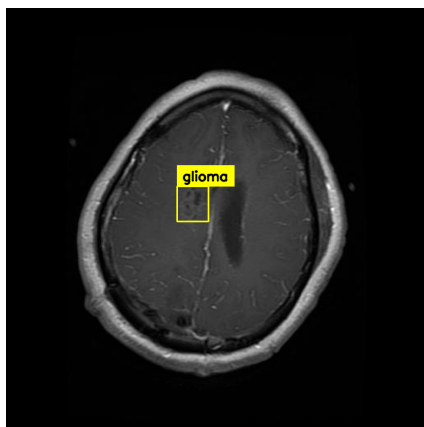
Напротив, версия 1.6.5 демонстрирует более поздний начальный скачок Average Recall (ближе к пятой эпохе), после чего кривая проявляет нестабильность с повторяющимися локальными минимумами (особенно после 20–30 эпох), повышенными амплитудами флуктуаций ($\sim 0.05\text{--}0.1$) и тенденцией к снижению к концу обучения; учитывая различия масштабов осей, где значения 1.6.5 визуальнo сжаты по сравнению с 1.3.0, эти колебания все равно указывают на переобучение регрессионной «головы» детектора под влиянием чрезмерно сложных аугментированных паттернов и повышенную чувствительность к сдвигу распределения данных, что в клинической практике чревато систематическими пропусками опухолей с атипичной морфологией или артефактами сканирования, подчеркивая необходимость тонкой настройки трансформерных детекторов для обеспечения робастной полноты обнаружения в медицинской визуализации.

Для иллюстрации качества работы моделей на конкретных клинических случаях на рисунке 4 представлены результаты детекции, полученные

с использованием RF-DETR v1.3.0.



a) успешная локализация крупной менингиомы



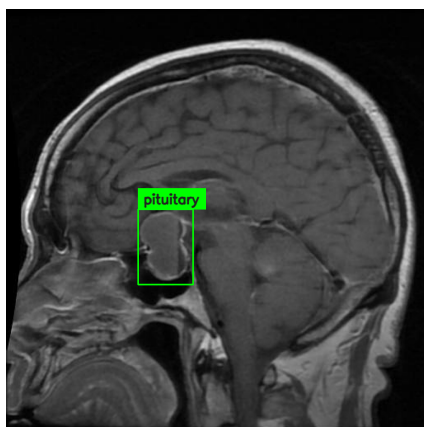
б) пропуск малой глиомы

Рисунок 4. Иллюстрация работы RF-DETR v1.3.0

На рисунке 5 показаны результаты детекции моделью Yolo26 с опти-



a) ложноположительное срабатывание на здоровой ткани



б) успешное нахождение опухоли гипофиза

Рисунок 5. Иллюстрация работы Yolo26 с оптимизатором MuSGD без аугментации яркости

мизатором MuSGD без аугментации яркости: а) ложноположительное срабатывание на здоровой ткани; б) успешное нахождение опухоли гипофиза. Зеленые ограничивающие прямоугольники обозначают верные детекции, красные – ложноположительные срабатывания, желтые рамки указывают пропущенные объекты. Отметим, что случаев ложноположительного срабатывания на здоровой ткани обученные нейросети не допускают вовсе. Пришлось снизить порог по уверенности в 10 раз, чтобы получить пример ложноположительного срабатывания.

3. Предложения по дальнейшему улучшению RF-DETR

Основываясь на проведенном анализе программного кода и научных работ из выполненного аналитического обзора, в том числе исследований по RF-DETR, Yolo26, HyMATOD, YOLOGX и др., предлагается доработать архитектуру RF-DETR, чтобы добиться лучших показателей качества, при этом дальнейшие модификации могут значительно усилить обобщаемость, скорость и адаптивность к специфическим доменам, таким как медицинская визуализация.

Для повышения точности на данных out-of-distribution рекомендуется интегрировать гибридные механизмы внимания, аналогичные используемым в HyMATOD [10], где показатель mAP вырос на 5–7% за счет пространственно-временной агрегации признаков в сценах, где множество объектов, текстур, фоновых элементов или перекрывающихся структур создают высокую плотность визуального шума, затрудняя точное выделение и локализацию целевых объектов.

Для задач медицинской визуализации, таких как детекция опухолей мозга на МРТ-изображениях, предлагается включить в промежуточный слой (neck) блок типа Gather-Distribution, который собирает и распределяет пространственные признаки на разных масштабах, дополнительно рекомендуется задействовать в «голове» блок SE-ResNeXt с механизмом Squeeze-and-Excitation для подчеркивания значимых канальных признаков. Эту гибридную схему стоит интегрировать с функцией потерь Focal-SIoU в «голове» детектора RF-DETR, которая позволит уделять больше внимания сложным ограничивающим прямоугольникам (например, при малых размерах или частичных перекрытиях) [11].

Заключение

Проведенное исследование показало, что современные трансформерные детекторы нового поколения способны обеспечивать высокую точность локализации внутричерепных объемных образований на МРТ-изображениях,

что подтверждает их высокую потенциальную клиническую ценность в условиях растущего потока нейровизуализационных данных и дефицита радиологов. Экспериментальное сравнение детекторов RF-DETR и Yolo26 на размеченном датасете MPT головного мозга продемонстрировало преимущество RF-DETR по интегральным метрикам качества (mAP и F1-меры при различных порогах IoU), а также более высокую скорость обработки (149 файлов MPT в секунду против 112 для Yolo26).

Прикладная ценность работы состоит в воспроизводимом сравнении двух архитектурных парадигм – трансформерной и сверточной – на стандартизированном медицинском датасете. Предложенный протокол оценки с оптимизацией порога уверенности по F1-мере ориентирован на клинический сценарий первичного скрининга и может быть использован для валидации других детекторов в медицинской визуализации; аналитическая сводка программных экосистем, представленная в обзорной части, дополняет количественное сравнение качественной оценкой инженерной доступности и лицензионных ограничений рассмотренных архитектур. Полученные количественные результаты позволяют обосновать выбор RF-DETR в качестве базовой модели для построения систем автоматизированной локализации опухолей головного мозга на МРТ-снимках.

Настоящее исследование имеет ряд ограничений, которые необходимо учитывать при интерпретации результатов.

Во-первых, экспериментальная валидация выполнена на единственном открытом датасете **Labeled MRI brain Tumor**, что не позволяет в полной мере оценить обобщающую способность моделей на данных из различных клиник, полученных на разных сканерах и с разными протоколами МРТ-съемки. Перспективным направлением является адаптация предложенного протокола оценки для публичных бенчмарков (BraTS, UPenn-GBM, TCGA).

Во-вторых, эксперимент полностью воспроизводим с использованием стандартных открытых библиотек: RF-DETR и Ultralytics YOLO, а также открытого датасета **Labeled MRI brain Tumor**; программная сводка в обзорной части систематизирует условия воспроизводимости для сторонних исследователей.

Авторский вклад сводится к тюнингу указанных моделей и подбору гиперпараметров, детально представленных в тексте статьи. Воспроизведение результатов не требует уникального программного кода и может быть выполнено стандартными скриптами указанных библиотек. В-третьих, задача детекции ограничивающих прямоугольников, будучи обоснованным первым этапом, не охватывает полной сложности природы опухолей;

дальнейшее развитие работы предполагает обучение с размытыми метками для учета неопределенности границ опухолевых образований.

Практическая значимость результатов состоит в том, что предложенный протокол обучения и оценки может быть использован как основа для построения воспроизводимой схемы внедрения систем автоматизированной детекции в клинический рабочий процесс с учетом доступных вычислительных ресурсов. Показанная эффективность RF-DETR открывает перспективы дальнейшей оптимизации и адаптации модели к другим задачам медицинской визуализации и к сценариям со сдвигом распределения данных. Перспективные направления развития включают интеграцию гибридных механизмов многократного внимания по типу HyMAT, использование модулей Gather-Distribution и SE-ResNeXt, а также внедрение расширенных функций потерь (например, Focal-SIoU) с последующей оценкой влияния таких модификаций на диагностическую точность и время работы врача-рентгенолога.

Поступила в редакцию 30.04.2026;
одобрена после рецензирования 19.05.2026;
принята к публикации 20.05.2026;
опубликована онлайн 26.06.2026.

Рекомендовал к публикации

к.ф.-м.н. А. Н. Виноградов

Информация об авторе:



Виталий Петрович Фраленко

Кандидат технических наук, ведущий научный сотрудник Исследовательского центра мультипроцессорных систем ИПС им. А.К. Айламазяна РАН. Область научных интересов: интеллектуальный анализ данных и распознавание образов, искусственный интеллект и принятие решений, параллельно-конвейерные вычисления, сетевая безопасность, диагностика сложных технических систем, графические интерфейсы.



0000-0003-0123-3773

e-mail: alarmod@pereslavl.ru

Декларация об отсутствии личной заинтересованности: *благополучие автора не зависит от результатов исследования.*

Список использованных источников

- [1] Robinson I., Robicheaux P., Popov M., Ramanan D., Peri N. *RF-DETR: Neural architecture search for real-time detection transformers*.– 2025.– 24 с. arXiv:2511.09554 [cs.CV] ↑^{265, 273}
- [2] Sapkota R., Cheppally R. H., Sharda A., Karkee M. *YOLO26: Key architectural enhancements and performance benchmarking for real-time object detection*.– 2025.– 15 с. arXiv:2509.25164 [cs.CV] ↑^{265, 274}
- [3] Chitphakdithai N., Chiang V. L., Duncan J. S. *Non-rigid registration of longitudinal brain tumor treatment MRI // Proceedings of the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2011 (Boston, MA, USA, 30 August 2011–03 September 2011)*.– IEEE.– 2011.– ISBN 978-1-4244-4121-1.– С. 4893–4896. ↑²⁶⁷
- [4] Drakopoulos F., Foteinos P., Liu Y., Chrisochoides N. P. *Toward a real time multi-tissue Adaptive Physics-Based Non-Rigid Registration framework for brain tumor resection // Frontiers in Neuroinformatics*.– 2014.– Т. 8.– 12 с. ↑²⁶⁷
- [5] Scheufe K., Mang A., Gholami A., Davatzikos C., Biros G., Mehl M. *Coupling brain-tumor biophysical models and diffeomorphic image registration // Computer Methods in Applied Mechanics and Engineering*.– 2019.– Т. 347.– С. 533–567. ↑²⁶⁸
- [6] Schaffer T., Brawanski A., Wein S., Tomé A. M., Lang E. W. *Brain tumor segmentation with special emphasis on the non-enhancing brain tumor compartment*.– 2026.– 31 с. arXiv:2602.21703 [cs.CV] ↑²⁶⁸
- [7] Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A. N., Kaiser L., Polosukhin I. *Attention is all you need*.– 2017.– 15 с. arXiv:1706.03762 [cs.CL] ↑²⁶⁹
- [8] Zhu X., Su W., Lu L., Li B., Wang X., Dai J. *Deformable DETR: Deformable transformers for end-to-end object detection // 9th International Conference on Learning Representations, ICLR 2021 (Virtual Event, Austria, May 3–7, 2021)*.– 2020.– ISBN 979-8-3313-0008-1.– С. 894–909. arXiv:2010.04159 [cs.CV] ↑²⁶⁹
- [9] Anasse B., Lahsen-Cherif I., El Khadimi A. *Deformable attention mechanisms applied to object detection, case of remote sensing // Procedia Computer Science*.– 2025.– Т. 270.– С. 5480–5489. ↑²⁷⁰
- [10] Moorthy S., Sachin Sakthi K. S., Arthanari S., Jeong J. H., Joo Y. H. *Hybrid multi-attention transformer for robust video object detection // Engineering Applications of Artificial Intelligence*.– 2025.– Т. 139, Part B.– ил. 109606. ↑^{270, 288}
- [11] Li C., Du Y., Zhang X., Wu P. *YOLOGX: an improved forest fire detection algorithm based on YOLOv8 // Frontiers in Environmental Science*.– 2025.– Т. 12.– ил. 1486212.– 15 с. ↑^{271, 288}
- [12] Lan J., Tao Y., Wang Zh., Yu H., Cui W. *Light-YOLOv8-Flame: A lightweight high-performance flame detection algorithm*.– 2025.– 12 с. arXiv:2504.08389 [cs.CV] ↑²⁷²



Application of RF-DETR and Yolo26 Detectors for Brain Tumor Localization Using MRI Data

Vitaly Petrovich **Fralenko**

Ailamazyan Program Systems Institute of RAS, Ves'kovo, Russia

Abstract. The study focuses on the application of artificial intelligence methods for detecting non-rigid objects in medical images, using the example of brain tumor detection from magnetic resonance imaging data.

The first part presents an analytical review of modern approaches to non-rigid registration and tumor segmentation, as well as transformer-based and hybrid object detector architectures with deformable attention, multi-channel feature aggregation, and specialized loss functions designed for processing deformable structures with blurred boundaries, as well as an analysis of software implementations and licensing constraints of the corresponding frameworks.

The second part presents an experimental comparison of the RF-DETR and Yolo26 neural networks on a labeled brain tumor dataset for bounding box detection. Notably, this task represents a standard procedure for localizing the region of interest and does not require special consideration of structural deformability or boundary uncertainty, unlike segmentation or non-rigid registration tasks. Quality assessment was performed using mAP0.50, mAP0.50–0.95, and F1-score metrics for IoU thresholds of 0.50 and 0.50–0.95 across confidence thresholds.

The obtained results showed that RF-DETR provides higher overall quality metrics and better speed compared to Yolo26, confirming that transformer-based detectors are a promising tool for automated analysis of brain MRI images. Based on analytical and experimental findings, recommendations were formulated for further development of RF-DETR in medical imaging tasks, aimed at improving the model's generalization ability, robustness to data distribution shifts, and clinical interpretability of results. (*In Russian*).

Key words and phrases: brain tumor, neural network, transformer, analytical review, object detection, RF-DETR, Yolo26, medical imaging

2020 Mathematics Subject Classification: 68-02; 68T45

Acknowledgments: The work was carried out within the framework of the State Budget Theme of the Ailamazyan Program Systems Institute of the Russian Academy of Sciences No. 125021302067-9 (completion deadline: 2025–2027)

For citation: Vitaly P. Fralenko. *Application of RF-DETR and Yolo26 Detectors for Brain Tumor Localization Using MRI Data*. Program Systems: Theory and Applications, 2026, 17:2(71), pp. 263–293. (*In Russ.*). https://psta.psriras.ru/read/psta2026_2_263-293.pdf

References

- [1] I. Robinson, P. Robicheaux, M. Popov, D. Ramanan, N. Peri. *RF-DETR: Neural architecture search for real-time detection transformers*, 2025, 24 pp.  [arXiv:2511.09554](#) [cs.CV]
- [2] R. Sapkota, R. H. Cheppally, A. Sharda, Karkee M.. *YOLO26: Key architectural enhancements and performance benchmarking for real-time object detection*, 2025, 15 pp.  [arXiv:2509.25164](#) [cs.CV]
- [3] N. Chitphakdithai, V. L. Chiang, J. S. Duncan. “Non-rigid registration of longitudinal brain tumor treatment MRI”, *Proceedings of the 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, EMBC 2011 (Boston, MA, USA, 30 August 2011–03 September 2011), IEEE, 2011, ISBN 978-1-4244-4121-1, pp. 4893–4896. 
- [4] F. Drakopoulos, P. Foteinos, Y. Liu, N. P. Chrisochoides. “Toward a real time multi-tissue Adaptive Physics-Based Non-Rigid Registration framework for brain tumor resection”, *Frontiers in Neuroinformatics*, **8** (2014), 12 pp. 
- [5] K. Scheufele, A. Mang, A. Gholami, C. Davatzikos, G. Biros, M. Mehl. “Coupling brain-tumor biophysical models and diffeomorphic image registration”, *Computer Methods in Applied Mechanics and Engineering*, **347** (2019), pp. 533–567. 
- [6] T. Schaffer, A. Brawanski, S. Wein, A. M. Tomé, E. W. Lang. *Brain tumor segmentation with special emphasis on the non-enhancing brain tumor compartment*, 2026, 31 pp.  [arXiv:2602.21703](#) [cs.CV]
- [7] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, I. Polosukhin. *Attention is all you need*, 2017, 15 pp.  [arXiv:1706.03762](#) [cs.CL]
- [8] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, Dai J.. “Deformable DETR: Deformable transformers for end-to-end object detection”, *9th International Conference on Learning Representations*, ICLR 2021 (Virtual Event, Austria, May 3–7, 2021), 2020, ISBN 979-8-3313-0008-1, pp. 894–909.  [arXiv:2010.04159](#) [cs.CV]
- [9] B. Anasse, I. Lahsen-Cherif, El Khadimi A.. “Deformable attention mechanisms applied to object detection, case of remote sensing”, *Procedia Computer Science*, **270** (2025), pp. 5480–5489. 
- [10] S. Moorthy, Sachin Sakthi K. S., S. Arthanari, J. H. Jeong, Y. H. Joo. “Hybrid multi-attention transformer for robust video object detection”, *Engineering Applications of Artificial Intelligence*, **139**, Part B (2025), id. 109606. 
- [11] C. Li, Y. Du, X. Zhang, Wu P.. “YOLOGX: an improved forest fire detection algorithm based on YOLOv8”, *Frontiers in Environmental Science*, **12** (2025), id. 1486212, 15 pp. 
- [12] J. Lan, Y. Tao, Zh. Wang, H. Yu, W. Cui. *Light-YOLOv8-Flame: A lightweight high-performance flame detection algorithm*, 2025, 12 pp.  [arXiv:2504.08389](#) [cs.CV]