

УДК 004.8

DOI: [10.26102/2310-6018/2026.54.3.012](https://doi.org/10.26102/2310-6018/2026.54.3.012)

Архитектуры глубокого обучения для сегментации мультифазных КТ-изображений

С.В. Самсоненко, И.Л. Каширина✉

МИРЭА – Российский технологический университет, Москва, Российская Федерация

Резюме. В статье проводится комплексный систематический анализ современных архитектур глубокого обучения для автоматической сегментации мультифазных КТ-изображений. Подробно рассматриваются специфические особенности мультифазных данных, главными из которых являются пространственные несовпадения (смещения) между фазами, вызванные движениями пациента, и различный характер накопления контрастного вещества в патологических тканях на разных фазах. Эти особенности делают прямую адаптацию классических методов сегментации неэффективной и требуют разработки специализированных архитектур. В статье прослеживается эволюция подходов: от базовых сверточных сетей (U-Net, 3D U-Net, nnU-Net) и гибридных моделей (TransUNet, UNETR), комбинирующих свертки и трансформеры, до специализированных решений. Особое внимание уделяется моделям с механизмами перекрестного внимания между фазами, таким как PA-ResSeg, M3Net и MULLET, которые позволяют осуществлять неявное выравнивание признаков и адаптивное слияние информации из разных фаз без явной регистрации (совмещения) изображений. В работе также анализируются сравнительные преимущества различных стратегий слияния данных с разных фаз (раннее, позднее, перекрестное взаимодействие), рассматриваются вопросы вычислительной эффективности и доступности открытых датасетов. Определены ключевые тенденции и перспективные направления развития области, включая применение фундаментальных моделей (MedSAM, VoxTell) и модально-агностичное обучение. Делается вывод о том, что дальнейший прогресс в области мультифазной сегментации КТ-изображений связан с созданием вычислительно эффективных архитектур, способных к интеграции в реальный клинический процесс для поддержки диагностических решений.

Ключевые слова: гибридные архитектуры, сегментация изображений, механизмы внимания, мультифазная КТ, слияние признаков, медицинская визуализация, глубокое обучение, компьютерное зрение, PA-ResSeg, M3Net.

Для цитирования: Самсоненко С.В., Каширина И.Л. Архитектуры глубокого обучения для сегментации мультифазных КТ-изображений. *Моделирование, оптимизация и информационные технологии*. 2026;14(3). URL: <https://moitvivr.ru/ru/journal/article?id=2211> DOI: 10.26102/2310-6018/2026.54.3.012

Deep learning architectures for multiphase CT image segmentation

S.V. Samsonenko, I.L. Kashirina✉

MIREA – Russian Technological University, Moscow, the Russian Federation

Abstract. The article provides a comprehensive systematic analysis of modern deep learning architectures for automatic segmentation of multiphase CT images. The specific features of multiphase data are considered in detail, the main of which are spatial mismatches (offsets) between phases caused by patient movements and the different nature of the accumulation of contrast agent in pathological tissues at different phases. These features make direct adaptation of classical segmentation methods ineffective and require the development of specialized architectures. The article traces the evolution of approaches: from basic convolutional networks (U-Net, 3D U-Net, nnU-Net) and hybrid models

(TransUNet, UNETR) combining convolutions and transformers to specialized solutions. Special attention is paid to models with mechanisms of cross-attention between phases, such as PA-ResSeg, M3Net and MULLET, which allow for implicit alignment of features and adaptive merging of information from different phases without explicit registration (alignment) of images. The paper also analyzes the comparative advantages of various data fusion strategies from different phases (early, late, cross-interaction), discusses issues of computational efficiency and availability of open datasets. Key trends and promising areas of development of the field have been identified, including the use of fundamental models (MedSAM, VoxTell) and modal-agnostic learning. It is concluded that further progress in the field of multiphase segmentation of CT images is associated with the creation of computationally efficient architectures capable of integration into the real clinical process to support diagnostic solutions.

Keywords: hybrid architectures, image segmentation, attention mechanisms, multiphase CT, feature fusion, medical imaging, deep learning, computer vision, PA-ResSeg, M3Net.

For citation: Samsonenko S.V., Kashirina I.L. Deep learning architectures for multiphase CT image segmentation. *Modeling, Optimization and Information Technology*. 2026;14(3). (In Russ.). URL: <https://moitvvt.ru/ru/journal/article?id=2211> DOI: 10.26102/2310-6018/2026.54.3.012

Введение

Мультифазная компьютерная томография (КТ) представляет собой серию трехмерных медицинских изображений одной анатомической области, полученных через определенные временные интервалы после введения контрастного вещества. Использование таких данных формирует задачу автоматической сегментации на основе многовременных снимков, где один и тот же объект имеет различные контрастные характеристики. В отличие от классической однофазной сегментации, мультифазные данные обладают информационной избыточностью, что теоретически позволяет более надежно выделять целевые структуры. Однако на практике этот подход сталкивается с серьезными вычислительными трудностями. К ним относятся пространственные смещения объектов между фазами, вызванные движениями пациента или деформациями органов, а также вариативность распределения контраста. Это обуславливает необходимость разработки специализированных алгоритмов для эффективного синтеза информации из нескольких источников [1].

Прямая адаптация существующих архитектур глубокого обучения, изначально разработанных для однофазных данных, оказывается малоэффективной из-за пространственной несогласованности изображений и ограниченной способности классических сверточных нейронных сетей (CNN) к моделированию глобальных зависимостей. Поскольку сверточные фильтры обрабатывают информацию локально в пределах ограниченного рецептивного поля, они не способны эффективно сопоставлять признаки объекта, смещенного между фазами.

Развитие архитектур на основе трансформеров и механизмов внимания (attention) открыло новые возможности для моделирования глобальных связей. Гибридные решения, такие как TransUNet, SegFormer и Swin-UNet, сочетающие сверточные слои с трансформерными энкодерами, уже продемонстрировали высокую эффективность при анализе КТ-изображений. Тем не менее, еще более значимые результаты показывают специализированные подходы, учитывающие специфику взаимодействия между фазами. Механизм внимания позволяет вычислять зависимости между всеми пространственными координатами одновременно, что помогает алгоритму находить соответствия между объектами без предварительной жесткой регистрации (совмещения) изображений. В последние годы были предложены методы, использующие перекрестное внимание (cross-attention) для адаптивного выравнивания и синтеза признаков: например, сеть PA-ResSeg с многоуровневым фазовым вниманием, M3Net с механизмом

нелокального межфазного взаимодействия и MULLET, применяющая трансформерную архитектуру для выделения признаков только в ключевых подобластях КТ-изображения.

Таким образом, актуальность разработки специализированных архитектур для мультифазной сегментации продиктована как теоретической сложностью обработки несогласованных многомерных данных, так и высокой практической значимостью автоматизации лучевой диагностики [2]. Целью данной работы является систематический анализ современных глубоких архитектур для сегментации мультифазных томографических изображений, сравнительная оценка их эффективности, а также обсуждение алгоритмических преимуществ различных механизмов внимания и определение перспективных направлений развития данной области.

Материалы и методы

Особенности мультифазных томографических данных. Мультифазная компьютерная томография предполагает последовательное сканирование анатомической области через определенные интервалы после внутривенного введения контрастного препарата. Стандартно выделяют четыре основные фазы: нативную (NC) – базовое сканирование, выполняемое до введения контрастного вещества; артериальную (ART) – примерно через 25–30 секунд после введения контраста; портально-венозную (PV) – через 60–70 секунд и отсроченную фазу (DL) – через 3–10 минут. Каждая фаза характеризуется специфическим распределением плотности контрастного вещества в тканях [3].

Необходимость мультифазного протокола обусловлена различной динамикой накопления и вымывания контраста в здоровых и патологических структурах. Использование полного набора фаз обеспечивает более полное и надежное описание анализируемого объекта (например, новообразования или другой патологии в онкологической диагностике). Исследования показывают, что совместный анализ артериальной и портально-венозной фаз значительно повышает показатели точности сегментации по сравнению с однофазным подходом, поскольку разные фазы могут выявлять разные аспекты морфологии и характеристик объекта [1]. При этом оптимальная информативность отдельных фаз зависит от типа и характеристик анализируемой структуры: одни типы структур лучше видны на артериальной фазе благодаря интенсивному накоплению контраста, другие – на портально-венозной, третьи – на отсроченной фазе. Следовательно, алгоритм сегментации должен уметь извлекать релевантную информацию из каждой фазы и синтезировать ее в единое согласованное представление об анализируемом объекте.

Для наглядной иллюстрации различий визуального представления очаговых образований печени на различных фазах контрастного усиления на Рисунке 1 приведен пример мультифазного КТ-исследования [3]. Представлены нативная (NC), артериальная (ART) и портально-венозная (PV) фазы четырех различных патологий: кист, фокальной нодулярной гиперплазии (FNH), гемангиом (HEM), гепатоцеллюлярной карциномы (HCC) и метастазов. Видно, что интенсивность контрастирования и визуальные границы очагов существенно различаются между фазами, что обуславливает необходимость их совместного анализа при автоматической сегментации.

Однако использование мультифазных данных сопряжено с проблемой пространственного несоответствия изображений. Несмотря на техническую синхронизацию при сканировании, положение структур в разных фазах может варьироваться на несколько миллиметров и более. Данное смещение вызвано двумя ключевыми факторами: физическим движением пациента между циклами сканирования и деформацией внутренних органов вследствие дыхания. Такие пространственные

несоответствия создают основную сложность при обработке мультифазных данных алгоритмами автоматической сегментации.

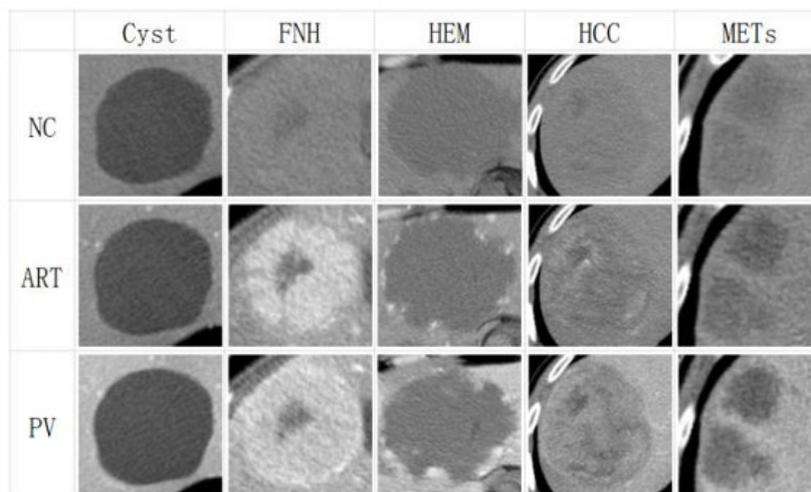


Рисунок 1 – Визуализация очаговых образований печени на различных фазах контрастного усиления КТ

Figure 1 – Imaging of focal liver lesions on different phases of contrast-enhanced CT

Проблема межфазных смещений напрямую определяет выбор архитектурных решений сегментационных моделей. Классические подходы, основанные на конкатенации изображений разных фаз как отдельных каналов многоканального изображения, предполагают идеальное пиксельное совпадение позиций объекта на разных фазах. В реальности такого совпадения нет, что приводит к артефактам в обучении модели: сверточные фильтры могут начать реагировать на граничные эффекты смещения и несоответствия вместо фактических признаков структуры. Это снижает обобщающую способность модели и ее устойчивость к вариациям входных данных.

Кроме того, границы и видимая форма объектов могут изменяться между фазами из-за неравномерного распределения контраста. Структура, четко очерченная в одной фазе, может иметь размытые контуры или сливаться с окружающими тканями в другой. Алгоритм сегментации должен уметь сопоставлять эти вариативные представления, идентифицируя их как части единого объекта.

Таким образом, мультифазная томографическая визуализация порождает задачу, которая требует не просто объединения информации из нескольких источников, но и решения проблемы пространственной и деформационной несогласованности между этими источниками при одновременном сохранении преимуществ, даваемых временной динамикой накопления контраста. Алгоритмы сегментации должны либо явно компенсировать эти несоответствия через предварительную регистрацию (выравнивание) фаз, либо неявно выполнять выравнивание на уровне признаков через механизмы, которые умеют находить соответствия между областями разных фаз несмотря на смещения и деформации. Выбор между этими стратегиями и различные реализации этих подходов определяют архитектурные особенности современных нейросетевых моделей, рассматриваемых в данной работе.

Обзор архитектур глубоких нейросетей для сегментации

Сверточные нейросети. Классической архитектурой для сегментации биомедицинских изображений является U-образный автокодировщик U-Net [4]. Концепция этой сети строится на сочетании сверточного энкодера для извлечения признаков и симметричного декодера для восстановления пространственного

разрешения. Эти два блока связаны между собой пропускными соединениями (skip-connections), которые работают по принципу конкатенации: признаки из энкодера объединяются с признаками декодера в виде дополнительных каналов тензора. Это позволяет сети сохранять детализированную пространственную информацию, которая обычно теряется при последовательном сжатии данных в глубоких слоях. Однако оригинальная модель U-Net была ориентирована на обработку двумерных срезов, что привело к разработке 3D U-Net [5], где классические операции были заменены на пространственные свертки для анализа всего объема КТ-исследования целиком.

Альтернативный подход к проектированию 3D сегментаторов был реализован в архитектуре V-Net [6], где вместо конкатенации признаков стали применяться остаточные соединения (residual connections). В отличие от классических пропускных связей U-Net, здесь используется механизм поэлементного суммирования входа и выхода сверточного блока. С математической точки зрения это заставляет слои обучаться вычислению лишь разности между текущим и целевым представлением, что предотвращает затухание градиентов и позволяет эффективно обучать более глубокие нейросети.

На основе синтеза этих подходов был создан фреймворк nnU-Net [7, 8], ставший современным стандартом в медицинской сегментации. Данная система автоматически адаптирует параметры сети – от размера патчей до конфигурации слоев – под специфику конкретного набора данных, избавляя исследователя от необходимости ручной настройки архитектуры.

Тем не менее, даже наиболее совершенные сверточные сети сохраняют ограничение, связанное с локальностью извлекаемых признаков. Из-за фиксированного размера рецептивного поля фильтров им сложно устанавливать глобальные пространственные взаимосвязи между удаленными областями трехмерного объема. Эта проблема становится особенно критичной при анализе мультифазных данных, где необходимо сопоставлять объекты со значительным смещением.

Трансформеры для сегментации. В 2020–2021 гг. в компьютерном зрении произошел качественный сдвиг, связанный с адаптацией архитектуры Vision Transformer (ViT) из области обработки естественного языка для анализа изображений [9]. Трансформеры привлекли внимание способностью учитывать глобальный контекст посредством механизма многоголового внимания. В отличие от локальной природы сверточных операций, внимание позволяет каждому элементу изображения взаимодействовать с элементами на произвольном расстоянии, что обеспечивает способность к глобальному моделированию зависимостей.

Одним из наиболее характерных примеров реализации полностью трансформерного подхода стала модель SETR (SEgmentation TRansformer), в которой изображение разбивается на последовательность небольших фрагментов (патчей) [10]. Эти патчи обрабатываются энкодером на базе трансформера, после чего декодер восстанавливает карту сегментации до исходного разрешения. Тем не менее, практика показала, что полностью трансформерные архитектуры часто сталкиваются с проблемой потери мелких деталей. Из-за последовательной обработки патчей разрешение признаков снижается. Без использования специализированных механизмов восстановления высокого разрешения это приводит к размытию контуров объектов, что недопустимо при высокоточном анализе медицинских снимков.

Гибридные архитектуры. Более практичным подходом оказалось комбинирование сверточных сетей и трансформеров в гибридные архитектуры. В таких моделях трансформер используется для захвата глобальных зависимостей в области высокоуровневых признаков, а сверточный декодер восстанавливает точную форму объектов на основе локальной информации.

Одной из первых гибридных сетей является TransUNet (2021), архитектура которой схематично показана на Рисунке 2 [11]. Входное изображение при обработке данной сетью сначала проходит через CNN-энкодер (на основе сети ResNet), который извлекает иерархические признаки на нескольких уровнях разрешения; признаки самого глубокого уровня линейно проецируются в последовательность патчей и подаются на вход трансформеру, где механизм внимания моделирует зависимости между всеми областями изображения, захватывая глобальный анатомический контекст. Затем выход трансформера преобразуется обратно в пространственное представление и поступает в декодер, построенный по принципу U-Net: на каждом этапе повышения разрешения признаки трансформера объединяются с соответствующими skip-соединениями от энкодера, что позволяет совместить глобальное понимание сцены с локальной точностью и восстановить детализированную маску сегментации на исходном разрешении. Такой синтез локальных и глобальных признаков особенно эффективен в медицинской визуализации, где важно одновременно учитывать как мелкие структуры (например, границы опухоли), так и их пространственное расположение относительно других анатомических объектов.

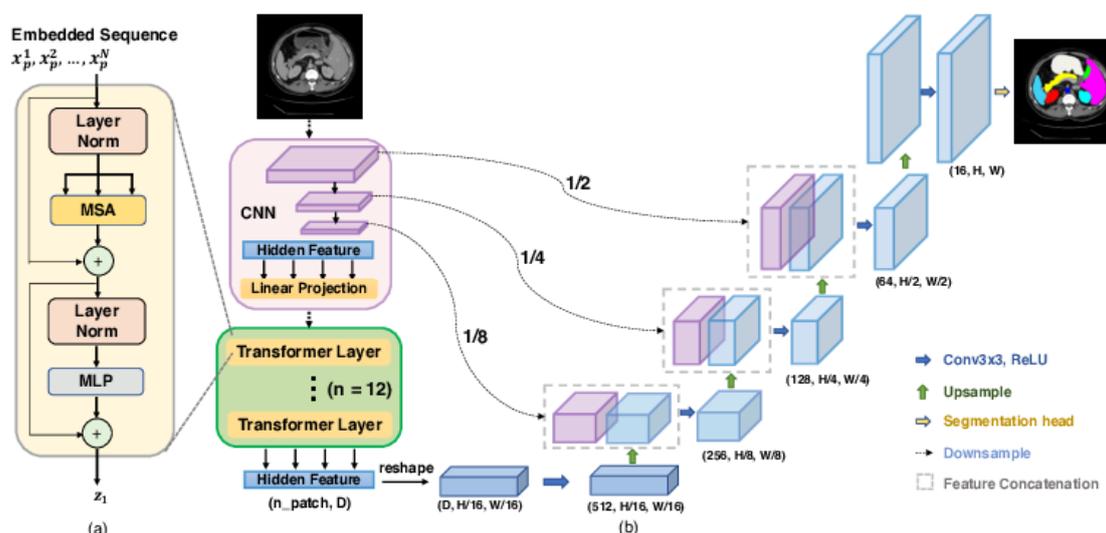


Рисунок 2 – Упрощенная схема гибридной архитектуры TransUNet
 Figure 2 – Simplified scheme of TransUNet hybrid architecture

Похожий принцип реализован в архитектуре UNETR (2021), где роль энкодера полностью берет на себя трансформер, работающий с патчами исходного изображения, но декодер остается сверточным, как в U-Net [12]. Это подчеркивает ключевую идею: глобальные зависимости удобнее моделировать средствами внимания, а восстановление точных карт сегментации – задача, для которой классический сверточный декодер по-прежнему более эффективен. Модификации типа Swin UNETR используют более экономичные варианты внимания (оконное, иерархическое), что делает модель UNETR масштабируемой на большие изображения и трехмерные данные без существенного роста вычислительных затрат [13].

К гибриднему семейству относятся и архитектуры, в которых блоки внимания встраиваются как «надстройка» к привычным CNN. Примером являются Attention U-Net и близкие подходы: здесь в пропускные соединения или внутрь сверточных блоков добавляются механизмы канального и пространственного внимания, которые учатся выделять важные области на изображениях и подавлять фон [14, 15]. Такие модели, по

сути, остаются сверточными, но приобретают способность переориентировать внимание на информативные участки изображения, что улучшает сегментацию.

Гибридные архитектуры представляют собой разумный компромисс [16]. Они снимают главное ограничение чисто сверточных сетей (локальность восприятия), но не требуют столь же больших объемов данных и вычислительных ресурсов, как полностью трансформерные модели.

Фундаментальные архитектуры. Параллельно с развитием специализированных CNN- и гибридных архитектур в медицине формируется направление фундаментальных (foundation) моделей для сегментации. В отличие от «традиционных» сетей, которые обучаются на одном конкретном датасете под одну задачу (например, сегментация конкретного органа на КТ), фундаментальный подход предполагает предварительное обучение одной крупной модели на большом и разнородном корпусе медицинских изображений разных модальностей (КТ, МРТ, ПЭТ, рентген, УЗИ) и различных анатомических областей. Такая модель выступает в роли универсального сегментатора, который затем можно адаптировать к конкретным задачам с минимальным объемом дообучения или даже использовать в режиме zero-shot (без обучения).

Примером таких архитектур является MedSAM (Medical Segment Anything Model) [17]. MedSAM является специализированной версией модели Segment Anything, обученной на более чем 1,5 млн пар изображение-маска из медицинских датасетов разных модальностей и органов. Благодаря такому охвату модель приобрела способность сегментировать новые изображения без дообучения, зачастую не хуже узкоспециализированных моделей. В MedSAM реализован механизм интерактивного сегментирования: алгоритм принимает на вход изображение и некий prompt – например, указание точки внутри интересующей области или ее грубую обводку – и мгновенно выдает соответствующую маску сегментации.

Еще более продвинутый класс представляют собой визуально-языковые модели для сегментации, которые способны воспринимать текстовые описания объекта. В 2025 г. была представлена модель VoxTell – универсальный сегментатор, принимающий на вход объемные медицинские снимки и произвольный текстовый запрос (от названия органа до целой клинической фразы). VoxTell обучен на 62 тысяч 3D-сканов (КТ, МРТ, ПЭТ) с более чем 1000 вариантов анатомических структур и патологий, и благодаря многоуровневому слиянию зрительных и текстовых признаков способен выделять на изображении структуру, описанную в запросе [18]. Он демонстрирует передовое качество в режиме zero-shot на новых датасетах. При этом у фундаментальных моделей есть ограничения, как и у всех других рассмотренных подходов.

Существующие подходы к мультифазной сегментации КТ-снимков. Мультифазные КТ-данные можно обработать двумя принципиально разными способами: предварительно совместить (зарегистрировать) изображения на разных фазах между собой или пытаться сегментировать напрямую без явной регистрации, полагаясь на алгоритм для совмещения информации. Рассмотрим подробнее оба варианта.

Сегментация с регистрацией фаз. Один из подходов к обработке мультифазных КТ-данных заключается в явном пространственном выравнивании (регистрации) изображений разных фаз перед их совместной сегментацией. Регистрация в этом контексте означает вычисление преобразования (аффинного или деформируемого), которое приводит все фазы к общей системе координат, уменьшая межфазные смещения и деформации [19].

Последовательная схема «регистрация→сегментация» реализует это в два шага. Сначала применяется алгоритм регистрации (классический или нейросетевой), который совмещает изображения по выбранному критерию сходства. Затем выровненные фазы

рассматриваются как согласованный многоканальный вход для сегментационной сети. Преимущество такого подхода в том, что сегментация выполняется по уже приведенным к одному положению данным, что частично устраняет артефакты от межфазовых смещений. Основные недостатки – накопление ошибок и разобценная оптимизация: искажения на этапе регистрации напрямую передаются на сегментацию, а сама регистрация не «знает» о том, какие структуры являются целевыми.

Совместные модели регистрации и сегментации решают эти задачи в единой архитектуре. Такой подход реализован в генеративно-сегментационной модели JSSR (Joint Synthesis, Segmentation and Registration) [20]. В ней три компонента обучаются совместно: генератор синтезирует изображение одной фазы в стиле другой, регистратор оценивает поле деформации для их совмещения, а сегментатор предсказывает маски на обоих изображениях. Обучение организовано без явной разметки смещений: используются косвенные критерии – схожесть реального и синтезированного изображений после деформации и согласованность сегментированных областей.

Схематически процесс можно описать так. Изображение фазы x подается в генератор Syn , который формирует псевдоизображение y_{fake} в стиле фазы y . Регистратор Reg вычисляет поле деформации τ , совмещающее y_{fake} с реальным y . Сегментатор Seg предсказывает маски на x , y_{fake} и y . Общая функция потерь включает: абсолютную или квадратичную метрику различия между y и деформированным y_{fake} ; штраф за негладкость поля τ ; сегментационную компоненту (например, Dice-потери), требующую совпадения масок после применения τ . Благодаря этому регистратор «подстраивается» не под произвольное совпадение интенсивностей, а под совмещение именно тех областей, которые важны для сегментации, что приводит к приросту качества по обеим задачам по сравнению с отдельной оптимизацией. На Рисунке 3 приведена схема такой системы.

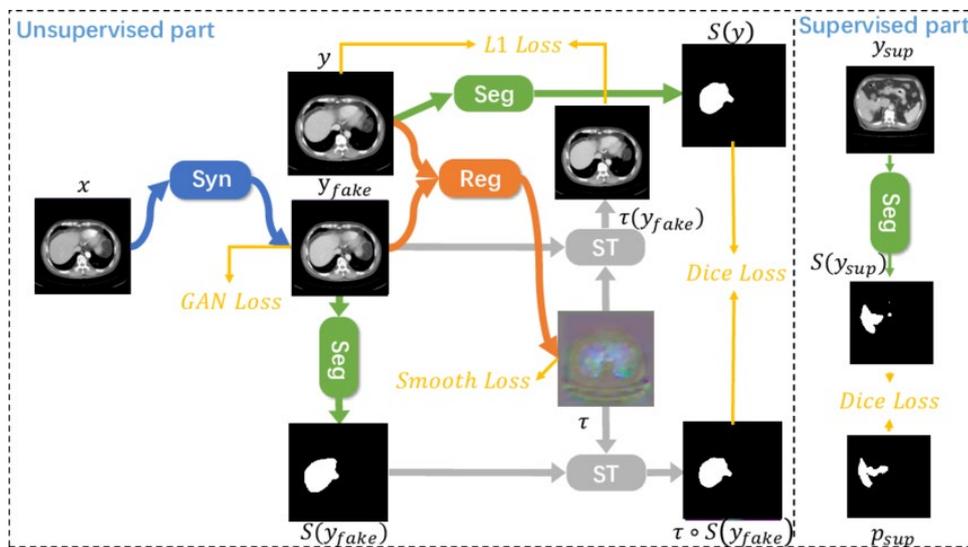


Рисунок 3 – Архитектура JSSR для одновременной регистрации и сегментации
 Figure 3 – JSSR architecture for simultaneous registration and segmentation

JSSR превосходит алгоритмы с последовательной схемой «регистрация → сегментация» по метрикам перекрытия областей на большом наборе патологических КТ с четырьмя фазами контрастирования. Тем не менее, такие комплексные модели сложны в реализации и требуют больших объемов данных для надежного обучения.

Сегментация без явной регистрации. Альтернативный подход к мультифазной КТ-сегментации основан на том, что фазы не выравниваются явно на уровне

изображений. Вместо этого архитектура сети строится так, чтобы сама модель училась сопоставлять и объединять информацию между фазами на уровне признаков. Это избавляет от отдельного этапа регистрации и снижает риск накопления ошибок, но требует более сложных сетевых механизмов. В литературе можно выделить три основных стратегии.

1) Раннее слияние (early fusion). Самый простой вариант – рассматривать мультифазное исследование как многоканальное изображение: отдельные фазы трактуются как разные каналы и одновременно подаются в сегментационную сеть (например, 3D U-Net или nnU-Net). Сеть с первых слоев «видит» совокупность фаз и может извлекать смешанные признаки. Такой подход минимально усложняет архитектуру и нередко дает выигрыш по сравнению с однофазной сегментацией, особенно при крупных объектах и небольших межфазовых смещениях. Основное ограничение – чувствительность к несовпадению структур: если объект заметно сдвинут между фазами, один и тот же сверточный фильтр видит в разных каналах несоответствующие фрагменты, что приводит к артефактам и ложным срабатываниям. Поэтому эффективность раннего слияния сильно зависит от качества исходной синхронизации сканов и стабильности протокола.

2) Позднее слияние (late fusion). Здесь каждая фаза обрабатывается своей подсетью, а объединение происходит уже на уровне полученных карт вероятностей или бинарных масок. Маски можно комбинировать логическими операциями (объединение, пересечение), усреднением или с помощью небольшой дополнительной сети, которая по нескольким картам принимает решение о финальной разметке. Преимущество такого подхода в том, что каждая ветвь подстраивается под свои особенности контрастирования: одна фаза может лучше «подсвечивать» границы, другая – внутреннюю структуру очага или органа. Недостаток в том, что совместное использование информации оказывается ограниченным: как правило, объединяются уже готовые решения, а не исходные признаки. Грубое объединение масок дает либо завышенную чувствительность с большим числом ложных областей, либо, наоборот, потерю мелких структур при слишком строгом пересечении.

3) Перекрестное взаимодействие признаков (cross-phase feature fusion). Данное направление связано с архитектурами, в которых взаимодействие фаз встроено непосредственно в структуру сети и реализуется на уровне признаков. В ранних работах такого типа была предложена архитектура Hyper-Pairing Network (HPN), представляющая собой трехмерный сегментатор по типу U-Net, в котором на каждом уровне энкодера карты признаков разных фаз попиксельно объединялись и совместно обрабатывались последующими сверточными слоями [21]. Это позволило явно учитывать различия контрастирования одного и того же анатомического участка на разных фазах и дало заметный прирост метрик качества сегментации по сравнению с использованием только одной фазы. При этом вклад каждой фазы задавался неявно параметрами сверток, без отдельного механизма управления важностью того или иного фазового компонента.

Следующий шаг развития данного подхода связан с введением фазового внимания. В архитектурах класса PA-ResSeg (Рисунок 4) используется двуветвевой энкодер по числу фаз, внутри которого на нескольких масштабах размещаются блоки, выполняющие как внутрифазное, так и межфазное перераспределение внимания. Внутрифазный компонент таких блоков анализирует каналы признаков в пределах одной фазы и усиливает наиболее информативные, подавляя фоновые и шумовые [3]. Межфазный компонент сопоставляет признаки различных фаз, находит соответствующие участки и передает релевантную информацию от одной ветви к другой. В результате представление объекта на каждой фазе постоянно уточняется с

учетом того, как он выглядит на других фазах, а слияние фазовой информации происходит не однократно, а на нескольких уровнях – от грубых низкочастотных карт до детальных границ. Дополнительные функции потерь, усиливающие вклад областей границ, позволяют еще более улучшить локализацию контуров. Архитектуры этого типа демонстрируют устойчивый прирост метрик качества по сравнению как с однофазными моделями, так и с простыми схемами раннего или позднего объединения фаз.

Схожие принципы реализуются и в моделях с нелокальным межфазовым взаимодействием, к числу которых относится архитектура M3Net [22]. В ней двуветвевой энкодер дополняется системой перекрестных связей, в которых используются специальные блоки для сопоставления пространственных позиций и каналов признаков между фазами. Пространственные блоки строят соответствия между удаленными областями трехмерного КТ-снимка, позволяя сопоставить одно и то же анатомическое место даже при смещении на несколько вокселей, а каналные блоки анализируют взаимосвязи между различными фазами и усиливают наиболее информативные их комбинации. Такой механизм можно рассматривать как мягкую регистрацию на уровне признаков: карты признаков фаз выравниваются по наиболее схожим участкам, за счет чего влияние геометрической несинхронности между фазами заметно снижается. Эксперименты на внутренних и внешних наборах данных подтверждают, что подобная организация межфазового взаимодействия обеспечивает высокую точность и обобщающую способность.

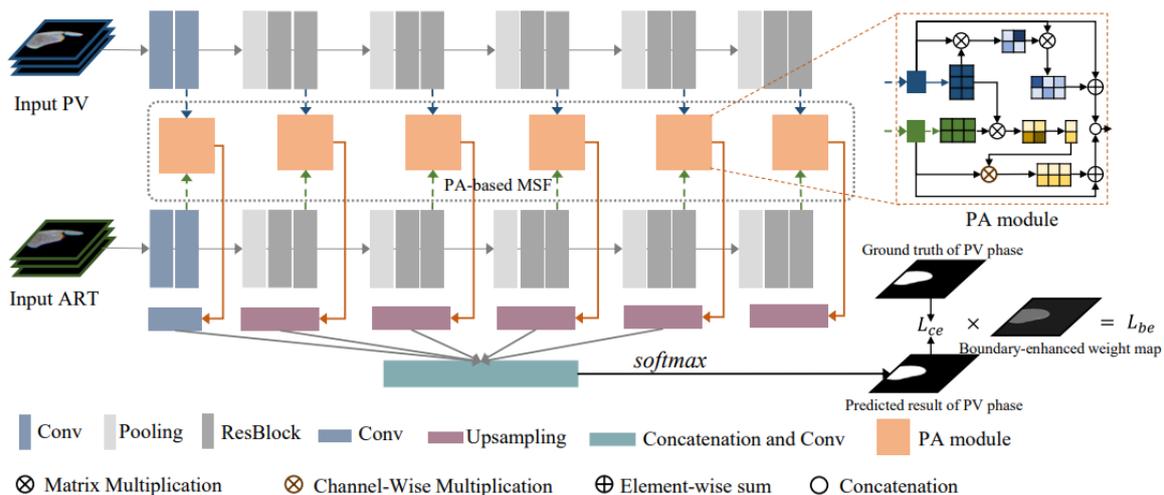


Рисунок 4 – Архитектура нейросетевой модели PA-ResSeg для сегментации опухолей печени на мультифазных КТ-изображениях

Figure 4 – Architecture of the PA-ResSeg neural network model for segmentation of liver tumors on multiphase CT images

Архитектура MULLET (Рисунок 5) представляет собой многоуровневую систему, спроектированную для явного учета взаимодействия между мультифазными контекстами [1]. Конвейер обработки состоит из четырех ключевых компонентов. На первом этапе каждая фаза независимо обрабатывается сверточным энкодером типа ResNet. Энкодер извлекает иерархические карты признаков, фиксируя как локальные особенности текстуры ткани, так и более высокоуровневые структурные характеристики анатомии. На этом шаге фазы обрабатываются параллельно, что позволяет каждой сети подстроиться под характер своего контрастирования. На втором этапе для каждой фазы выделяются области интереса (ROI), содержащие целевую структуру (например, опухоль). Эти области кодируются с помощью механизма самовнимания, который для

каждой ROI из одной фазы выбирает наиболее информативные пространственные позиции и признаковые каналы. Глобальное усреднение затем превращает каждую закодированную ROI в компактное векторное представление. На выходе этого этапа получаются наборы признаков, соответствующие выделенным регионам обеих фаз. Третий, ключевой этап, использует трансформерный блок, в котором благодаря механизму внимания строятся связи между областями интереса разных фаз. На этапе исследования мультифазного контекста используется два типа внимания. Внутрифазное внимание строит отношения между ROI одной и той же фазы, позволяя модели по-разному весить вклад разных частей одной фазы. Межфазное внимание явно сопоставляет области интереса из разных фаз, находит их соответствие несмотря на возможные смещения и передает информацию между фазами. В результате для каждой ROI формируется уточненное представление, которое учитывает как внутреннюю структуру самой ROI, так и соответствующие ей области на других фазах. Трансформер работает с уже упрощенным представлением (ROI в виде векторов), что делает вычисления эффективными и позволяет моделировать сложные долгосрочные зависимости между фазами. На выходе этого блока получается объединенное мультифазное представление для каждой области интереса. На заключительном этапе объединенные признаки подвергаются процессу восстановления пространственного разрешения и дополнительной обработке сверточными слоями. При этом происходит преобразование объединенной информации обратно в пространственное представление, соответствующее исходному разрешению среза. На завершающем этапе каждый воксель (трехмерный пиксель) классифицируется как принадлежащий целевой структуре или фону, что дает финальную маску сегментации.

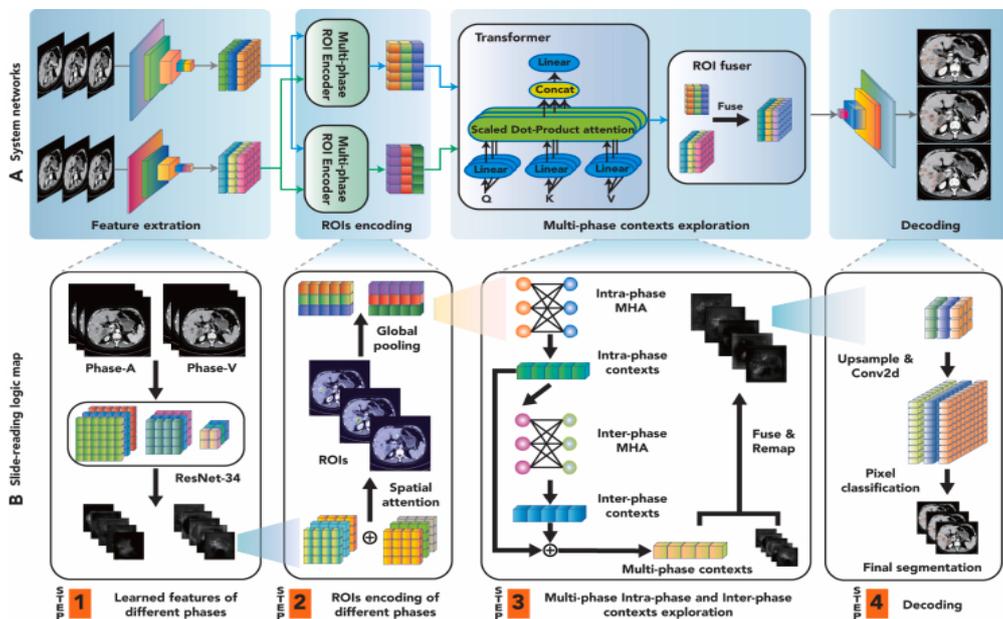


Рисунок 5 – Нейросетевая система для сегментации очагов печени на мультифазной КТ
 Figure 5 – Neural network system for liver lesion segmentation on multiphase CT

Модально-агностичное обучение. Это отдельное направление развития мультифазных моделей, при котором архитектура изначально не привязывается к конкретному набору фаз. В этом подходе в ходе обучения модель последовательно видит различные комбинации доступных фаз: полный набор, отдельные фазы, их случайные подмножества. При каждой конфигурации входных данных от сети требуется обеспечивать одинаково высокое качество сегментации целевых структур [23]. Такой

режим обучения заставляет алгоритм сосредоточиться на отличительных признаках анатомии и патологии, которые проявляются независимо от конкретного типа или набора контрастирования, и снижает его зависимость от частного фазового протокола. В результате сеть формирует усредненное представление органа и его поражений, которое остается стабильным при вариациях исходных данных. Такая стратегия оказывается особенно полезной в ситуациях, когда один из этапов контрастирования отсутствует вследствие технических причин, когда качество отдельной фазы снижено или когда исследования проводятся по разным протоколам. В этих случаях «гибкая» модель способна достичь приемлемого качества, опираясь лишь на имеющуюся информацию, без существенного падения показателей.

Открытые датасеты для мультифазной сегментации. Развитие архитектур с перекрестным вниманием и гибридных моделей для мультифазной сегментации напрямую зависит от доступности качественных обучающих данных. Однако в отличие от однофазных задач, где существуют обширные общедоступные коллекции (например, Medical Segmentation Decathlon [24]), открытые мультифазные датасеты КТ с разметкой патологических очагов являются крайне ограниченными. Среди них выделяется MCT-LTDiag (Multi-phase CT Dataset for Liver Tumor Diagnosis) – крупнейший на сегодня открытый набор данных для мультифазной сегментации печени [25]. Датасет состоит из 517 пациентов с полным набором четырех фаз контрастирования, каждая из которых представляет трехмерный объем печени в полном разрешении. Очаги поражения размечены экспертами-радиологами на всех фазах, что позволяет проводить обучение и валидацию алгоритмов, явно учитывающих мультифазную информацию. Датасет включает 4 типа опухолей и размещен на платформе Harvard Dataverse.

WAW-TACE представляет собой открытый датасет мультифазных КТ-исследований 233 пациентов с 377 размеченными опухолевыми масками [26]. Отличительной особенностью этого датасета является включение автоматизированных сегментаций 104 анатомических структур для каждой фазы, что позволяет исследовать не только целевую структуру, но и роль контекстной информации в алгоритмах сегментации. Данные размещены на платформе Zenodo.

TriALS (Triphasic-aided Liver Lesion Segmentation) представляет собой открытый бенчмарк, также содержащий полные четырехфазные КТ-исследования печени [27]. Датасет разработан с целью валидации методов, способных эффективно использовать мультифазную информацию при наличии вариаций в доступных фазах и протоколах сканирования. В отличие от статических датасетов, этот ресурс постоянно пополняется данными из разнородных клинических центров.

Несмотря на появление этих специализированных коллекций, их совокупный объем все еще не сопоставим с масштабами данных в классическом компьютерном зрении. Текущего количества размещенных наборов недостаточно для полноценного обучения тяжелых трансформерных архитектур. Это диктует необходимость поиска более эффективных методов переноса знаний или синтеза признаков, способных работать в условиях ограниченных выборок.

Метрики качества сегментации КТ-изображений. Качество сегментации оценивается метриками, сравнивающими предсказанную маску с экспертной разметкой. Основу составляют показатели, основанные на величинах: TP (True Positives) – правильно выделенные пиксели патологии; FP (False Positives) – ложные срабатывания; FN (False Negatives) – пропущенные участки патологии. Ключевой метрикой является коэффициент Дайса:

$$DSC = \frac{2|P \cap G|}{|P| + |G|} = \frac{2TP}{2TP + FP + FN}, \quad (1)$$

где P – предсказанная маска, G – эталон. Dice интерпретируется как мера схожести (1,0 – идеальное совпадение). Близкой по смыслу является метрика IoU (Intersection over Union):

$$IoU = \frac{|P \cap G|}{|P \cup G|} = \frac{TP}{TP + FP + FN}, \quad (2)$$

при этом:

$$DSC = \frac{2 \cdot IoU}{1 + IoU}. \quad (3)$$

Dice несколько выше численно (например, $IoU = 0,80$ соответствует $DSC = 0,89$). В медицинских публикациях чаще приводят Dice как более привычный показатель сегментации, что обусловлено его более высокой численной чувствительностью к изменениям в точности выделения границ.

Несмотря на широкое применение, коэффициенты Дайса и IoU имеют существенный недостаток: они оценивают лишь площадь или объем пересечения, но игнорируют пространственное положение ошибок. Для модели не имеет значения, находятся ли ошибочно классифицированные пиксели рядом с границей объекта или на значительном удалении от него, хотя с клинической точки зрения это принципиально разные погрешности. Данная проблема становится критической при сегментации мелких или сложной формы очагов, где даже небольшое смещение контура может привести к серьезным диагностическим искажениям при сохранении высокого значения Dice. Для компенсации этого ограничения используются дистанционные метрики, наиболее распространенной из которых является расстояние Хаусдорфа (HD). В отличие от объемных показателей, она измеряет фактическое физическое расстояние между поверхностями предсказанной и эталонной масок. Математически она представляет собой максимальное из всех кратчайших расстояний между точками двух контуров, указывая на «худший случай» отклонения границы.

Результаты и обсуждение

Анализ современных подходов к мультифазной сегментации показывает, что эффективная интеграция информации из нескольких фаз контрастирования требует специализированных архитектурных решений, выходящих за рамки простого конкатенирования входных каналов. В Таблице 1 приведен сравнительный анализ архитектур глубоких нейросетей для сегментации.

Ключевым прорывом стали механизмы перекрестного внимания между фазами, реализованные в архитектурах вроде PA-ResSeg, M3Net и MULLET. Эти модели решают две фундаментальные задачи: адаптивное выравнивание признаков при межфазных смещениях и селективное усиление информативных контрастных паттернов без отдельного этапа жесткой регистрации. Преимущество механизмов внимания заключается в их способности к неявной регистрации на уровне признаков. В отличие от классических сверточных сетей, где фильтр обрабатывает фиксированное окно пикселей, механизм внимания вычисляет веса соответствия между произвольными позициями разных фаз, компенсируя смещения до 5–7 мм, вызванные движениями пациента или дыхательными артефактами. Эксперименты, приведенные в литературе, демонстрируют, что архитектуры с межфазным вниманием превосходят базовые схемы раннего и позднего слияния на 8–15 % по метрике Dice, достигая значений 0,85–0,92 на задачах сегментации органов брюшной полости.

Важной особенностью современных решений является многоуровневая организация взаимодействия фаз. Гибридные архитектуры реализуют обмен признаками

на нескольких масштабах энкодера: глобальное позиционирование объекта формируется на глубоких уровнях через пространственное внимание, а уточнение границ – на промежуточных уровнях через каналы внимания. Такая иерархия позволяет модели одновременно учитывать анатомический контекст и локальные текстурные особенности, повышая робастность к вариациям протоколов сканирования.

Таблица 1 – Сравнение архитектур глубоких нейросетей для сегментации изображений
Table 1 – Comparison of deep neural network architectures for image segmentation

Класс архитектур + типичные представители	Характерные особенности реализации	Требовательность к данным и ресурсам	Пригодность для 3D и мультифазной КТ
Сверточные: U-Net, V-Net, 3D U-Net, nnU-Net	2D/3D-свертки, skip-соединения, локальное рецептивное поле; хорошо отлаженные реализации.	Относительно невысокие; ресурсы растут примерно линейно с размером патча/объема.	Естественная поддержка 3D (3D-свертки, объемные патчи). Для мультифазной КТ требуют явных схем слияния фаз, плохо компенсируют межфазовые смещения без специальных модулей.
Трансформерные: SETR	Энкодер полностью на механизме внимания по патчам; глобальное поле зрения уже на ранних слоях	Квадратичный рост сложности внимания по числу патчей; высокая потребность в объеме данных и памяти.	3D-расширения возможны, но еще более ресурсоемки. Без специальных блоков не учитывают структуру фаз; на практике почти всегда требуют гибридизации или межфазного внимания.
Гибридные: TransUNet, UNETR, Swin-UNETR, Swin-UNet, SegFormer, Attention U-Net, PA-ResSeg	Сочетание сверточного декодера с трансформерными или attention-блоками на глубоких уровнях; локальные признаки извлекаются CNN, глобальный контекст моделируется через внимание	Требования выше, чем у чистых CNN, но ниже, чем у полностью трансформерных моделей; хорошо масштабируются за счёт иерархии признаков.	Хорошо подходят для 3D; позволяют естественно внедрять межфазное взаимодействие (cross-phase, phase-attention). На практике – наиболее удобный класс для мультифазной КТ (PA-ResSeg, M3Net, MULLET и др.).
Фундаментальные: MedSAM, VoxTell	Предварительное обучение на крупном наборе медицинских изображений разных модальностей и органов; поддержка интерактивных и/или текстовых запросов; выступают как универсальное «ядро» сегментации.	Чрезвычайно высокие требования к данным и ресурсам на этапе обучения; эксплуатация возможна на стандартном оборудовании.	Потенциально пригодны для 3D и мультифазной КТ, обеспечивают zero/few-shot адаптацию к новым органам и протоколам; однако требуют существенной донастройки под конкретные фазовые протоколы и схемы слияния фаз.

Перспективным направлением выступает совместная оптимизация сопряженных задач. Архитектура JSSR демонстрирует синергетический эффект от объединения регистрации, синтеза контраста и сегментации в единый вычислительный граф: градиенты от решения задачи сегментации направляют регистратор к выравниванию именно тех анатомических областей, которые критичны для точности маски. Подобные

подходы открывают путь к комплексным системам, где сегментация интегрируется с дополнительными диагностическими задачами в едином конвейере обработки.

Дальнейшее развитие области связано с тремя ключевыми трендами. Во-первых, применение фундаментальных моделей (MedSAM, VoxTell), предобученных на мультимодальных медицинских данных, что потенциально позволяет достигать приемлемого качества при минимальном количестве размеченных примеров новой задачи. Во-вторых, модально-агностичное обучение, при котором сеть обучается на случайных подмножествах фаз, формируя устойчивое представление объекта независимо от конкретного режима контрастирования – это критично для развертывания в медицинских учреждениях с нестандартными протоколами сканирования. В-третьих, интеграция визуально-языковых интерфейсов, позволяющих взаимодействовать с сегментатором через текстовые запросы, что упрощает адаптацию под специфические диагностические сценарии без переобучения модели.

В задачах сегментации мультифазных КТ снимков остаются значимые нерешенные проблемы. Вычислительная сложность трехмерных трансформерных архитектур ограничивает их применение в условиях типичной клинической инфраструктуры: обработка полного объема мультифазной КТ (4 фазы × 200–500 срезов) на современном многоядерном CPU занимает от пяти до десяти минут, что часто неприемлемо для рабочего процесса радиолога. Решения этой проблемы включают оптимизацию через дистилляцию знаний в легкие сверточные архитектуры, использование иерархического внимания (оконные трансформеры) и квантизацию моделей до INT8. Другая проблема – сегментация мелких и множественных очагов (<10 мм), где даже мультифазная информация оказывается недостаточной. Здесь перспективны гибридные конвейеры: детекция подозрительных регионов с последующей локальной сегментацией высокого разрешения.

Существенной проблемой остается ограниченная доступность мультифазных датасетов с разметкой сегментации. Большинство передовых моделей валидируются на закрытых институциональных коллекциях, что затрудняет объективное сравнение и тиражирование результатов. Создание открытых бенчмарков, охватывающих разнообразные органы и патологии при полном наборе фаз контрастирования, станет ключевым драйвером прогресса в этой области.

Таким образом, мультифазная сегментация эволюционирует от узкоспециализированной задачи к общей проблеме мультимодального обучения с пространственно-временными зависимостями. Успешные архитектуры объединяют три компонента: механизмы неявного выравнивания признаков, иерархическую интеграцию контрастной информации и адаптивную селекцию релевантных паттернов. Дальнейшее развитие потребует баланса между вычислительной эффективностью, робастностью к вариациям данных и практичностью развертывания в условиях реальной клинической инфраструктуры – задач, находящихся на стыке компьютерного зрения, обработки медицинских изображений и прикладного машинного обучения.

Заключение

Автоматическая сегментация патологических очагов печени на мультифазных КТ-изображениях – активно развивающаяся область, сочетающая достижения медицинской визуализации и глубокого обучения. В данной работе рассмотрены основные подходы к решению этой задачи. Проведенный анализ показывает, что учет мультифазной информации принципиально повышает качество сегментации по сравнению с однофазными алгоритмами, позволяя более полно и точно выявлять патологические изменения. Ключевым фактором успеха стали архитектуры глубоких

нейросетей с механизмами внимания и трансформерными блоками, которые эффективно интегрируют данные с нескольких фаз контрастирования.

Такие модели, как PA-ResSeg, M3Net, MULLET учатся находить соответствия между фазами и усиливать значимые признаки патологических изменений на КТ-снимках, игнорируя артефакты несовпадения изображений. В результате достигается высокая чувствительность (до 94–100 % для злокачественных очагов) при сохранении высокой специфичности, а метрики Dice превышают 0,85–0,90, что сравнимо или превосходит уровень экспертов-радиологов в стандартных условиях.

Однако есть два существенных барьера, сдерживающих переход от исследовательских прототипов к промышленным решениям. Во-первых, острый дефицит общедоступных мультифазных датасетов с разметкой сегментации. Во-вторых, вычислительная сложность трехмерных трансформерных архитектур, требующая оптимизации для развертывания в условиях типичной ИТ-инфраструктуры медицинского учреждения. Решение этих задач позволит преодолеть разрыв между академическими достижениями и практическим внедрением автоматизированных систем анализа мультифазных КТ в практику медицинских учреждений.

Вклад данной работы заключается в систематизации быстро развивающейся области мультифазной сегментации: классификации подходов к интеграции фазовой информации, выделении архитектур, доказавших эффективность на практике.

Можно ожидать, что уже в ближайшие годы мультифазная сегментация найдет применение в системах поддержки принятия решений: уже сейчас прототипы помогают разгрузить врача и определяют ряд значимых патологий. По мере того, как эти технологии будут совершенствоваться и проходить клиническую валидацию, они станут стандартным инструментом в радиологии.

СПИСОК ИСТОЧНИКОВ / REFERENCES

1. Wu L., Wang H., Chen Y., et al. Beyond radiologist-level liver lesion detection on multi-phase contrast-enhanced CT images by deep learning. *iScience*. 2023;26(11). <https://doi.org/10.1016/j.isci.2023.108183>
2. Руденко А.В., Руденко М.А., Каширина И.Л. Применение искусственных нейронных сетей для поиска объектов на медицинских изображениях. *Моделирование, оптимизация и информационные технологии*. 2024;12(3). <https://doi.org/10.26102/2310-6018/2024.46.3.013>
Rudenko A.V., Rudenko M.A., Kashirina I.L. The use of artificial neural networks to search for objects in medical images. *Modeling, Optimization and Information Technology*. 2024;12(3). (In Russ.). <https://doi.org/10.26102/2310-6018/2024.46.3.013>
3. Xu Y., Cai M., Lin L., et al. PA-ResSeg: A phase attention residual network for liver tumor segmentation from multiphase CT images. *Medical Physics*. 2021;48(7):3752–3766. <https://doi.org/10.1002/mp.14922>
4. Ronneberger O., Fischer Ph., Brox Th. U-Net: Convolutional networks for biomedical image segmentation. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015: 18th International Conference: Proceedings: Part III, 05–09 October 2015, Munich, Germany*. Cham: Springer; 2015. P. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
5. Çiçek Ö., Abdulkadir A., Lienkamp S.S., Brox Th., Ronneberger O. 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In: *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016: 19th International Conference: Proceedings: Part II, 17–21 October 2016, Athens, Greece*. Cham: Springer; 2016. P. 424–432. https://doi.org/10.1007/978-3-319-46723-8_49

6. Milletari F., Navab N., Ahmadi S.-A. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In: *2016 Fourth International Conference on 3D Vision (3DV), 25–28 October 2016, Stanford, CA, USA*. IEEE; 2016. P. 565–571. <https://doi.org/10.1109/3DV.2016.79>
7. Isensee F., Jaeger P.F., Kohl S.A.A., Petersen J., Maier-Hein K.H. nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*. 2021;18:203–211. <https://doi.org/10.1038/s41592-020-01008-z>
8. Куликов А.А., Каширина И.Л., Савкина Е.Ф. Сегментация объемных образований печени на мультифазных КТ-изображениях с использованием фреймворка nnU-Net. *Моделирование, оптимизация и информационные технологии*. 2025;13(1). <https://doi.org/10.26102/2310-6018/2025.48.1.040>
Kulikov A.A., Kashirina I.L., Savkina E.F. Segmentation of liver volumetric lesions in multiphase CT images using the nnU-Net framework. *Modeling, Optimization and Information Technology*. 2025;13(1). (In Russ.). <https://doi.org/10.26102/2310-6018/2025.48.1.040>
9. Dosovitskiy A., Beyer L., Kolesnikov A., et al. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In: *9th International Conference on Learning Representations, ICLR 2021, 03–07 May 2021, Virtual Event, Austria*. 2021. <https://doi.org/10.48550/arXiv.2010.11929>
10. Zheng S., Lu J., Zhao H., et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In: *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 20–25 June 2021, Nashville, TN, USA*. IEEE; 2021. P. 6877–6886. <https://doi.org/10.1109/CVPR46437.2021.00681>
11. Chen J., Lu Y., Yu Q., et al. *TransUNet: Transformers make strong encoders for medical image segmentation*. arXiv. URL: <https://doi.org/10.48550/arXiv.2102.04306> [Accessed 15th January 2026].
12. Hatamizadeh A., Tan Y., Nath V., et al. UNETR: Transformers for 3D medical image segmentation. In: *Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 03–08 January 2022, Waikoloa, HI, USA*. IEEE; 2022. P. 1748–1758. <https://doi.org/10.1109/WACV51458.2022.00181>
13. Hatamizadeh A., Nath V., Tang Y., et al. Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. In: *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: 7th International Workshop, BrainLes 2021, Held in Conjunction with MICCAI 2021: Revised Selected Papers: Part I, 27 September 2021, Virtual Event*. Cham: Springer; 2022. P. 272–284. https://doi.org/10.1007/978-3-031-08999-2_22
14. Oktay O., Schlemper J., Folgoc L.L., et al. *Attention U-Net: Learning where to look for the pancreas*. arXiv. URL: <https://doi.org/10.48550/arXiv.1804.03999> [Accessed 15th January 2026].
15. Alirri O.I. Dual attention U-Net for liver tumor segmentation in CT images. *International Journal of Computers, Communications & Control*. 2024;19(2).
16. Достовалова А.М., Горшенин А.К., Старичкова Ю.В., Арзамасов К.М. Сравнительный анализ модификаций нейросетевых архитектур U-Net в задаче сегментации медицинских изображений. *Digital Diagnostics*. 2024;5(4):833–853. <https://doi.org/10.17816/DD629866>
Dostovalova A.M., Gorshenin A.K., Starichkova J.V., Arzamasov K.M. Comparative analysis of modifications of U-Net neuronal network architectures in medical image segmentation. *Digital Diagnostics*. 2024;5(4):833–853. <https://doi.org/10.17816/DD629866>

17. Ma J., He Y., Li F., et al. Segment anything in medical images. *Nature Communications*. 2024;15. <https://doi.org/10.1038/s41467-024-44824-z>
18. Rokuss M., Langenberg M., Kirchhoff Y., et al. VoxTell: Free-text promptable universal 3D medical image segmentation. arXiv. URL: <https://doi.org/10.48550/arXiv.2511.11450> [Accessed 15th January 2026].
19. Старичкова Ю.В., Питинов А.В., Газанова Н.Ш. Разработка метода регистрации мультифазных КТ-изображений с использованием аффинных преобразований. В сборнике: *Медэлектроника–2024. Средства медицинской электроники и новые медицинские технологии: сборник научных статей XIV Международной научно-технической конференции, 05–06 декабря 2024 года, Минск, Беларусь*. Минск; 2024. С. 237–240.
Starichkova Yu.V., Pitinov A.V., Gazanova N.Sh. Development of a method for registration of multiphase CT images using affine transformations. In: *Medelectronics–2024. Medical electronics and new medical technologies: Proceedings of the XIV International Scientific and Technical Conference, 05–06 December 2024, Minsk, Belarus*. Minsk; 2024. P. 237–240. (In Russ.).
20. Liu F., Cai J., Huo Y., et al. JSSR: A joint synthesis, segmentation, and registration system for 3D multi-modal image alignment of large-scale pathological CT scans. In: *Computer Vision – ECCV 2020: 16th European Conference: Proceedings: Part XIII, 23–28 August 2020, Glasgow, UK*. Cham: Springer; 2020. P. 257–274. https://doi.org/10.1007/978-3-030-58601-0_16
21. Zhou Y., Li Y., Zhang Zh., et al. *Hyper-Pairing Network for Multi-Phase Pancreatic Ductal Adenocarcinoma Segmentation*. arXiv. URL: <https://doi.org/10.48550/arXiv.1905.00367> [Accessed 20th January 2026].
22. Qu T., Wang X., Fang Ch., et al. M³Net: A multi-scale multi-view framework for multi-phase pancreas segmentation based on cross-phase non-local attention. *Medical Image Analysis*. 2022;75. <https://doi.org/10.1016/j.media.2021.102232>
23. Lteif D., Appapogu D., Bargal S.A., Plummer B.A., Kolachalama V.B. Anatomy-guided, modality-agnostic segmentation of neuroimaging abnormalities. *Human Brain Mapping*. 2025;46(14). <https://doi.org/10.1002/hbm.70329>
24. Antonelli M., Reinke A., Bakas S., et al. The Medical Segmentation Decathlon. *Nature Communications*. 2022;13. <https://doi.org/10.1038/s41467-022-30695-9>
25. Wu X., Su H., Hua Y., et al. A multi-phase CT dataset for automated differential diagnosis of liver tumors. *Scientific Data*. 2026;13. <https://doi.org/10.1038/s41597-025-06343-4>
26. Bartnik K., Bartczak T., Krzyziński M., et al. WAW-TACE: A hepatocellular carcinoma multiphase CT dataset with segmentations, radiomics features, and clinical data. *Radiology: Artificial Intelligence*. 2024;6(6).
27. Elbatel M., Yi Q., Huang X., et al. Triphasic-aided Liver Lesion Segmentation in Non-contrast CT (TriALS) Challenge. In: *Medical Image Computing and Computer Assisted Intervention 2025 (MICCAI)*. <https://doi.org/10.5281/zenodo.15087646>

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Самсоненко Станислав Владимирович, Stanislav V. Samsonenko, Assistant, Department ассистент, кафедра технологий искусственного интеллекта, МИРЭА – Российский технологический университет, Москва, Российская Федерация. of Artificial Intelligence Technologies, MIREA – Russian Technological University, Moscow, Russian Federation.
e-mail: samsonenko@mirea.ru
ORCID: [0009-0006-9304-015X](https://orcid.org/0009-0006-9304-015X)

Каширина Ирина Леонидовна, доктор технических наук, профессор, кафедра технологий искусственного интеллекта, МИРЭА – Российский технологический университет, Москва, Российская Федерация. **Irina L. Kashirina**, Doctor of Engineering Sciences, Professor, Department of Artificial Intelligence Technologies, MIREA – Russian Technological University, Moscow, the Russian Federation.

e-mail: kash.irina@mail.ru

ORCID: [0000-0002-8664-9817](https://orcid.org/0000-0002-8664-9817)

Статья поступила в редакцию 02.02.2026; одобрена после рецензирования 16.03.2026; принята к публикации 24.03.2026.

The article was submitted 02.02.2026; approved after reviewing 16.03.2026; accepted for publication 24.03.2026.