

ОБНАРУЖЕНИЕ И КЛАССИФИКАЦИЯ ОБЪЕКТОВ НА ТРЕХМЕРНЫХ ИЗОБРАЖЕНИЯХ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ

© 2024 г. И. А. Матвеев^{b, *}, А. А. Юрченко^{a, b, **}

^aФедеральное государственное автономное образовательное учреждение высшего образования “Московский физико-технический институт (национальный исследовательский ун-т)” МФТИ, Москва, Россия

^bФИЦ ИУ РАН, Москва, Россия

*e-mail: matveev@frccsc.ru

**e-mail: yurchenko.aa@phystech.edu

Поступила в редакцию 13.09.2024 г.

После доработки 27.08.2024 г.

Принята к публикации 16.10.2024 г.

Рассматривается задача автоматического обнаружения и классификации объектов на трехмерных одноканальных растровых изображениях. Объекты могут иметь малую контрастность и значительную изменчивость формы, что затрудняет построение модели в явном виде. Для решения используются методы машинного обучения по размеченному базе прецедентов. Предлагается двухэтапная обработка, на первом шаге производится обнаружение объектов на изображении, на втором – уменьшение доли ложноположительных результатов и классификация. Применяется подход глубокого обучения с единственным входом и тренировкой на одновременное решение нескольких задач. Решается актуальная практическая задача построения клинически применимой автоматической системы поддержки принятия решений при обнаружении и определении категорий переломов ребер по данным снимков компьютерной томографии. Проведены вычислительные эксперименты на общедоступном наборе данных RibFrac. Показано, что предложенная система обеспечивает значение чувствительности обнаружения 0.935 при среднем количестве ложноположительных предсказаний на одно изображение 4.7. Приводится сравнение полученного алгоритма с существующими методами на основании количественных мер.

Ключевые слова: компьютерная томография, детекция переломов ребер, классификация переломов ребер, глубокое обучение

DOI: 10.31857/S0002338824060091, EDN: SUBUGN

DETECTION AND CLASSIFICATION OF OBJECTS IN THREE-DIMENSIONAL IMAGES USING DEEP LEARNING METHODS

И. А. Матвеев^{b, *}, А. А. Юрченко^{a, b, **}

^aMoscow Institute of Physics and Technology
(National Research University), Moscow, Russia

^bFederal Research Center “Computer Science and Control”
of the Russian Academy of Sciences, Moscow, Russia

*e-mail: matveev@frccsc.ru

**e-mail: yurchenko.aa@phystech.edu

The issue of automatic object detection and categorization in three-dimensional, single-channel, raster images is considered. Objects may have low contrast and substantial shape variability, making it challenging to explicitly construct a model. The proposed solution employs machine learning techniques based on a labeled database of use scenarios. A two-step algorithm is presented, with the first stage being the detection of objects within the image and the second being the reduction of false positives and object categorization. Deep

learning approach is applied with a single input and trained for the simultaneous solution of multiple tasks. The practical goal of developing a clinically viable automatic decision support system to detect and classify rib fractures based on CT scans is being solved. Computational experiments were conducted on the publicly available RibFrac dataset. The proposed system was shown to achieve a detection sensitivity of 0.935, with an average number of false positive predictions per image of 4.7. The resulting algorithm was compared with existing methods using quantitative measures.

Keywords: computed tomography, rib fracture detection, rib fracture classification, deep learning

Введение. Задачи обнаружения, локализации и классификации объемных объектов на трехмерных изображениях значительно сложнее своих двумерных аналогов как с точки зрения формализации и алгоритмизации, так и в вычислительном отношении. Человек воспринимает большинство трехмерных объектов лишь как их поверхности, но не как действительно трехмерную структуру. В связи с этим большинство “классических” (не основанных на машинном обучении) решений при обработке трехмерных данных сразу сводятся к построению поверхностей. Кроме того, сама алгоритмизация работы с трехмерными данными существенно сложнее. В результате разнообразие и качество решения задач для трехмерных изображений существенно уступают таковым для двухмерных. Методы машинного обучения, опирающиеся на извлечение взаимосвязей на базе обучающих данных (набора прецедентов), свободны от таких ограничений. По этой причине методы машинного обучения даже на основании небольшого набора прецедентов достигают результатов, сравнимых или превосходящих по качеству получаемые “классическими” подходами. Все сказанное в полной мере относится к медицинской компьютерной томографии (КТ).

КТ дает возможность получить трехмерное изображение структур тела человека с высокой детализацией и является одним из наиболее эффективных методов обнаружения переломов ребер, позволяя находить значительно больше травм в области грудной клетки в сравнении с рентгенографией. Однако идентификация переломов ребер на тонкосрезовых КТ-изображениях с аксиальным пространственным разрешением 0.5–1.5 мм – трудоемкая задача для медиков. В силу наличия у ребер сложной формы и их диагонального расположения на срезах КТ-изображений, некоторые виды переломов тяжело диагностируются и имеют низкий процент обнаружения в клинической практике.

На практике для анализа КТ-изображений грудной клетки при обнаружении переломов ребер используются алгоритмы криволинейной развертки ребер, основанные на извлечении осевой линии каждого ребра с их последующим проецированием в двухмерную плоскость [1] или преобразованиях каждого аксиального среза изображения через цилиндрическую проекцию [2]. Данные методы позволяют снизить время анализа одного исследования врачом-диагностом, однако имеют более низкую чувствительность по сравнению с исходным трехмерным КТ-изображением [3, 4]. Также было отмечено [4], что применение двухмерной развертки приводит к возникновению ложноположительных результатов, некорректной оценке величины смещения ребер при переломах с диастазом, а также снижению коэффициента согласованности мнений экспертов.

Последние достижения в области искусственного интеллекта и компьютерного зрения привели к быстрому внедрению данных технологий в клиническую диагностику и анализ медицинских изображений. Глубокое обучение позволило достичь высоких результатов во многих задачах медицинской диагностики. В последние годы активно развивался нейросетевой подход к проблеме семантической сегментации переломов ребер на основании снимков КТ. Большая часть данных методов базируется на использовании глубоких сверточных нейросетей 3D U-Net [5], DeepLab v3+ [6], 3D ResNet [7] и их комбинаций, например FracNet [8] и FracNet+ [9].

Привлечение дополнительной информации о сегментационных масках ребер для обнаружения переломов дает значительное улучшение качества работы алгоритмов. Так, в работе [8] рассматривается метод предсказания переломов ребер с помощью скользящего окна по предварительно вычисленным областям расположения ребер. Продолжая данную идею, авторы статьи [9] предлагают добавлять сегментационные маски ребер, предварительно приведенные к необходимому пространственному разрешению и преобразованные набором сверточных модулей, на каждом слое энкодера и декодера сети U-Net. Для выделения области ребер могут применяться алгоритмы глубокого обучения, основанные на сегментации трехмерных воксельных изображений или разреженных облаков точек, например моделями PointNet [10] или DGCNN [11].

Знание о расположении ребер может быть привлечено при обучении модели для решения нескольких задач. Например, в SA-FracNet [12] предлагается одновременное решение задачи сегментации переломов и определение их границы с учетом формы ребер. В работе [13] приводится схема обучения, сочетающая решение задачи детекции переломов на основании двухмерных аксиальных срезов и локализации ребер с использованием трехмерного КТ-изображения. Данный подход позволяет добиться повышения эффективности обучения и точности прогнозирования по сравнению с обучением моделей по отдельности.

С увеличением разнообразия решаемых методами глубокого обучения задач начинает развиваться подход, представляющий из себя автоматический поиск подходящей архитектуры в решении конкретной задачи. Одним из наиболее популярных фреймворков (наборов инструментов), используемых при решении задач обработки двухмерных и трехмерных медицинских изображений, является nnU-Net [14]. Фреймворк представляет из себя метод семантической сегментации, который автоматически адаптируется к заданному набору данных. На основании анализа статистик предоставленных обучающих примеров производится автоматическая настройка соответствующего алгоритма сегментации на базе нейросетевой модели U-Net [5]. Настройка подразумевает определение оптимальной конфигурации сети, алгоритмов предварительной обработки изображений и аугментаций данных, используемых в процессе обучения. Инструмент Auto3Dseg [15] также реализует методы поиска оптимальной архитектуры среди заданного класса сегментационных моделей, включающих в себя Swin UNETR [16], UNet [5] и SegResNet [17]. Алгоритмы были апробированы на широком наборе медицинских тестовых данных и продемонстрировали высокие показатели мер качества сегментации и детекции во многих соревнованиях по локализации структур организма и обнаружению патологий.

С ростом объемов открытых наборов данных с медицинскими изображениями различных модальностей становится распространенной практика применения предобученных моделей в решении более узких медицинских задач, например STU-Net [18], которая построена на основании фреймворка nnU-Net. Предварительно обученная на крупномасштабном наборе данных TotalSegmentator [19] модель STU-Net была применена авторами работы [9] в решении задачи сегментации переломов ребер и продемонстрировала более высокие результаты в сравнении с подходом, реализующим обучение аналогичной архитектуры с нуля.

Предварительное обучение базовых моделей для последующего решения целевой задачи может быть выполнено с применением техник обучения с частичным привлечением учителя (англ. semi-supervised learning). Использование данного подхода позволяет включить в обучение значительно больше КТ-изображений, привлекая объемы немаркированных снимков, число которых зачастую значительно превосходит количество размеченных сканов. В работе [13] предлагается применение техники, основанной на контрастном обучении (англ. contrastive learning) на уровне пикселей, для предварительной настройки весов базовой модели с последующей тонкой настройкой параметров сети для решения задачи сегментации переломов ребер.

1. Постановка задачи. На вход подается КТ-изображение области I , являющееся трехмерной матрицей размерами $W \times H \times D$, содержащей целые числа из диапазона от -1024 до 3071 , которые задают значение плотности каждого voxel изображения по шкале денситометрических показателей Хаунсфилда (англ. hounsfield units (HU)). Требуется выделить области изображения, на которых присутствуют объекты (переломы ребер), и определить класс каждого объекта из заданного списка возможных клинических категорий переломов: консолидированный перелом, перелом с диастазом (смещением), перелом без диастаза.

С точки зрения решения задачи сегментации результат работы алгоритма может быть описан в виде сегментационной маски объектов, представляющей собой трехмерную матрицу F , которая соответствует размерам исходного изображения $I - W \times H \times D$, и списка классов объектов C . Значения voxelей сегментационной маски F являются целыми числами в диапазоне от 0 до N , где N – количество обнаруженных на изображении объектов. Вoxelы маски, имеющие значение, равное $i \in [1, N]$, описывают область локализации объекта i , принадлежащего к классу $C[i]$. Значение O voxelей маски соответствует заднему плану и не описывает объекты. Обозначая через F_i набор координат, соответствующих области сегментационной маски $F_i = \{(x, y, z), \text{ если } F(x, y, z) = i\}$, получаем область локализации объекта на исходном изображении $I_{F_i} = \{I(x, y, z) : (x, y, z) \in F_i\}$ и ее класс $C[i]$.

1.1. Задача детекции. В большинстве работ, посвященных решению задачи локализации и классификации переломов ребер, рассматриваются методы семантической сегментации, использующие в качестве целевых объектов предсказания voxelные сегментационные маски.

Однако начальная постановка задачи сегментации как выделения области исходного КТ-изображения имеет несколько недостатков, усложняющих получение результатов. Во-первых, внешний вид переломов сильно варьируется между различными классами. Переломы с диастазом имеют различной степени выраженности разрывы между двумя частями одного ребра. Сегментация патологии такого вида может интерпретироваться неоднозначно – к сегментационной маске перелома могут быть отнесены различные объемы кости на концах частей ребра и окружающего вещества, находящегося в разрыве. В то же время консолидированные переломы, являющиеся полноценным восстановлением кости после перелома, не имеют четких границ. Во-вторых, на данный момент не существует единой методологии, регламентирующей правила нанесения разметки для различных классов переломов. Мaska области перелома (ее форма, объем и ориентация) сильно зависит от специалиста, выполняющего разметку. Разметка существующих наборов данных КТ-снимков пациентов с переломами составляется широким кругом рентгенологов и врачей-диагностов, каждый из которых придерживается собственного метода определения объема и формы области перелома. Так, в открытых наборах данных переломы одного и того же вида, размеченные разными специалистами, могут иметь объемы сегментационных масок, отличающиеся в 3 и более раз. В некоторых случаях экспертом может быть размечена вся область ребра, содержащая перелом, в других – только непосредственная область перелома. Часто эксперты делают разный отступ разметки от границ ребра. Описанные различия в сегментационной разметке переломов со смещениями представлены на рис. 1. На изображениях 1, 3 и 7 маска выступает за границы кости самого ребра. Для изображений 5, 6 и 8 маска перекрывает всю или практически всю область ребра, в то время как для переломов 2 и 4 к сегментационной маске отнесены меньшие объемы костной ткани.

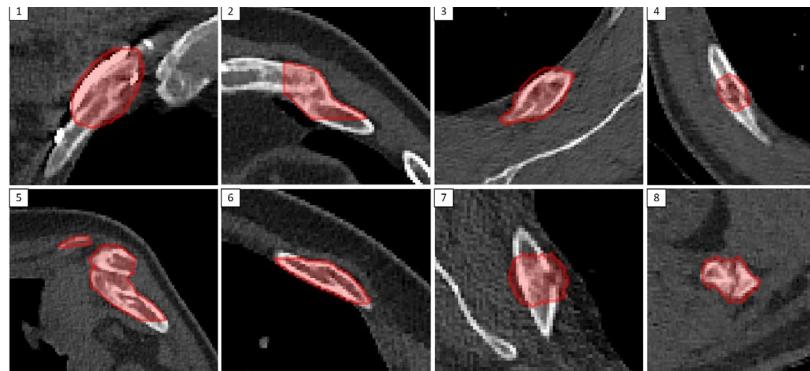


Рис. 1. Примеры, демонстрирующие различие в сегментационной разметке переломов со смещениями.

Исходя из этого, задача переформулирована как задача определения параллелепипеда, ограничивающего область интереса (bounding box), что принято называть *задачей детекции*. При необходимости можно перейти к сегментационным маскам путем выделения области объекта, но уже внутри ограничивающего параллелепипеда, а не на всем изображении.

В терминах задачи детекции результат представляется в виде набора ограничивающих параллелепипедов B , каждый элемент которого задается шестью координатами – минимальными и максимальными координатами вершин ограничивающего параллелепипеда по трем осям:

$$B[i] = \left(x \in [x_{\min}^{(i)}, x_{\max}^{(i)}], y \in [y_{\min}^{(i)}, y_{\max}^{(i)}], z \in [z_{\min}^{(i)}, z_{\max}^{(i)}] \right), i \in \overline{1, N}, \quad (1.1)$$

и списка классов объектов C . Количество элементов списков B и C совпадает и равно количеству обнаруженных объектов N . Обозначим как $I_{B[i]}$ область изображения I , заключенную внутри параллелепипеда $B[i]$. Формально

$$I_{B[i]} = \{I(x, y, z) : (x, y, z) \in B[i]\}. \quad (1.2)$$

Область $I_{B[i]}$ является локализацией объекта i , относящегося к классу $C[i]$.

1.2. Переход к задаче детекции. Преобразование от сегментационных масок к ограничивающим параллелепипедам достаточно просто и сводится к вычислению минимальных и максимальных координат вокселей маски по трем осям. Пусть F – сегментационная маска объектов, B – список ограничивающих параллелепипедов, задаваемый изначально

как пустой. Пусть $F_i = \{(x, y, z) : M(x, y, z) = i\}$ – сегментационная маска перелома i в координатах M . Тогда прямоугольные границы маски перелома i с учетом отступа в 1voxель могут быть записаны как

$$x_{\min}^{(i)} = \max(0, \min\{x : (x, y, z) \in F_i\} + 1),$$

$$x_{\max}^{(i)} = \min(W', \max\{x : (x, y, z) \in F_i\} + 1).$$

Аналогично определяются остальные границы параллелепипеда (1.1) – $y_{\min}^{(i)}, y_{\max}^{(i)}, z_{\min}^{(i)}, z_{\max}^{(i)}$.

2. Архитектура системы. Представленная система методов решения задачи локализации и классификации переломов ребер на КТ-изображениях имеет следующие основные блоки:

1) предварительная обработка КТ-изображения, включающая выделение области ребер методом сегментации разреженного облака точек с применением модели PointNet++, приведение к целевому пространственному разрешению и нормирование интенсивности;

2) определение областей локализаций переломов ребер в виде набора ограничивающих параллелепипедов с применением детекционной модели;

3) формирование итоговых детекционных предположений с учетом определенных экспериментально оптимальных гиперпараметров отбора ограничивающих параллелепипедов;

4) разделение полученных локализаций переломов на клинические категории и удаление ложноположительных предсказаний с помощью классификационной модели.

Отдельно стоит отметить, что дополнительный выход детекционной сети, реализующий решение задачи семантической сегментации ребер, используется только в процессе обучения и может быть исключен на этапе применения модели при диагностировании переломов.

Общая архитектурная схема предлагаемой системы приведена на рис. 2.

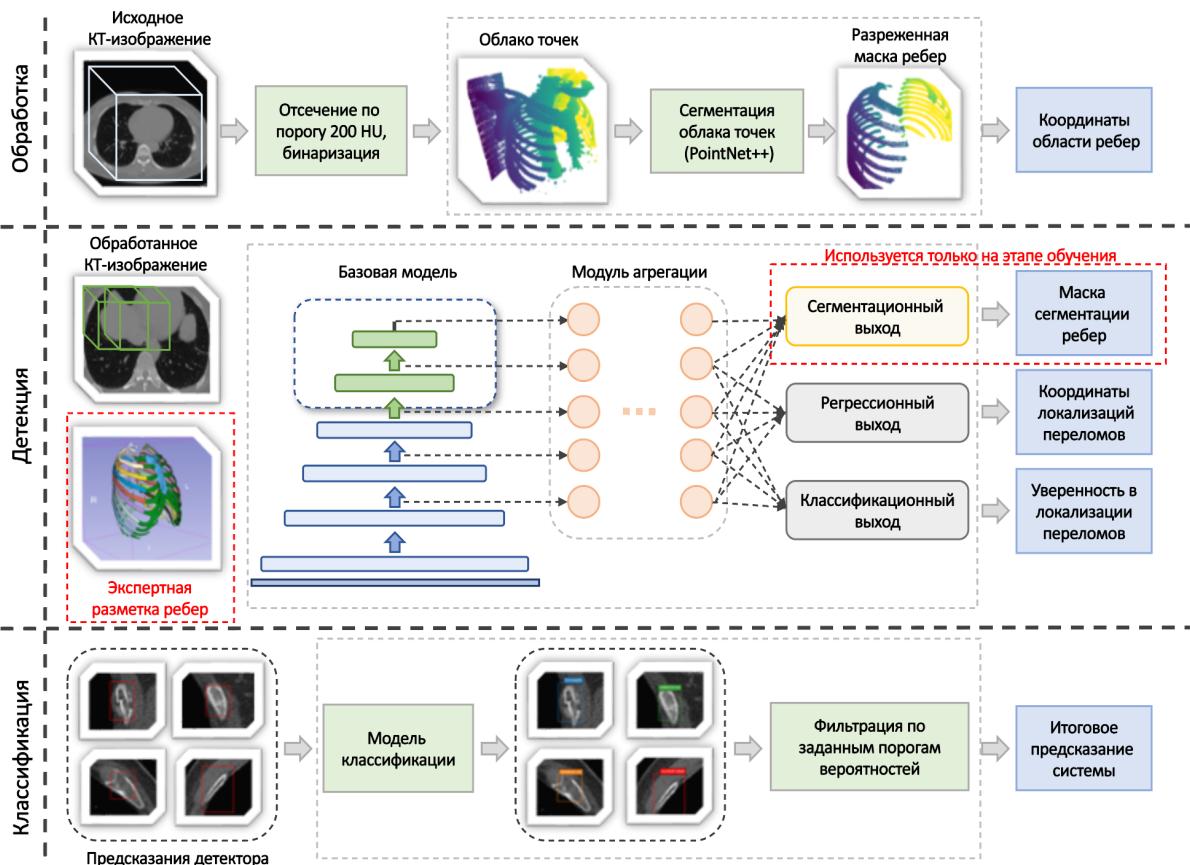


Рис. 2. Архитектурная схема предлагаемого двухэтапного метода, включающего детектирование переломов с последующей их классификацией.

2.1. П р е д о б р а б о т к а д а н н ы х . Первоначальным этапом обработки КТ-исследований является приведение всех изображений к единому пространственному разрешению. Данное преобразование необходимо для выравнивания расстояний расположения анатомических структур относительно друг друга. В качестве целевого пространственного разрешения было выбрано медианное значение спейсинга для КТ-сканов из отобранной части исследований датасета RibFrac – (1.25, 0.75, 0.75). Преобразование КТ-изображений к указанному пространственному разрешению осуществлялось методом трилинейной интерполяции, преобразование соответствующих сегментационных масок путем интерполяции методом ближайших соседей.

Также важным этапом обработки исходных КТ-изображений, непосредственно влияющим на качество и скорость работы алгоритмов обнаружения целевых патологий, является предварительное выделение области изображения, в котором анатомически предполагается локализация искомой патологии. В задаче обнаружения переломов ребер естественным ограничением области поиска выступает граница внутри грудной клетки, включающая в себя ребра. Непосредственное выделение области ребер позволяет исключить другие костные структуры (такие как ключицы, нижние элементы черепа и части позвоночника) из рассмотрения, что уменьшает общую ошибку обнаружения переломов и снижает количество ложноположительных предсказаний. Более того, это дает возможность уменьшить размер исходного изображения, тем самым сокращая время работы и количество используемой памяти, что особенно важно при обработке трехмерных изображений методами глубокого обучения с помощью сверточных нейросетей.

Для выделения целевой области на КТ-изображении может быть использован ряд методов, основанных на априорных знаниях о плотности структур и тканей, эвристических правилах, морфологических операциях, а также алгоритмах компьютерного зрения. Зачастую при сегментации области интереса не требуется субвоксельная точность работы алгоритма. В силу данного факта при применении методов компьютерного зрения в решении данной задачи могут быть использованы изображения, предварительно переведенные в более низкое пространственное разрешение методами интерполяции. Таким образом, сегментация области грудной клетки, содержащей в себе ребра, может быть выполнена посредством применения сегментационных нейросетевых методов, например U-Net-подобных архитектур.

В статье для выделения области ребер предлагается использовать подход, основанный на сегментации облака точек алгоритмами глубокого обучения. Данный метод широко применяется в задачах, связанных с сегментацией ребер, выделением центр-линий ребер и обнаружением переломов. Авторы работ [22, 23] рассматривают модель сегментации разреженного облака точек PointNet++ [10] непосредственно для решения задачи определения маски каждого ребра и последующего выделения центральных линий. В работах [8, 9] применяется данный метод как вспомогательный при решении задачи сегментации переломов ребер.

Преобразование исходного КТ-изображения в облако точек осуществляется путем отсечения вокселей по порогу 200 HU с последующей бинаризацией. Предварительно производится понижение линейного пространственного разрешения КТ-изображения в 2 раза в целях ускорения работы алгоритма и сглаживания отдельных вокселей высокой плотности. Полученное облако точек сегментируется на два класса (ребра и фон) с помощью модели PointNet++, обученной на случайных подвыборках размером 30 тыс. элементов из исходного облака точек. В процессе обучения модели минимизировалась функция потерь перекрестной энтропии.

Обучение модели происходило с использованием сегментационных масок ребер, представленных в датасете RibSeg v2 [22]. По аналогии с преобразованием исходных КТ-сканов осуществлялось приведение разметки ребер к целевому пространственному разрешению с последующим понижением разрешения в 2 раза путем интерполяции методом ближайших соседей.

Предсказание модели на этапе использования для всего облака точек проводилось путем последовательного выбора некоторого количества непересекающихся наборов по 30 тыс. точек в каждом с последующим применением обученной модели для каждого из них. В силу отсутствия требования классификации каждой точки количество выборок не обязательно должно покрывать все облако точек – для достижения необходимой точности локализации области ребер может быть использовано небольшое число случайных выборок. Эмпирически установлено, что количество выборок, равное трем, обеспечивает баланс между скоростью работы и точностью локализации области ребер. Общая структура алгоритма сегментации ребер на базе разреженного облака точек приведена на рис. 3.

На основе полученной маски сегментации ребер вычисляются координаты ограничивающего параллелепипеда с добавлением отступа в 5 вокселей по каждому направлению и производится обрезание исходного изображения. Использование отступов позволяет уменьшить возможные погрешности модели сегментации и не терять информацию об анатомических

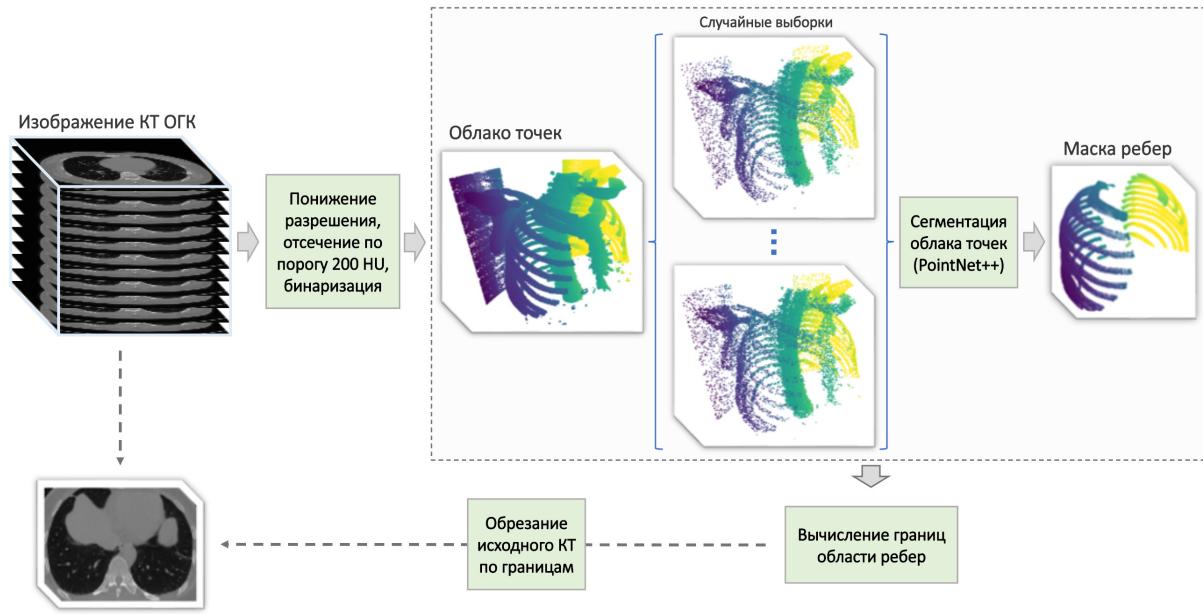


Рис. 3. Сегментация ребер с использованием преобразования исходного КТ-изображения в разреженное облако точек и модели PointNet++.

структур на краях обрезанного изображения. Аналогичные преобразования применяются к соответствующим сегментационным маскам переломов и ребер.

В рамках разработки модели семантической сегментации ребер также был проанализирован метод сегментации воксельного КТ-изображения с помощью нейросетевой модели SegResNet, предварительно обученной на наборе данных TotalSegmentator [19]. Среднее время предсказания алгоритма с учетом предварительной обработки исследования, работы модели и финальной обработки результата, включающей фильтрацию малых областей сегментации и применение морфологических операций, составило 15–20 с, что в рамках разрабатываемой системы является превышением допустимого времени работы. Использование архитектур на основании сети U-Net при аналогичном качестве требует большего времени на формирование предсказания.

Таким образом, предварительная обработка КТ-изображения и сегментационных масок переломов может быть представлена в виде следующего алгоритма:

1. Исходное изображение I и сегментационные маски переломов F размерами $W \times H \times D$ преобразуются к пространственному разрешению (1.25, 0.75, 0.75) методами трилинейной интерполяции и ближайших соседей соответственно. Результат данного шага – изображение I' и маски переломов F' размерами $W' \times H' \times D'$.

2. Методом трилинейной интерполяции разрешение изображения I' понижается в 2 раза и выполняется его преобразование в облако точек O путем отсечения вокселей по порогу 200 HU; $O = \{(x, y, z) : I'_{1/2}(x, y, z) > 200\}$.

3. К трем случайным выборкам размерами 30 тыс. элементов в каждом из результирующего облака точек применяется обученная модель сегментации PointNet++, классифицирующая каждую точку изображения на ребра ($c_r = 1$) и фон ($c_{bg} = 0$). Формируется набор меток классов $S = \{s_i = \text{PointNet}(o_i) \in [0, 1] : o_i \in O\}$. Тогда маска ребер может быть определена как $R = \{o_i \in O : s_i = 1\}$.

4. Задаются границы маски ребер с учетом отступа 5 вокселей по каждому из пространственных направлений:

$$x_{\min} = \max(0, 2 \cdot \min\{x : (x, y, z) \in R\} - 5),$$

$$x_{\max} = \min(W', 2 \cdot \max\{x : (x, y, z) \in R\} + 5).$$

Аналогично находятся координаты $y_{\min}, y_{\max}, z_{\min}, z_{\max}$.

5. Изображение I' и маски переломов F' обрезаются в соответствии с полученными координатами:

$$I'' = \{I'(x, y, z) : x_{\min} \leq x \leq x_{\max}, y_{\min} \leq y \leq y_{\max}, z_{\min} \leq z \leq z_{\max}\},$$

$$F'' = \{F'(x, y, z) : x_{\min} \leq x \leq x_{\max}, y_{\min} \leq y \leq y_{\max}, z_{\min} \leq z \leq z_{\max}\}.$$

2.2. М етод сегментации. В силу того, что на первом этапе работы предлагаемого алгоритма обнаружения переломов ребер требуется только их локализация, рассматривается задача одноклассовой семантической сегментации методами глубокого обучения.

В качестве начальной точки разработки сегментационной модели представлено использование алгоритмов автоматического определения наиболее подходящей архитектуры для заданного набора данных. В решении задачи сегментации переломов ребер приведены 3D-реализации моделей SegResNet с механизмом обучения Deep Supervision [24], U-Net и Swin UNETR v2 [25]. Оптимальные гиперпараметры сети U-Net были подобраны с помощью методов фреймворка nnU-Net [14], модель SegResNet показала лучшие результаты на основании алгоритма выбора модели методами Auto3DSeg [15].

Конфигурация модели SegResNet: число начальных фильтров (каналов на первом слое) – 16, количество блоков энкодера и декодера – (1, 2, 2, 4) и (1, 1, 1) соответственно, глубина используемых карт признаков для механизма Deep Supervision – 3. На каждом блоке энкодера пространственное разрешение изображения уменьшается в 2 раза, количество каналов соответственно увеличивается в 2 раза.

Конфигурация модели SwinUNETR: количество модулей на слоях – (2, 2, 2, 2), число выходов механизма внимания на каждом слое – (3, 6, 12, 24), размер признакового описания скрытого состояния – 24. Применялась вторая версия данной сети.

Конфигурация модели U-Net: количество каналов на слоях энкодера и декодера – (16, 32, 64, 128, 256, 384), функция активации LeakyReLU с наклоном в области отрицательных значений, равным 0.01, размеры сверток – $3 \times 3 \times 3$, шаг ядер свертки на первом слое энкодера – $1 \times 1 \times 1$, на последующих – $2 \times 2 \times 2$. Использовалась модификация сети с применением механизма Skip Connection [26].

Основываясь на результатах исследований, демонстрирующих прирост качества методов сегментации переломов ребер при привлечении дополнительной информации о ребрах [8, 9], был предложен вариант обучения моделей с добавлением второго канала с информацией о voxелях изображения в областях, соответствующих сегментационным маскам ребер. Пусть I – исходное КТ-изображение размером $W \times H \times D$, R – соответствующая изображению маска семантической сегментации ребер. Тогда дополнительный канал I^c может быть представлен в виде трехмерной матрицы размером $W \times H \times D$ с элементами $I^c(x, y, z)$ следующего вида:

$$I^c(x, y, z) = \begin{cases} I(x, y, z), & \text{если } R(x, y, z) \neq 0, \\ -1024 & \text{иначе.} \end{cases} \quad (2.1)$$

В процессе обучения дополнительный канал конкатенировался с исходным изображением и использовался в качестве входа в модель.

При обучении сегментационных моделей функциями потерь выступала сумма функции потерь Focal Loss (2.7) и функции потерь Dice (2.6). Вычисление функции потерь на дополнительных выходах сети SegResNet осуществлялось с предварительным повышением разрешения соответствующих карт признаков до размера целевой сегментационной маски переломов. Компонента i функции потерь, вычисленная для выхода глубины i , нормировалась на значение $1/2^i$. Итоговая функция потерь определялась в виде суммы компонент с различных выходов декодера:

$$L_{total}(y_{true}, y_{pred}) = \sum_{i=0}^{d-1} \frac{1}{2^i} L_i(y_{true}, y_{pred}^{(i)}). \quad (2.2)$$

Здесь $\{y_{pred}^{(i)} : i \in [1, \dots, d]\}$ – выходы модели с последних d слоев декодера.

2.3. А рх и т е к т у р ы д е т е к т о р о в . Первым этапом работы алгоритма обнаружения и классификации переломов является локализация областей, содержащих данную травму. Результат работы данного шага в терминах задачи детекции может быть описан в виде набора ограничивающих параллелепипедов B (1.1) и степени уверенности алгоритма P в наличии перелома в каждой из предсказанных областей. С учетом нормировки уверенность модели $P[i]$

для предсказания с номером i в нахождении перелома в области изображения $I_{B[i]}$ (1.2) может быть интерпретирована как вероятность нахождения патологии в данной области. На этом этапе требуется обнаружение как можно большего числа переломов, т.е. высокое значение чувствительности алгоритма.

Классическим подходом к построению нейросетевой модели детекции является представление архитектуры в виде комбинации трех основных составляющих, включающих в себя базовую модель (англ. backbone), модуль извлечения и агрегации признаковых описаний с различными пространственными разрешениями (англ. neck) и выходы для предсказания координат ограничивающих параллелепипедов и уверенности модели в локализации объекта в заданной области (англ. regression and classification heads). Задача базовой модели заключается в извлечении информативных признаковых описаний различных пространственных разрешений из исходного изображения. Модуль извлечения и агрегации преобразует полученные выходы нижних и верхних уровней базовой модели путем комбинации карт признаков различного пространственного разрешения между собой. Классификационный и регрессионный выходы на основании объединенных карт признаков формируют финальные предсказания, извлекая информацию о классах и координатах объектов на изображении.

В качестве базовой детекционной модели предлагается применять anchor-based [27] детектор, концептуальная схема которого продемонстрирована на рис. 4. Выбор модели, использующей предварительно заданный набор ограничивающих параллелепипедов, основан на эмпирических выводах и результатах сравнения anchor-based и anchor-free детекционных архитектур с точки зрения мер качества полноты и чувствительности [28, 29].

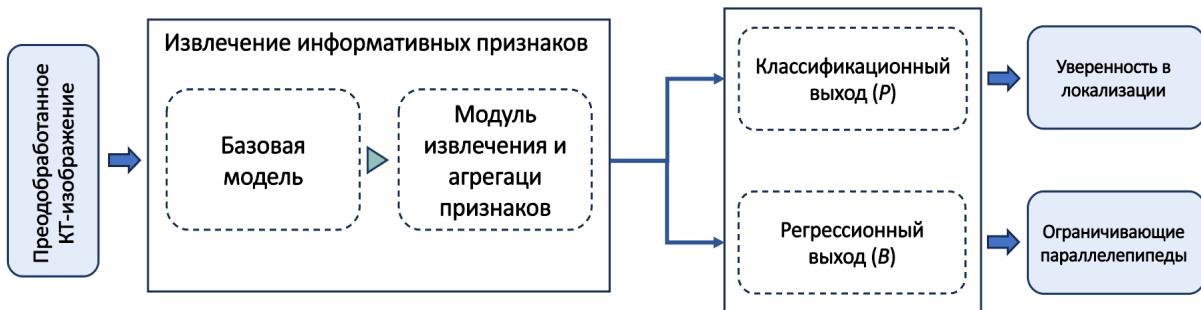


Рис. 4. Концептуальная схема модели детекции переломов.

В рамках разработки и тестирования методов детекции были реализованы 3D-модификации архитектур RetinaNet [30] и EfficientDet [31] с применением в качестве базовых моделей ResNet (residual network) [7], SEResNeXt (squeeze-and-excitation residual network with external transformations) [32, 33], HRNet (high-resolution net) [34] и EfficientNet [35] (B1-B5) соответственно. Сверточные ядра первого слоя в моделях ResNet, SEResNeXt и HRNet были заменены на свертки размером $3 \times 3 \times 3$ и шагом $1 \times 1 \times 1$ с целью увеличения пространственного разрешения начальной и последующих карт признаков в 2 раза. Рассматривались модули агрегации признаков FPN (feature pyramid network) [36] и BiFPN (bidirectional feature pyramid network) [31] с применением различного числа выходных слоев с базовыми моделями, количество которых варьировалось от двух до четырех. Количество карт признаков, используемых в выходных слоях (модулях регрессии и классификации), также варьировалось в диапазоне от трех до пяти. Финальными модулями, преобразующими полученные карты признаков различных пространственных разрешений в логиты классификации и регрессии на ограничивающих параллелепипедах, выступали сверточные блоки, состоящие из четырех последовательных слоев. Выход с последнего пятого слоя в данных модулях не применяется в силу малой пространственной размерности карт признаков данного слоя, получающихся при использовании сегмента КТ-изображения с разрешением $W \times H \times D = 192 \times 192 \times 192$.

Формирование дополнительных карт признаков, используемых в модулях агрегации, осуществлялось с помощью добавления дополнительных блоков к последнему слою базовой модели. Увеличение количества признаковых описаний различных пространственных разрешений на единицу производилось с применением операции MaxPooling с размером ядра $3 \times 3 \times 3$ и шагом $2 \times 2 \times 2$ к последнему выходу сети, на два — путем последовательного применения двух сверточных модулей с финальной операцией MaxPooling. Архитектурная схема детекционной модели приведена на рис. 5.

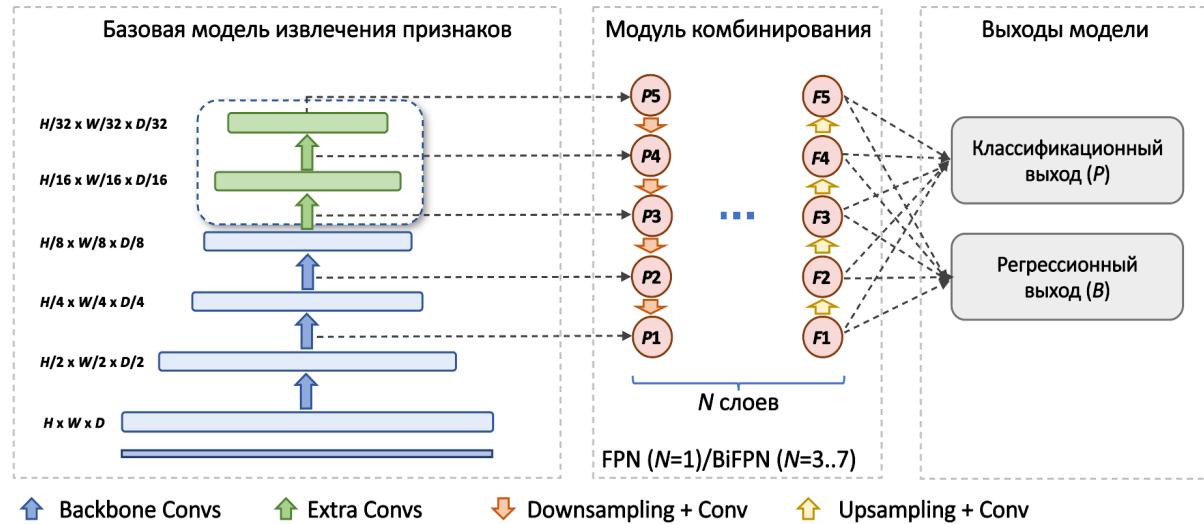


Рис. 5. Общая архитектура модели детекции переломов: базовая модель, модуль комбинирования карт признаков и модули классификации и регрессии.

2.4. Использование дополнительного выхода для сегментации областей интереса. Применение подхода обучения нейросетей на несколько задач в некоторых случаях позволяет добиться более высокого качества решения целевой задачи, а также получить более устойчивую модель [37] за счет выделения более информативного и общего признакового описания исходных данных. Как было отмечено ранее, интуитивным шагом в решении задачи обнаружения переломов ребер является привлечение дополнительной информации в виде масок семантической сегментации ребер [9]. Однако подходы, использующие данную информацию непосредственно на шаге предсказания модели, имеют существенный недостаток. Качество итогового предсказания модели будет зависеть от качества работы алгоритма сегментации ребер, и, соответственно, ошибки на данном этапе могут приводить к увеличению итоговой ошибки работы системы. Более того, увеличивается суммарное время работы алгоритма. Таким образом, предлагается использовать информацию о семантических масках ребер только на этапе обучения модели. В исследовании рассматривались сегментационные маски ребер, представленные в датасете RibSeg v2 [22].

В качестве дополнительного выхода детекционной модели предлагается применять модуль сегментации ребер на изображениях пониженного разрешения. Понижение разрешения осуществляется в 2 раза для соответствия пространственному разрешению первой карты признаков, формирующейся на выходе базовой модели и модуля агрегации. Среди вариантов реализации дополнительного сегментационного выхода были рассмотрены простые реализации декодеров, представляющие из себя объединение карт признаков с различными пространственными разрешениями в финальную сегментационную маску, и U-Net-подобные архитектуры, такие как 3D U-Net, 3D U-Net++ [38] и 3D Attention U-Net [39]. Картами признаков энкодеров данных модулей выступают выходы с модуля извлечения и комбинирования признаков детекционной модели, предварительно преобразованные сверточными модулями.

Особое внимание в данном случае будет уделено модификации сети Attention U-Net, схема которой представлена на рис. 6. Карты признаков каждого слоя декодера получаются путем конкатенации экстраполированных карт признаков декодера с предыдущего слоя и карт признаков энкодера, прошедших через модуль Attention Gate. Механизм пространственного внимания позволяет сети сфокусироваться на важных регионах изображения, игнорируя нерелевантные или шумные участки. С точки зрения подхода к построению архитектуры на решение нескольких задач использование данного приема позволит модели выучивать важные признаки как для задачи детекции переломов ребер, так и для задачи сегментации костных структур ребер. Более того, информация о расположении ребер теоретически может заметно снизить число ложноположительных предсказаний, локализованных вне ребер, так как у модели будет информация о структуре костной системы грудной клетки.

Положим, что $\{f^l\}_{l=1}^n$ — карты признаков, получаемые на выходе из модуля извлечения признаков детекционной модели, например BiFPN. Здесь n — количество выходов модуля

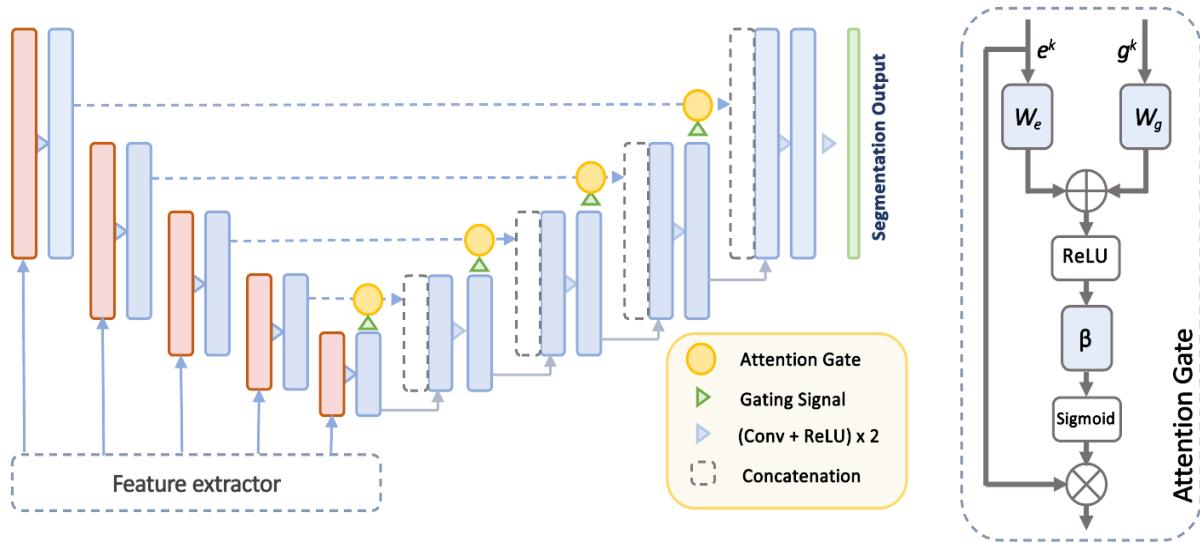


Рис. 6. Архитектура модифицированного варианта сети Attention U-Net с механизмом пространственного внимания Attention Gate.

извлечения признаков. Тогда картами признаков энкодера будет являться набор $\{e^l\}_{l=1}^n$, в котором e^l получается из f^l путем последовательного применения модулей свертки, батч-нормализации [40] и функций активации.

Механизм внимания Attention Gate может быть формализован в следующем виде. Пусть e^l – карта признаков энкодера на слое l размером $C_e^l \times W_e^l \times H_e^l \times D_e^l$, d^l – карта признаков на слое декодера l размером $C_d^l \times W_d^l \times H_d^l \times D_d^l$ (нумерация в обратном порядке по мере уменьшения пространственного размера карт признаков). Входной сигнал g^l (gating signal) для модуля внимания формируется на основании карты признаков декодера d^l путем последовательного выполнения операций интерполяции, свертки, батч-нормализации и функции активации. Размер g^l равен $C_g^l \times W_g^l \times H_g^l \times D_g^l$ и совпадает с пространственным разрешением соответствующей карты признаков энкодера. С помощью свертки $1 \times 1 \times 1$ полученные карты признаков выравниваются по числу каналов. Данные карты признаков g_1^l размера $C_{inner}^l \times W_e^l \times H_e^l \times D_e^l$ и e_1^l размера $C_{inner}^l \times W_e^l \times H_e^l \times D_e^l$ складываются с последующим применением функции активации ReLU. С помощью последовательного использования операций свертки, батч-нормализации и функции активации-сигмоиды формируется карта внимания размерами $1 \times W_e^l \times H_e^l \times D_e^l$, которая поэлементно умножается на карту признаков энкодера e^l . Формирование карты признаков d^{l-1} следующего слоя декодера осуществляется путем конкатенации полученного результата с картой признаков g^l :

$$\begin{aligned} Q_a &= \beta^T \left(\sigma_2 \left(\mathbb{W}_e^T e^l + \mathbb{W}_g^T g^l + b_g \right) \right) + b_\beta, \\ d^{l-1} &= \left[\sigma_1(Q_a) * e^l, g^l \right], \\ \mathbb{W}_e &\in \mathbb{R}^{C_e \times C_{inner}}, \mathbb{W}_g \in \mathbb{R}^{C_g \times C_{inner}}, \beta \in \mathbb{R}^{C_{inner} \times 1}, b_\beta \in \mathbb{R}, b_g \in \mathbb{R}^{C_{inner}}. \end{aligned} \quad (2.3)$$

Здесь σ_1 – функция активации ReLU, σ_2 – функция активации сигмоида, “*” – операция поэлементного произведения.

2.5. Функция потерь общей модели включает в себя три компоненты в соответствии с тремя различными выходами сети: классификационным, регрессионным (на ограничивающих параллелепипедах) и сегментационным. В силу того, что на данном шаге алгоритма не производится классификация переломов ребер по клиническим категориям, классификационный выход модели предсказывает вероятности только для одного класса, которые могут интерпретироваться как уверенность модели в локализации перелома в рамках

соответствующего ограничивающего параллелепипеда. На основании этого в классификационном выходе сети была применена функция потерь бинарной перекрестной энтропии:

$$BCELoss(y_{true}, y_{pred}) = -\frac{1}{N} \sum_{i=1}^N [y_{true} \log y_{pred} + (1 - y_{true}) \log(1 - y_{pred})]. \quad (2.4)$$

В качестве функции потерь, использующейся в регрессионном выходе сети, был применен сглаженный $L1$ лосс [41] с коэффициентом β , равным $1/9$:

$$SmoothL1Loss(x_n, y_n, \beta) = \begin{cases} 0.5(x_n - y_n)^2 / \beta, & \text{если } |x_n - y_n| < \beta, \\ |x_n - y_n| - 0.5\beta & \text{иначе.} \end{cases} \quad (2.5)$$

Агрегирование данной функции потерь производилось путем усреднения значений по всем предсказаниям и по шести измерениям каждого предсказания.

Целевой функцией оптимизации для сегментационного выхода модели выступала равновесная сумма функций потерь Dice Loss [42] и Focal Loss [30] с коэффициентом γ , равным двум:

$$DiceLoss(y_{true}, y_{pred}) = 1 - \frac{2 \sum (y_{true} * y_{pred}) + \epsilon}{\sum y_{true} + \sum y_{pred} + \epsilon}. \quad (2.6)$$

Здесь “*” в числителе – операция поэлементного произведения. Суммы рассчитываются по всем вокселям предсказанной и действительной масок сегментации объектов. Коэффициент сглаживания $\epsilon = 1$ используется для обеспечения численной стабильности вычисления функции потерь Dice:

$$FocalLoss(y_{true}, y_{pred}, \gamma) = -\alpha_\tau (1 - p_\tau)^\gamma \log(p_\tau), \text{ где } p_\tau = \begin{cases} y_{pred}, & y_{true} = 1, \\ 1 - y_{pred} & \text{иначе.} \end{cases} \quad (2.7)$$

Коэффициент α_τ обеспечивает возможность балансирования классов.

В случае решения задачи многоклассовой сегментации производится усреднение функций потерь по каждому из классов.

2.6. Классификация обнаженных объектов. В некоторых случаях для решения задач многоклассовой сегментации или детекции структур, образований или патологий, имеющих схожую природу и разделенных пространственно друг от друга (например, переломы ребер, узелки в легких или патологии лимфатических узлов), используются двухстадийные алгоритмы [43, 44]. Результатом работы первого этапа являются локализованные цеплевые структуры, не дифференцированные по различным классам. На втором этапе выполняется решение задачи классификации предсказаний, полученных на первом шаге. В ряде задач такой подход позволяет повысить полноту обнаружения патологий и в то же время увеличить меры качества классификации.

В статье предлагается использовать модель классификации итоговых предсказаний алгоритма детекции переломов ребер не только для разделения переломов на требуемые клинические категории, но и с целью уменьшения количества ложноположительных предсказаний первого шага алгоритма. Выходом такого классификатора для одного предсказания детектора будет являться вектор $\tilde{p} = [\tilde{p}_1, \tilde{p}_2, \tilde{p}_3, \tilde{p}_4]$ с компонентами, характеризующими вероятности следующих классов: отсутствие перелома, перелом со смещением, перелом без смещения, консолидированный перелом.

Результат работы модели классификации и примеры различных классов переломов с учетом возможного варианта ложноположительного предсказания детектора приведены на рис. 7. Изображение 1 – консолидированный перелом, 2, 3 – переломы без смещения, 5, 6 – переломы со смещением, 4 – ложноположительное предсказание детекционной модели.

2.7. Меры качества детекции и классификации. На различных этапах работы алгоритма присутствуют разные требования к мерам качества предсказаний. Так, на первом шаге локализации патологий необходимо обнаружение как можно большего числа переломов, т.е. максимизация полноты AR (average recall). В силу того, что решается задача детекции, данная мера качества рассчитывается для определенного уровня значения IoU пересечения предсказаний с реальными локализациями объектов. При вычислении AR используются N (гиперпараметр алгоритма) предсказаний с наибольшей уверенностью модели:

$$Recall@(\text{IoU} = t) = \frac{TP}{TP + FN}. \quad (2.8)$$

Здесь TP – количество верно предсказанных локализаций переломов, для которых IoU с ограничивающим параллелепипедом какого-либо реального перелома больше порога t , FN – коли-

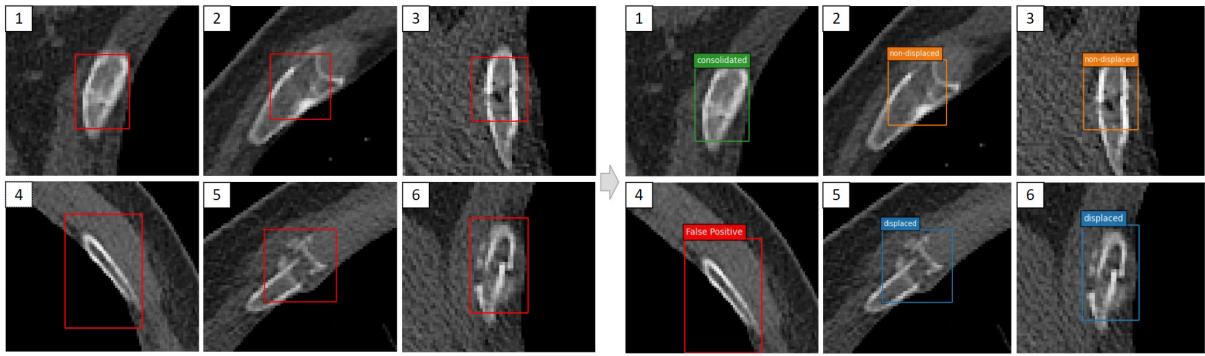


Рис. 7. Примеры различных классов переломов с учетом возможного варианта ложноположительного предсказания детектора.

чество переломов, не обнаруженных на уровне порога t по величине IoU . Также рассматривалась мера качества mAR , являющаяся усреднением AR для различных порогов пересечения по IoU :

$$mAR = \sum_{t \in [t_1, \dots, t_n]} \text{Recall}@(\text{IoU} = t). \quad (2.9)$$

Список порогов t , по которым производится усреднение меры AR , выбирается исходя из требований к желаемой чувствительности локализации переломов. В данной статье целевыми мерами качества работы первого этапа алгоритма были выбраны мера AR для $N = 100$ предсказаний с наибольшей уверенностью алгоритма на уровне значения $IoU = 0.3$ и мера mAR с порогами от 0.1 до 0.5 и шагом 0.05. Выбор данных значений порогов для расчета мер полноты обусловлен результатами визуального анализа предсказаний, для которых значение меры IoU с каким-либо ограничивающим параллелепипедом действительной локализации перелома находилось на уровне от 0.1 до 0.5.

На втором этапе работы алгоритма требуются высокие значения мер точности (precision) и полноты (recall) классификации переломов и объектов, не являющихся переломами, т.е. максимизация макроусредненного по всем классам (отсутствие перелома, перелом со смещением, перелом без смещения, консолидированный перелом) значений данных мер качества:

$$\text{Precision} = \frac{1}{4} \sum_{i=1}^4 \frac{TP_i}{TP_i + FP_i}, \quad (2.10)$$

$$\text{Recall} = \frac{1}{4} \sum_{i=1}^4 \frac{TP_i}{TP_i + FN_i}. \quad (2.11)$$

Также рассматривалась макроусредненная мера качества $F1$:

$$F1 = \frac{1}{4} \sum_{i=1}^4 2 \frac{\text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}. \quad (2.12)$$

Для определения качества детекции применялись меры IoU и $Dice$, вычисленные для N детекционных предположений с наибольшей уверенностью модели:

$$IoU(y_{true}, y_{pred}) = \frac{|y_{true} \cap y_{pred}|}{|y_{true} \cup y_{pred}|}, \quad (2.13)$$

$$Dice(y_{true}, y_{pred}) = \frac{|y_{true} \cap y_{pred}|}{|y_{true}| + |y_{pred}|}. \quad (2.14)$$

Основной мерой качества работы алгоритмов обнаружения патологий, в том числе переломов ребер, принято считать меру FROC (англ. free-response receiver operating characteristic), оценивающую свободный ответ при различных порогах принятия решения. В качестве порогов рассматриваются значения 0.5, 1, 2, 4 и 8 среднего количества ложноположительных

предсказаний переломов алгоритма на одном КТ-изображении, в качестве ответа – значение чувствительности AR (2.8) при $N = 100$ и $IoU = 0.3$.

2.8. Формирование итогового предсказания. Выбор гиперпараметров алгоритма обнаружения и классификации переломов для формирования финального предсказания системы включает в себя определение следующих величин: T_{conf} – степень уверенности алгоритма (confidence), по которой происходит отсечение предсказанных локализаций переломов; N_{map} – максимальное количество детекций, приходящихся на одну карту признаков на каждом уровне с различными пространственными разрешениями; T_{nms} – уровень порога по IoU для работы алгоритма non-maximum suppression; N_{img} – максимальное количество детекций на одно изображение; $T_{non-frac}$ – значение уровня вероятности не принадлежности ни одному из клинических категорий переломов, предсказанный классификатором.

При заданных гиперпараметрах формирование финального предсказания системы может быть представлено в виде следующего алгоритма:

1. Для каждого уровня карт признаков, на которых осуществляется предсказание переломов ребер, удаляются предсказания со значением уверенности модели меньше T_{conf} .

2. Для каждого уровня карт признаков сохраняются только N_{map} предсказаний с наибольшим значением уверенности модели детекции.

3. Для каждого изображения применяется алгоритм non-maximum suppression со значением порога по $IoU = T_{nms}$.

4. Для каждого изображения оставляются только N_{img} предсказаний детектора.

5. Оставляются только те предсказания, для которых уровень вероятности принадлежности к классу, описывающему отсутствие перелома, меньше $T_{non-frac}$.

Данные параметры определяются с помощью методов поиска по сетке, генетических алгоритмов или Байесовской оптимизации. Для ускорения поиска формируются предварительные наборы предсказаний, полученные при эмпирически определенных краевых значениях порогов, которые обеспечивают формирование наиболее широкого набора предсказаний. В качестве целевой оптимизируемой меры качества предлагается рассматривать меру FROC.

3. Набор данных. Основным открытым набором данных, используемых исследователями при разработке и тестировании методов решения задачи обнаружения и классификации переломов ребер, является датасет RibFrac [20]. Датасет содержит 660 снимков компьютерной томографии области грудной клетки пациентов с различными классами переломов ребер. Каждый КТ-скан RibFrac сопоставлен соответствующим воксельным маскам сегментации переломов ребер, аннотированным одним из четырех клинических категорий переломов: перелом-пряжка, перелом с диастазом, перелом без диастаза или сегментарный перелом. Отдельной категорией в данном наборе исследований выделяются сопутствующие находки и переломы, класс которых не может быть обозначен из-за неопределенности или трудностей диагностики.

Предварительно набор данных разделен авторами на три части – обучающая, валидационная и тестовая. Обучающая часть содержит 420 КТ-изображений, валидационная – 80 и тестовая – 120. Маски переломов ребер для тестовой части датасета являются закрытыми и используются для оценки качества алгоритмов, предложенных участниками соревнования RibFrac Challange [21].

Исходные КТ-изображения из набора данных RibFrac были получены с применением различных КТ-сканеров и настроек их работы и имеют различные значения пространственного разрешения и количества аксиальных срезов. Медианный спейсинг (пространственное расстояние между вокселями в миллиметрах) КТ-изображений в датасете равен, пиксельное разрешение одного среза исследования – 512 на 512. Медианное количество аксиальных срезов на одно исследование равно 357.

В работе была использована частичная выборка изображений из набора данных RibFrac, содержащая 250 исследований, повторно проанализированных и исправленных специалистами-медиками. В процессе валидации были скорректированы и дополнены существующие маски сегментации переломов ребер и добавлена информация о величине диастаза в миллиметрах для переломов со смещением. Более того, было произведено изменение меток классификации переломов на три следующих класса: переломы с диастазом, переломы без диастаза и консолидированные переломы. Корректирование исходной разметки набора данных RibFrac обусловлено присутствием неточностей изначальной диагностики патологий ребер и требованиями, предъявляемыми к разрабатываемой системе.

В клинической практике выделяют три основных типа переломов: переломы с диастазом, переломы без диастаза и консолидированные переломы. Переломы без диастаза – это переломы, которые не вызывают смещения костей. Их бывает сложно идентифицировать при

рентгенографии. Часто переломы ребер без смещения можно обнаружить только при последующих осмотрах, когда уже проявились признаки заживления. Переломы с диастазом – это переломы со значительным отклонением в положении двух частей поврежденного ребра. При возникновении таких переломов могут возникнуть травмы окружающих тканей и структур. Величиной диастаза для данной категории переломов называется расстояние максимального расхождения частей ребра или его фрагментов. Консолидированные переломы – это полное или частичное восстановление целостности кости после ее перелома с формированием сначала первичной, а затем и вторичной костной мозоли.

Дальнейший анализ методов и подходов будет осуществляться с использованием описанной выше выборки данных. На рис. 8. приведены гистограммы распределения количества КТ-изображений для отобранный части исследований из набора данных RibFrac по коронарному и сагиттальному пространственным разрешениям, количеству аксиальных срезов и общему количеству переломов на одно исследование. Статистики рассчитаны до преобразований.

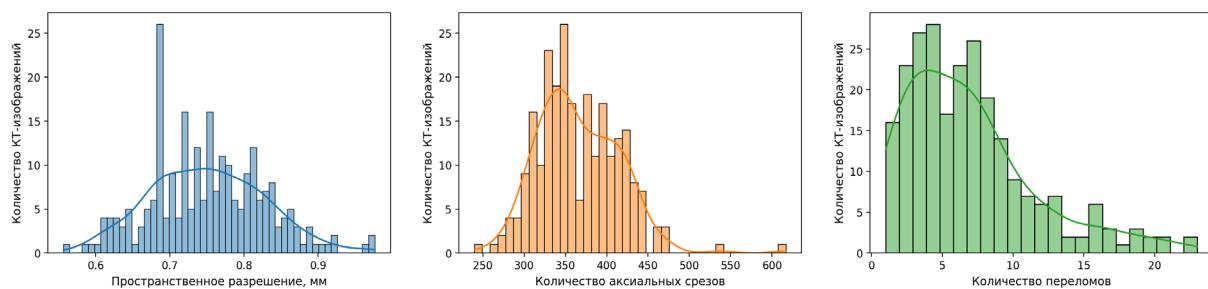


Рис. 7. Гистограммы распределения количества КТ-изображений для отобранный части исследований из набора данных RibFrac.

Для решения задач предварительной обработки данных и исследования моделей локализации, задействующих информацию о структуре ребер, рассматривался датасет RibSeg v2 [22], содержащий воксельные сегментационные маски ребер, разделенные на 24 класса, и информацию о центральных линиях каждого ребра. Разметка в наборе данных RibSeg v2 формировалась на основании исследований датасета RibFrac. Таким образом, каждое КТ-изображение набора данных RibFrac имеет соответствующую разметку ребер из датасета RibSeg v2.

3.1. Формирование отрицательных примеров. Важным аспектом при обучении модели снижения числа ложноположительных предсказаний является генерация отрицательных примеров, представляющих из себя сегменты изображения, не локализующие целевую патологию. Существуют различные способы сэмплирования таких примеров. Некоторые из методов описаны, например, в работе [45].

В качестве основной составляющей числа отрицательных примеров уместно рассматривать ложноположительные предсказания алгоритма детекции переломов ребер на первом шаге. Предсказание разумно считать ложным, если значение IoU (intersection over union) предсказания меньше 0.01 для любого истинного перелома на изображении. В силу того, что ложноположительные предсказания модели детекции почти всегда выделяют область высокой плотности, локализующую какую-либо часть костной ткани (чаще всего ребер), естественным будет включение в число отрицательных примеров случайно выбранных областей ребер, не локализующих никакой из переломов на изображении ($IoU < 0.01$). Размеры выбираемых из изображения областей могут быть определены случайным образом или через набор базовых параллелепипедов, использующихся в процессе обучения детекционной модели. Дополнительно добавляется некоторая доля от общего числа отрицательных примеров случайно выбранных из изображения произвольных регионов, не содержащих переломы.

Формирование отрицательных примеров может выполняться в динамическом режиме в процессе обучения для каждого изображения обучающей выборки. В совокупности с использованием пространственных аугментаций такой метод повышает разнообразие и вариативность набора отрицательных регионов. Динамическое формирование примеров позволяет учитывать ложноположительные предсказания, возникновение которых обусловлено пространственными аугментациями или изменениями интенсивности, в силу того, что модель детекции применяется к каждому преобразованному изображению заново. Однако данный

подход имеет существенный недостаток, заключающийся в низкой скорости обучения модели в силу необходимости обрабатывать все КТ-исследование целиком на каждой итерации обучения с применением предсказания модели локализации переломов.

В качестве альтернативного варианта генерации отрицательных сэмплов предлагается использовать предварительное выделение определенного набора областей изображения. Для обеспечения вариативности формируемого набора примеров рекомендуется вырезать области с учетом добавления дополнительного случайного отступа к границам исходных параллелепипедов. Данный шаг позволит модели классификации учитывать больше информации об окружающих анатомических структурах относительно применения фиксированного размера ограничивающих параллелепипедов, что в совокупности с использованием пространственных аугментаций увеличивает обобщающую способность модели. При таком подходе предсказание модели детекции переломов осуществляется для каждого изображения только один раз, что в теории может снижать вариативность ложноположительных предсказаний, вызываемых трансформациями исходного КТ-исследования.

Пусть $I^{(i)}$ – изображение обучающего набора с номером i , являющееся трехмерной матрицей размерами $W^{(i)} \times H^{(i)} \times D^{(i)}$, $R_{\text{ribs}}^{(i)}$ – сегментационная маска ребер, сопоставленная данному изображению, B_{anchors} – базовый набор ограничивающих параллелепипедов, используемый при обучении модели детекции, $B_{\text{real}}^{(i)}$ – набор ограничивающих параллелепипедов, описывающий действительные локализации переломов ребер на изображении $I^{(i)}$, $B_{\text{pred}}^{(i)}$ – предсказанный детектором набор ограничивающих параллелепипедов переломов ребер для изображения $I^{(i)}$, отсортированный в порядке убывания уверенности модели.

Коэффициенты $\alpha = [\alpha_1, \alpha_2]$, $\alpha_1 + \alpha_2 = 1$ определяются как доли количества формируемых отрицательных примеров, являющихся ложноположительными предсказаниями детектора и случайными элементами в областях ребер, и число N – как общее количество генерируемых отрицательных примеров на одно изображение. Тогда формирование набора отрицательных примеров для обучения модели классификации может быть описано в виде следующего алгоритма:

для изображений $I^{(i)}$ в обучающей выборке

$$\begin{aligned} B &\leftarrow B + \left\{ B_{\text{pred},j}^{(i)} : j \leq \lfloor \pm_1 N \rfloor \right\}, k = 0; \\ \text{до тех пор, пока } k < \max\left(\lceil \alpha_2 N \rceil, N - \lceil B_{\text{pred}}^{(i)} \rceil\right), \\ (x_c, y_c, z_c) &= \mathcal{RS}\left(\{(x, y, z) : (x, y, z) \in R_{\text{ribs}}^{(i)}\}\right), \\ (b_x, b_y, b_z) &= \text{Size}(\mathcal{RS}(B_{\text{anchors}})) + \mathcal{U}_{\{x, y, z\}}[0, 10], \\ \tilde{B} &= \left(x \in \left[x_c - \frac{b_x}{2}, x_c + \frac{b_x}{2} \right], y \in \left[y_c - \frac{b_y}{2}, y_c + \frac{b_y}{2} \right], z \in \left[z_c - \frac{b_z}{2}, z_c + \frac{b_z}{2} \right] \right); \\ \text{если } \max\left\{ \text{IoU}(\tilde{B}, B_j) : B_j \in B_{\text{real}}^{(i)} \right\} &< 0.01, \text{ тогда } B \leftarrow B + \tilde{B}, k \leftarrow k + 1. \end{aligned} \tag{3.1}$$

Здесь $[x]$ – операция целочисленного округления x вниз, $[x]$ – операция целочисленного округления x вверх, \mathcal{RS} – операция взятия случайного элемента множества, Size – операция вычисления размера параллелепипеда по трем осям, результатом которой является трехмерный вектор, $\mathcal{U}_{\{x, y, z\}}[a, b]$ – трехмерный вектор, каждый элемент которого взят из равномерного распределения на заданном отрезке $[a, b]$.

Помимо положительных примеров из числа размеченных специалистами масок переломов, ограниченных параллелепипедами описанным в разд. 1.2 методом, могут быть использованы предсказания детектора, для которых существует разметка с $\text{IoU} < 0.5$. Дополнительное расширение ограничивающих параллелепипедов может быть выполнено по аналогии с методом, применяемым для обработки генерируемых отрицательных регионов. Соотношение общего числа отрицательных и положительных сэмплов является гиперпараметром модели классификации.

4. Экспериментальная часть. Основной акцент в данном исследовании сделан на разработке двухстадийной системы детектирования переломов ребер с помощью КТ-изображений органов грудной клетки (ОГК). В процессе исследования рассматривались следующие задачи:

1. Построение архитектуры обнаружения переломов ребер в терминах решения задачи семантической сегментации.

2. Разработка и определение оптимальной конфигурации модели детекции, использующей заданный набор ограничивающих параллелепипедов (anchor-based):

- подбор оптимальных нейросетевых архитектур, примеряемых в качестве базовой модели и модуля комбинирования признаковых описаний различных пространственных разрешений;
- добавление и конфигурирование дополнительного выхода детекционной модели, использующегося для решения задачи семантической сегментации ребер на этапе обучения;
- определение оптимальной конфигурации гиперпараметров выбора ограничивающих параллелепипедов на этапах обучения и предсказания.

3. Построение модели классификации переломов и уменьшения числа ложноположительных предсказаний, полученных на первом шаге. Выбор оптимальной стратегии формирования отрицательных примеров, применяемых при обучении модели классификации.

4. Подбор оптимальных гиперпараметров всей системы локализации и классификации переломов ребер.

5. Сравнение результатов полученной системы с существующими методами решения задачи сегментации переломов ребер с точки зрения мер качества чувствительности и FROC.

4.1. П р е д о б р а б о т к а и а у г м е н т а ц и я д а н н ы х . Для обучения моделей использовался выделенный и провалидированный врачами-рентгенологами набор из 250 КТ-исследований датасета RibFrac, предварительно обработанный алгоритмом, который описан в разд. 2.1. В качестве обучающего набора были случайным образом отобраны 80% изображений. Оставшиеся 20% КТ-сканов применялись для валидации. Пространственное разрешение исследований данного набора – (1.25, 0.75, 0.75), медианный размер одного изображения в осях, Oz , Oy , Ox равен (392, 277, 284). Предварительно каждое изображение нормировалось в диапазоне от –240 до 1040 HU, что являлось 0.025 и 0.975 квантилями распределения плотности в денситометрических показателях, рассчитанной в областях переломов по всем изображениям. При обучении моделей рассматривались случайные сегменты изображения размерами $192 \times 192 \times 192$ вокселей, содержащие сегментационные маски переломов ребер.

В качестве аугментаций исходного набора данных использовались случайные изменения интенсивности методами сдвига и шкалирования, добавление случайного Гауссова шума и Гауссова сглаживания. Пространственными аугментациями выступали случайные аффинные преобразования с максимальным углом поворота по каждой оси $\pi/15$ и максимальным относительным растяжением (сжатием) 0.15, а также случайное приближение или отдаление с амплитудой в диапазоне [0.9, 1.1]. Для изображений валидационного набора применялось только нормирование плотности по шкале Хаунсфилда.

Для обучающей части набора данных нормировка КТ-изображения по диапазону $HU \in [-240, 1040]$ осуществлялась после применения аугментаций интенсивности. Выделение сегмента размерами $192 \times 192 \times 192$ вокселей происходило после использования пространственных трансформаций. Аналогичные пространственные преобразования применялись для сегментационных масок ребер и переломов ребер. При помощи трансформаций данного вида на КТ-изображениях использовался метод трилинейной интерполяции, на сегментационных масках – интерполяция методом ближайших соседей. Вычисление ограничивающих параллелепипедов для преобразованных в процессе применения аугментаций сегментационных масок производилось с помощью алгоритма, который представлен в разд. 1.2.

4.2. Р е ш е н и е з а д а ч и с е г м е н т а ц и и . Изначальный подход к решению задачи локализации переломов ребер был реализован с помощью методов семантической сегментации. В качестве базовых архитектур рассматривались модели, описанные в разд. 2.2.

Обучение моделей производилось в течение 300 эпох на видеокарте Nvidia RTX 3090 с 24 Гб оперативной памяти. Оптимизация выполнялась с помощью градиентного спуска методом Adam с параметрами $\beta_1 = 0.9$, $\beta_2 = 0.999$ и коэффициентом регуляризации весов $3e - 5$. Скорость обучения $1e - 3$. Обучение производилось с использованием смешанной точности вычислений с плавающей точкой, исходные данные преобразовывались в формат половинной точности. Размер пакета (батча) – 1. Формирование результата на этапе предсказания модели осуществлялось с применением метода скользящего окна, соответствующего размерам регионов изображения, используемых в процессе обучения – вокселей $192 \times 192 \times 192$. Доля линейного перекрытия между различными окнами – 0.625.

Полученные предсказания моделей сегментации обрабатывались в два шага: на первом этапе предсказания корректировались с помощью морфологических операций замыкания и размыкания, на втором этапе удалялись сегментационные маски объемом менее 50 вокселей с целью исключения шумов. В процессе эксперимента были рассмотрены варианты обучения моделей с использованием дополнительного канала, содержащего информацию о вокселях КТ-изображения, которые соответствуют сегментационной маске ребер.

Вычисление мер качества FROC и Recall производилось на уровне значения IoU, равном 0.2. Результаты соответствующих экспериментов приведены в табл. 1.

Таблица 1. Результаты сравнения сегментационных моделей с учетом и без учета использования дополнительного канала с масками сегментации ребер

| Канал для ребер | Модель | FROC | Recall | Dice |
|------------------|------------|--------------|--------------|--------------|
| Не использовался | SegResNet | 0.647 | 0.771 | 0.567 |
| | Swin UNETR | 0.445 | 0.56 | 0.319 |
| | nnU-Net | 0.601 | 0.691 | 0.54 |
| Использовался | SegResNet | 0.683 | 0.795 | 0.594 |
| | Swin UNETR | 0.489 | 0.61 | 0.413 |
| | nnU-Net | 0.624 | 0.725 | 0.553 |

Полученные меры качества детекции были расценены как неудовлетворительные в рамках решения данной задачи. В связи с этим было принято решение изменить подход к формализации задачи и переформулировать ее в терминах решения задачи детекции. Подробно переход описан в разд. 1.2.

4.3. О б у ч е н и е д е т е к ц и о н н о й м о д е л и. Начальным шагом разработки системы обнаружения переломов ребер являлось экспериментальное определение оптимальной архитектуры модели детекции, использующей заданный набор ограничивающих параллелепипедов.

Сопоставление реальных примеров переломов с предсказаниями модели в процессе обучения детектора производилось с помощью алгоритма адаптивного выбора обучающей выборки ATSS (англ. adaptive training sample selection) [29], который представляет собой метод автоматического определения положительных и отрицательных выборок в соответствии со статистическими характеристиками ограничивающих параллелепипедов. Рассматривался вариант алгоритма ATSS с дополнительным требованием на нахождение центра предсказанного параллелепипеда внутри ограничивающего параллелепипеда реального перелома для отнесения объекта к числу положительных примеров.

Определение набора положительных и отрицательных примеров, по которым в дальнейшем будет производиться расчет классификационной и регрессионной функций потерь, производился с помощью стратегии жесткой отрицательной выборки [46]. Размер пакета (батча) на изображение был выбран равным 64, минимальное количество отрицательных примеров, выбираемых случайным образом из 160 отрицательных примеров с наибольшим значением уверенности модели, равен 8, соотношение количества положительных и отрицательных примеров – 0.3.

Вычисление оптимального набора ограничивающих параллелепипедов было выполнено с помощью алгоритма, предложенного авторами nnDetection [47]. Данный алгоритм максимизирует значение IoU между набором заданных параллелепипедов и реальными локализациями переломов ребер по всем изображениям обучающей выборки данных. Базовым набором выбрали список из шести параллелепипедов.

Обучение моделей производилось в течение 250 эпох на видеоускорителе A100 с 80 Гб оперативной памяти. Оптимизация выполнялась с помощью стохастического градиентного спуска SGD (англ. stochastic gradient descent) совместно с использованием накопления момента с коэффициентом при экспоненциальном скользящем среднем $\beta = 0.9$ и коэффициентом регуляризации весов $3e - 5$. Скорость обучения $1e - 3$. Все расчеты производились с применением смешанной точности вычислений с плавающей точкой. Исходные данные предварительно преобразовывались в формат половинной точности *fp16*. Размер пакета (батча) варьировался от 1 до 2 в зависимости от модели.

Формирование результата на этапе предсказания модели осуществлялось с применением метода скользящего окна, соответствующего размерам регионов изображения, используемых в процессе обучения – $192 \times 192 \times 192$ вокселей. Доля линейного перекрытия между различными окнами составляла 0.5.

4.4. О п р е д е л е н и е о п т и м а л ь н о й а р х и т е к т у р ы. На первом этапе выбора детекционной модели были реализованы 3D-модификации архитектур RetinaNet и EfficientDet с использованием в качестве базовых моделей ResNet, SEResNeXt, High-Resolution Net

и EfficientDet (B1-B5) с базовой сетью EfficientNet (B1-B5) соответственно. Сверточные ядра первого слоя в моделях ResNet, SEResNeXt и EfficientNet были заменены на ядра размером $3 \times 3 \times 3$ и шагом $1 \times 1 \times 1$ с целью повышения пространственного разрешения карт признаков. Эмпирически было определено, что данная замена оказывает значимое влияние на качество работы детекционной модели. В качестве модулей агрегации признаков различных пространственных разрешений исследовались FPN и BiFPN с привлечением различного числа выходных слоев с базовых моделей, количество которых варьировалось от 2 до 4. Количество карт признаков, применяемых в выходных модулях классификации и регрессии на ограничивающих параллелепипедах, варьировалось в диапазоне от 3 до 5. В исследовании рассматривались лишь некоторые наиболее осмысленные и относительно различные варианты комбинаций модулей и используемых слоев из возможных. Основной мерой качества, по которой происходил выбор оптимальной архитектуры, являлось значение полноты (Recall) на уровне $IoU = 0.3$ – сокращенно $Recall@(IoU = 0.3)$.

Изначально были зафиксированы значения количества выходных слоев из базовых моделей и из модулей извлечения признаков на величинах 3 и 4 соответственно, анализировались только архитектуры. Данный выбор обусловлен анализом размера карт признаков на этих слоях. Для входного изображения размером $192 \times 192 \times 192$ вокселей размеры карт признаков с выходных слоев 1–4 модулей извлечения признаков равны $96 \times 96 \times 96$, $48 \times 48 \times 48$, $24 \times 24 \times 24$ и $12 \times 12 \times 12$ соответственно. Рассмотрение карт признаков с более низким пространственным разрешением или использование более глубоких слоев базовых моделей на данном этапе исследования не было целесообразно.

Сравнение в данном эксперименте выполнялось с помощью детекционных мер FROC и $Recall@(IoU = 0.3)$, а также мер качества локализации IoU и Dice. Результаты эксперимента приведены в табл. 2.

Таблица 2. Результаты сравнения базовых архитектур и модулей комбинирования карт признаков различного пространственного разрешения

| Модуль агрегации | Базовая модель | FROC | Recall | Dice | IoU |
|------------------|---------------------|--------------|--------------|--------------|--------------|
| FPN | ResNet-50 | 0.815 | 0.887 | 0.687 | 0.523 |
| | SEResNeXt-50 | 0.801 | 0.903 | 0.683 | 0.519 |
| | HRNet-50 | 0.785 | 0.892 | 0.678 | 0.513 |
| BiFPN | ResNet-50 | 0.798 | 0.8933 | 0.694 | 0.531 |
| | SEResNeXt-50 | 0.809 | 0.906 | 0.703 | 0.542 |
| | EfficienNet-B1 | 0.765 | 0.887 | 0.664 | 0.497 |
| | EfficienNet-B3 | 0.771 | 0.9024 | 0.667 | 0.501 |
| | EfficienNet-B5 | 0.733 | 0.854 | 0.657 | 0.489 |

Исследование влияния количества выходных слоев из базовых моделей и количества слоев модуля комбинирования карт признаков различных пространственных разрешений было проведено для сети ResNet с модулем FPN, SEResNeXt с модулем BiFPN и архитектуры EfficientDet-B3 (EfficientNet-B3 с BiFPN). Для блока FPN различие между числом входных и выходных признаков равно 1, так как в качестве дополнительного используется только 1-й слой, получаемый применением операции MaxPooling к последнему выходу базовой модели. Нумерация слоев производилась от выхода с наибольшим пространственным разрешением карт признаков. Сравнение в данном эксперименте проводилось только с точки зрения меры $Recall@(IoU = 0.3)$. Результаты приведены в табл. 3.

4.5. Использование дополнительного выхода сегментации. Второй этап исследования архитектуры детекционной модели включал определение эффекта применения дополнительного выхода модели для решения задачи семантической сегментации ребер. Сравнение производилось для архитектур ResNet-50 с модулем FPN и SEResNeXt-50 с модулем BiFPN. Количества выходных слоев базовой модели и модулей извлечения признаков были выбраны равными 3 и 4 соответственно на основании результатов предыдущего эксперимента. В случае использования модуля BiFPN для сегментационного выхода сети применялись выходы со всех 5 его слоев. Результаты эксперимента приведены в табл. 4.

Таблица 3. Результаты сравнения различного числа выходных слоев из базовых моделей и количества признаковых описаний из модулей FPN и BiFPN по мере $Recall @ (IoU = 0.3)$

| Слои базовой модели, слои модуля комбинирования | | | | | |
|---|---------------|--------|--------|---------------|--------|
| Архитектура | (2, 3) | (2, 4) | (2, 5) | (3, 4) | (3, 5) |
| ResNet + FPN | 0.8902 | — | — | 0.8872 | — |
| SEResNeXt + BiFPN | 0.8971 | 0.8993 | 0.8734 | 0.9063 | 0.8913 |
| EfficientNet + BiFPN | 0.8811 | — | — | 0.9024 | 0.8689 |

Таблица 4. Результаты сравнения архитектур различных конфигураций дополнительного выхода сегментации ребер

| Дополнительный выход | Архитектура | FROC | Recall | Dice | IoU |
|----------------------|-------------|--------------|--------------|--------------|--------------|
| Не используется | ResNet | 0.815 | 0.887 | 0.687 | 0.523 |
| | SEResNeXt | 0.809 | 0.906 | 0.703 | 0.542 |
| Upsampling | ResNet | 0.829 | 0.934 | 0.697 | 0.535 |
| | SEResNeXt | 0.816 | 0.921 | 0.711 | 0.551 |
| U-Net | ResNet | 0.83 | 0.939 | 0.698 | 0.536 |
| | SEResNeXt | 0.818 | 0.927 | 0.709 | 0.549 |
| Attention U-Net | ResNet | 0.839 | 0.944 | 0.701 | 0.54 |
| | SEResNeXt | 0.837 | 0.948 | 0.714 | 0.556 |

Можно видеть, что при добавлении дополнительного выхода для сегментации ребер наблюдается существенное повышение меры полноты детекции переломов. Наилучший результат получается при применении модификации сети Attention U-Net.

4.6. Финальная архитектура детектора. В качестве финальной детекционной модели предлагается использовать anchor-based детектор, основанный на 3D-реализации сети SEResNeXt-50 с модулем извлечения признаков BiFPN и дополнительным сегментационным выходом, представляющим из себя модификацию сети Attention U-Net. В качестве выхода из модели SEResNeXt выбираются карты признаков со слоев 1, 2 и 3 с количеством каналов в слоях 256, 512 и 1024 соответственно. Дополнительные карты признаков для входа в модуль BiFPN получаются путем добавления сверточных блоков на выходе последнего слоя базовой модели. В модулях классификации и регрессии на ограничивающих параллелепипедах на вход подаются карты признаков с 1-го по 4-й слой модуля BiFPN. Модулями, преобразующими карты признаков в логиты классификации и регрессии на ограничивающих параллелепипедах, выступают сверточные блоки, состоящие из 4 слоев. В модуле сегментации ребер используются выходы со всех 5 слоев BiFPN.

4.7. Генерация положительных и отрицательных примеров. В рамках разработки и обучения модели классификации предсказаний переломов ребер, полученных на предыдущем шаге работы алгоритма, рассматривались различные стратегии формирования отрицательных и положительных примеров (разд. 3.1). В эксперименте рассматривались два варианта генерации отрицательных обучающих примеров. В обоих вариантах в качестве положительных примеров были выбраны реальные локализации переломов ребер и корректные предсказания детекционной модели с $IoU > 0.5$. Варианты генерации отрицательных примеров, применяемых в процессе обучения классификатора, могут быть описаны следующим образом:

1) формирование отрицательных примеров только из числа ложноположительных предсказаний модели, т.е. таких, для которых значение IoU с любым истинным параллелепипедом, ограничивающим перелом ребра, составляет менее 0.05;

2) формирование отрицательных примеров в виде комбинации ложноположительных предсказаний детекционной модели, случайных областей ребер, не локализующих никакой

из переломов на изображении (для которых значение IoU с любой истинной локализацией равно 0), а также некоторой доли случайно выбранных областей изображений.

В случае, когда для генерации примеров не использовались предсказания детекционной модели, размер выбираемой из изображения области определялся с помощью алгоритма из разд. 3.1. с применением набора шести предварительно заданных параллелепипедов, применяемых при обучении модели локализации переломов.

4.8. О б у ч е н и е м о д е л и к л а с с и ф и к а ц и и . В качестве классификационной модели использовалась сверточная сеть ResNet-18. Выбор данной архитектуры обусловлен малыми размерами ограничивающих параллелепипедов, по которым производилось выделение регионов изображения. Рассматривались две стратегии обучения модели-классификатора.

1. Динамическая – обучение с помощью пакетов (батчей), формируемых на основании положительных и отрицательных примеров, генерируемых по изображению в моменте обучения на каждой итерации.

2. Статическая – обучение с использованием случайного сэмплирования из заранее сформированного набора положительных и отрицательных примеров по всем изображениям обучающей выборки. На каждое изображение формировалось до 100 отрицательных примеров при применении варианта генерации с помощью случайных областей ребер.

Положительные и отрицательные примеры объединялись в 1 батч размером 16, каждый элемент данного набора дополнялся до максимального размера изображения в батче (с учетом делимости на 16) симметричным нулевым отступом. Соотношение положительных и отрицательных примеров равно 0.3, минимальное число отрицательных примеров в одном батче – 8. В случае, когда количество изначально полученных отрицательных или положительных примеров было недостаточно для удовлетворения поставленных условий, они дополнялись аугментированными вариантами изображений соответствующего класса.

В качестве функции потерь использовалась перекрестная энтропия с весами классов, обратно пропорциональных мощности классов в обучающих примерах. Меры качества, представленные в табл. 5, найдены путем макроусреднения соответствующих мер по всем классам.

Таблица 5. Результаты сравнения методов обучения классификационной модели для различных вариантов формирования положительных и отрицательных примеров

| Метод генерации | Метод генерации | Precision | Recall | F1 |
|------------------------|-----------------|--------------|--------------|--------------|
| Регионы ребер | Динамический | 0.744 | 0.785 | 0.763 |
| | Статический | 0.732 | 0.761 | 0.748 |
| Предсказания детектора | Динамический | — | — | — |
| | Статический | 0.753 | 0.797 | 0.773 |
| Предсказания + ребра | Динамический | — | — | — |
| | Статический | 0.756 | 0.801 | 0.778 |

Применение ложноположительных предсказаний детектора при формировании отрицательных примеров на каждой итерации обучения (динамический метод) не осуществлялось из-за необходимости использования детекционной модели на каждом шаге, что приводило к значительному увеличению времени обучения модели классификации.

По результатам эксперимента наиболее оптимальной стратегией формирования отрицательных примеров для обучения классификатора является применение ложноположительных предсказаний детектора в совокупности со случайными регионами с центром в точках ребер. Меры качества классификации для различных клинических категорий переломов с использованием стратегии смешанной генерации отрицательных примеров во время обучения приведены в табл. 6.

4.9. П о д б о р г и п е р п а р а м е т р о в с и с т е м ы . Выбор гиперпараметров моделей детекции и классификации предполагает определение оптимальных значений порогов (разд. 2.8). В целях ускорения проведения вычислительных экспериментов предварительно производилась генерация предсказаний для граничных значений $T_{conf} = 0.05$, $N_{map} = 10\ 000$, $N_{img} = 1000$ при различных значениях порога для алгоритма non maximum suppression $T_{nms} \in [0.3, 0.4, 0.5, 0.6]$. Дальнейший подбор оптимальных гиперпараметров выполнялся ме-

Таблица 6. Меры качества классификации для различных клинических категорий переломов и ложноположительных предсказаний

| Клиническая категория | Precision | Recall | F1 |
|-----------------------|-----------|--------|-------|
| С диастазом | 0.905 | 0.849 | 0.876 |
| Без диастаза | 0.643 | 0.67 | 0.656 |
| Консолидированный | 0.552 | 0.671 | 0.606 |
| Не перелом | 0.986 | 0.981 | 0.983 |

тодом поиска по сетке с целевой функцией оптимизации FROC. Значение порога $T_{non-frac}$ определялось эвристически. В ходе эксперимента были выбраны следующие значения соответствующих параметров:

$$T_{conf} = 0.83, N_{map} = 1000, T_{nms} = 0.5, N_{img} = 100, T_{non-frac} = 0.4.$$

4.10. Сравнение результатов. Большинство существующих подходов ориентированы на решение задачи сегментации переломов ребер, поэтому сравнение данных методов с предлагаемым подходом с точки зрения мер *IoU* или *Dice* не информативно. Общее качество системы обнаружения переломов наилучшим образом описывается мерами качества FROC, максимальной чувствительностью при заданном минимально допустимом значении *IoU* и средним количеством ложноположительных предсказаний на изображение. Пороги, на основании которых производится усреднение меры FROC, были выбраны равными 0.5, 1, 2, 4 и 8 в соответствии со значениями данных порогов, предлагаемых участникам соревнования RibFrac Challange. Порог по значению *IoU* был выбран на уровне 0.3. Для сравнения авторами [9] предлагается использовать порог $IoU = 0.2$. Более высокое значение порога по *IoU* при решении задачи детекции относительно задачи сегментации обусловлено различиями в объемах перекрытия прямоугольных и произвольных криволинейных сегментационных областей.

Важно отметить, что прямое сравнение методов в данном случае не будет корректным, так как обучение и тестирование производились для различных частей набора данных RibFrac. Более того, разметка данных при разработке FracNet++ [9] и предлагаемого метода дополнительно корректировалась медиками. Наиболее показательным являлось бы сравнение методов на закрытой части датасета RibFrac, однако соревнование на данный момент времени закрыто для участия.

Результаты сравнения предлагаемого метода с алгоритмами сегментации переломов ребер FracNet++ [9] и SA-FracNet [12] с точки зрения детекционных мер качества приведены в табл. 7.

Таблица 7. Результаты сравнения предложенного метода с методами сегментации переломов FracNet++ [8] и SA-FracNet [12] (сравнение было выполнено на различных частях набора данных RibFrac с учетом дополнительного корректирования разметки и не может быть рассмотрено как прямое сравнение методов)

| Архитектура | Detection FROC | Max Sensitivity | Average FP |
|-----------------------------|----------------|-----------------|------------|
| Предлагаемый метод (1 шаг) | 0.809 | 0.948 | 24.3 |
| Предлагаемый метод (2 шага) | 0.837 | 0.935 | 4.7 |
| FracNet++ | 0.83 | 0.898 | 17.73 |
| SA-FracNet | — | 0.926 | — |

Заключение. Предложен и реализован двухстадийный метод решения задачи сегментации переломов ребер, в котором обнаружение и классификация разделены по различным стадиям. На первом этапе производится детектирование локализаций переломов ребер с помощью anchor-based модели, основанной на 3D-модификации нейросети SEResNeXt с модулем комбинирования признаков различных пространственных разрешений BiFPN и дополнительным выходом в виде 3D Attention U-Net, используемым для предсказания семантических масок

ребер на этапе обучения модели. Результатом работы первого шага алгоритма являются предварительные предсказания переломов, локализованные параллелипипедами в трехмерном пространстве КТ-изображения. На втором этапе предсказанные локализации переломов классифицируются с целью определения клинической категории перелома ребра и уменьшения числа ложноположительных предсказаний детектора. Более того, предлагается метод формирования отрицательных примеров для обучения модели классификации, обеспечивающий оптимальную скорость работы и повышающий точность обнаружения ложноположительных предсказаний детекционной модели.

Экспериментально демонстрируется, что разработанная система обеспечивает высокое качество обнаружения переломов ребер на отобранной и дополнительно скорректированной части набора данных RibFrac с максимальным значением чувствительности обнаружения 0.935 при среднем количестве ложноположительных предсказаний на одно изображение, равном 4.7.

СПИСОК ЛИТЕРАТУРЫ

1. Topker M., Ringl H., Lazar M. The Ribs Unfolded – a CT Visualization Algorithm for Fast Detection of Rib Fractures: Effect on Sensitivity and Specificity in Trauma Patients // Eur Radiol. 2015. V. 25. P. 1865–1874.
2. Urbaneja A., De Verbijer J., Formery A. S. et al. Automatic Rib Cage Unfolding with CT Cylindrical Projection Reformat in Polytraumatized Patients for Rib Fracture Detection and Characterization: Feasibility and Clinical Application // European J. Radiology. 2019. V. 110. P. 121–127.
3. Jang Y.C., Lee C.W., Huang C.C. Diagnostic Accuracy for Acute Rib Fractures: A Cross-sectional Study Utilizing Automatic Rib Unfolding and 3D Volume Rendered Reformation // Acad. Radiol. 2024. V. 31. № 4. P. 1538–1547.
4. Erdemir A., Onur M.R., Idilman I. et al. Pros and Cons of Rib Unfolding Software: a Reliability and Reproducibility Study on Trauma Patients // Turkish J. Trauma and Emergency Surgery. 2023. V. 29. № 6. P. 717–723.
5. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation // arXiv preprint. 2015. arXiv:1505.04597.
6. Chen L.C., Zhu Y. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation // Computer Vision – ECCV. 2018. V. 11211. P. 833–851.
7. Kaiming H. Deep Residual Learning for Image Recognition // arXiv preprint. 2015. arXiv:1512.03385.
8. Jin L., Yang J., Kuang K. et al. Deep-Learning-Assisted Detection and Segmentation of Rib Fractures from CT Scans: Development and Validation of FracNet // EBioMedicine. 2020. V. 62. № 12. P. 103106.
9. Yang J., Shi R., Jin. L. et al. Deep Rib Fracture Instance Segmentation and Classification from CT on the RibFrac Challenge // arXiv preprint. 2024. arXiv:2402.09372.
10. Charles R.Q. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space // arXiv preprint. 2017. arXiv:1706.02413.
11. Wang Y., Sun. Y., Liu Z. et al. Dynamic Graph CNN for Learning on Point Clouds // ACM Transactions on Graphics. 2019. V. 38. № 5. P. 146.
12. Cao Z., Xu L., Chen D. et al. A Robust Shape-Aware Rib Fracture Detection and Segmentation Framework With Contrastive Learning // IEEE Transactions on Multimedia. 2023. V. 1. P. 1–8.
13. Wu M., Chai Z., Qian G. et al. Development and Evaluation of a Deep Learning Algorithm for Rib Segmentation and Fracture Detection from Multicenter Chest CT Images // Radiology: Artificial Intelligence. 2021. V. 3. № 9. P.e200248.
14. Isensee F., Wald T., Ulrich C. et al. nnU-Net Revisited: A Call for Rigorous Validation in 3D Medical Image Segmentation // arXiv preprint. 2024. arXiv:2404.09556.
15. Cardoso M.J., Li W., Brown R. et al. MONAI: An Open-source Framework for Deep Learning in Healthcare // arXiv preprint. 2022. arXiv:2211.02701.
16. Hatamizadeh A., Nath V., Tang Y. et al. Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images // arXiv preprint. 2022. arXiv:2201.01266.
17. Myronenko A. 3D MRI Brain Tumor Segmentation Using Autoencoder Regularization // arXiv preprint. 2018. arXiv:1810.11654.
18. Huang Z., Wang H., Deng Z. et al. STU-Net: Scalable and Transferable Medical Image Segmentation Models Empowered by Large-Scale Supervised Pre-training // arXiv preprint. 2023. arXiv:2304.06716.
19. D'Antonoli T.A., Berger L.K., Indrakanti A.K. et al. TotalSegmentator MRI: Sequence-Independent Segmentation of 59 Anatomical Structures in MR Images // arXiv preprint. 2024. arXiv:2405.19492.
20. Jin L., Yang J., Kuang K. et al. Deep-Learning-Assisted Detection and Segmentation of Rib Fractures from CT Scans: Development and Validation of FracNet // EBioMedicine. 2020. V. 62. № 12. P. 103106.
21. Yang J., Shi R., Jin L. et al. Deep Rib Fracture Instance Segmentation and Classification from CT on the RibFrac Challenge // arXiv preprint. 2024. arXiv:2402.09372.
22. Jin L., Gu S., Wei D. et al. RibSeg v2: A Large-scale Benchmark for Rib Labeling and Anatomical Centerline Extraction // IEEE Transactions on Medical Imaging (TMI). 2024. V. 43. № 1. P. 570–581.
23. Yang J., Gu S., Wei D. et al. RibSeg Dataset and Strong Point Cloud Baselines for Rib Segmentation from CT Scans // Intern. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI). Strasbourg, 2021. P. 611–621.
24. Li C., Zia M.Z., Tran Q.H. et al. Deep Supervision with Intermediate Concepts // arXiv preprint. 2018. arXiv:1801.03399.

25. He Z., Nath V., Yang D. et al. SwinUNETR-V2: Stronger Swin Transformers with Stagewise Convolutions for 3D Medical Image Segmentation // Medical Image Computing and Computer Assisted Intervention (MICCAI). Vancouver, 2023. P. 416–426.
26. Drozdzal M., Vorontsov E., Chartrand G. et al. The Importance of Skip Connections in Biomedical Image Segmentation // arXiv preprint. 2016. arXiv:1608.04117.
27. Redmon J., Divvala S., Girshick R. et al. You Only Look Once: Unified, Real-Time Object Detection // IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). Las Vegas, 2016. P. 779–788.
28. Tian Z., Shen C., Chen H. et al. FCOS: Fully Convolutional One-Stage Object Detection // IEEE/CVF Intern. Conf. on Computer Vision (ICCV). Seoul, 2019. P. 9626–9635.
29. Zhang S., Chi C., Yao Y. et al. Bridging the Gap Between Anchor-based and Anchor-free Detection via Adaptive Training Sample Selection // IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). Seattle, 2020. P. 9756–9765.
30. Lin T.Y., Goyal P., Girshick R. et al. Focal Loss for Dense Object Detection // IEEE Intern. Conf. on Computer Vision (ICCV). Venice, 2017. P. 2999–3007.
31. Tan M., Pang R., Quoc V.L. EfficientDet: Scalable and Efficient Object Detection // IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). Seattle, 2020. P. 10778–10787.
32. Hu J., Shen L., Albanie S. et al. Squeeze-and-Excitation Networks // IEEE/CVF Conf. on Computer Vision and Pattern Recognition. Salt Lake City, 2018. P. 7132–7141.
33. Xie S., Girshick R., Dollar P. et al. Aggregated Residual Transformations for Deep Neural Networks // IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). Honolulu, 2017. P. 5987–5995.
34. Wang J., Sun K., Cheng T. et al. Deep High-Resolution Representation Learning for Visual Recognition // arXiv preprint. 2019. arXiv: 1908.07919.
35. Tan M., Le Q.V. EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks // arXiv preprint. 2019. arXiv: 1905.11946.
36. Lin T.Y., Dollar P., Girshick R.B. et al. Feature Pyramid Networks for Object Detection // IEEE Conf. on Computer Vision and Pattern Recognition (CVPR). Honolulu, 2017. P. 936–944.
37. Liu S., Johns E., Davison A.J. End-To-End Multi-Task Learning With Attention // IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). Long Beach, 2019. P. 1871–1880.
38. Zhou Z., Siddiquee M., Tajbakhsh N. et al. UNet++: A Nested U-Net Architecture for Medical Image Segmentation // Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. 2018. V. 11045. P. 3–11.
39. Oktay O., Schlemper J., Folgoc L.L. et al. Attention U-Net: Learning Where to Look for the Pancreas // arXiv preprint. 2018. arXiv: 1804.03999.
40. Ioffe S., Szegedy C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift // arXiv preprint. 2015. arXiv: 1502.03167.
41. Fu C.Y., Shvets M., Berg A.C. RetinaMask: Learning to Predict Masks Improves State-of-the-art Single-Shot Detection for Free // arXiv preprint. 2019. arXiv: 1901.03353.
42. Sudre C.H., Li W., Vercauteren T. et al. Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations // Lecture Notes in Computer Science. 2017. V. 2017. P. 240–248.
43. Wang C., MacGillivray T., Macnaught G. et al. A Two-stage 3D Unet Framework for Multi-class Segmentation on Full Resolution Image // arXiv preprint. 2018. arXiv: 1804.04314.
44. Jeon U., Kim H., Hong H., Wang J.H. Two-stage Meniscus Segmentation Framework Integrating Multiclass Localization Network and Adversarial Learning-based Segmentation Network in Knee MR Images // Medical Imaging 2021: Computer-Aided Diagnosis. 2021. V. 11597. P. 1159714.
45. Jin D., Ma X., Zhang C. et al. Towards Overcoming False Positives in Visual Relationship Detection // arXiv preprint. 2020. arXiv:2012.12510.
46. Shi W., Chen J., Feng F. et al. On the Theories Behind Hard Negative Sampling for Recommendation // Proc. ACM Web Conf. Austin, 2023. P. 812–822.
47. Baumgartner M., Jaeger P.F., Isensee F. et al. NnDetection: A Self-configuring Method for Medical Object Detection // Lecture Notes in Computer Science. 2021. V. 12905. P. 530–539.