



Применение компьютерного зрения для локализации позвонков на срединных сагиттальных срезах компьютерной томографии

А.Р. Теплякова, ORCID: 0000-0001-8612-3850 <anastasija-t23@mail.ru>

Р.В. Шершнева, ORCID: 0009-0000-2324-5893 <rvshershnev@yandex.ru>

*Обнинский институт атомной энергетики — филиал федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский ядерный университет «МИФИ» (ИЯТЭ НИЯУ МИФИ),
Россия, 249039, г. Обнинск, тер. Студгородок, д. 1.*

Аннотация. Автоматизация рутинных операций, связанных с анализом медицинских снимков, представляет собой важную задачу, так как позволяет снизить нагрузку на рентгенологов. Выбор снимков компьютерной томографии, соответствующих уровням конкретных позвонков, для оценки состава тела пациента обычно осуществляется вручную, что требует дополнительных временных затрат. Цель исследования – разработка подхода к решению задачи локализации позвонков на срединных сагиттальных срезах компьютерной томографии для автоматического выбора аксиальных срезов, используемых для оценки состава тела. В основе разработанного подхода лежит применение модели мультиклассовой сегментации, имеющей архитектуру семейства U-Net, и методов компьютерного зрения для предобработки снимков и постобработки масок сегментации. С целью оценки влияния типов входных данных и архитектур моделей на точность сегментации рассмотрено 20 конфигураций подхода. Выявлено, что предлагаемый метод предобработки входных данных, основанный на формировании трёхканальных снимков, повышает точность мультиклассовой сегментации для четырёх архитектур из пяти рассмотренных (Dense U-Net демонстрирует максимальное значение коэффициента схожести Дайса – 0,8858). Также установлено, что предложенный метод расширения обучающей выборки, основанный на пропуске аксиальных срезов при формировании сагиттальных срезов, повышает точность мультиклассовой сегментации для моделей с архитектурами ResU-Net и Dense U-Net. На базе предложенного подхода реализован программный модуль, решающий задачи автоматического определения положений позвонков шейного, грудного и поясничного отделов позвоночника на срединном сагиттальном срезе компьютерной томографии, их визуализации и определения индексов аксиальных срезов, соответствующих центрам тел позвонков. Разработанный модуль интегрирован с программой для визуализации и анализа медицинских файлов DICOM. Разработанный модуль может применяться в качестве вспомогательного инструмента при решении диагностических задач.

Ключевые слова: компьютерное зрение; глубокое обучение; компьютерная томография; сегментация; локализация; предобработка; аугментация; позвонок; архитектура U-Net; медицинские снимки.

Для цитирования: Теплякова А.Р., Шершнева Р.В. Применение компьютерного зрения для локализации позвонков на срединных сагиттальных срезах компьютерной томографии. Труды ИСП РАН, том 37, вып. 4, часть 1, 2025 г., стр. 233–248. DOI: 10.15514/ISPRAS-2025-37(4)-14.

Application of Computer Vision for Localization of Vertebrae on Midsagittal Computed Tomography Slices

A.R. Teplyakova, ORCID: 0000-0001-8612-3850 <anastasija-t23@mail.ru>

R.V. Shershnev, ORCID: 0009-0000-2324-5893 <rvshershnev@yandex.ru>

*Obninsk Institute for Nuclear Power Engineering,
1, Studgorodok territory, Obninsk, 249039, Russia.*

Abstract. Automation of routine operations related to medical image analysis is an important task, as it reduces the workload of radiologists. The selection of computed tomography images corresponding to the levels of specific vertebrae for assessing the patient's body composition is usually done manually, which requires additional time. The purpose is to develop an approach to solving the problem of vertebrae localization on midsagittal computed tomography slices for automatic selection of axial slices used to assess body composition. We developed an approach based on the use of a multiclass segmentation model with the U-Net family architecture and computer vision methods for images preprocessing and segmentation masks postprocessing. In order to assess the impact of input data types and model architectures on segmentation accuracy, we considered 20 approach configurations. We found that the proposed method of preprocessing input data, based on the formation of three-channel images, increases the accuracy of multiclass segmentation for four architectures out of five considered (Dense U-Net demonstrates the maximum Dice similarity coefficient of 0.8858). We also found that the proposed training set augmentation method based on skipping axial slices when forming sagittal slices improves the multiclass segmentation accuracy for models with the ResU-Net and Dense U-Net architectures. Based on the proposed approach, we implemented a software module that solves the problems of automatic determination of the positions of the cervical, thoracic and lumbar vertebrae on the midsagittal computed tomography slice, their visualization and determination of the axial slice indices corresponding to the vertebral body centers. We integrated the developed module with the program for visualization and analysis of DICOM medical files. The developed module can be used as an auxiliary tool in solving diagnostic problems.

Keywords: computer vision; deep learning; computed tomography; segmentation; localization; preprocessing; augmentation; vertebra; U-Net; medical images.

For citation: Teplyakova A.R., Shershnev R.V. Application of Computer Vision for Localization of Vertebrae on Midsagittal Computed Tomography Slices. *Trudy ISP RAN/Proc. ISP RAS*, vol. 37, issue 4, part 1, 2025, pp. 233-248 (in Russian). DOI: 10.15514/ISPRAS-2025-37(4)-14.

1. Введение

Системы поддержки принятия врачебных решений, в основе функционирования которых лежит применение методов компьютерного зрения, способны выступать в качестве источника второго мнения и снижать рутинную нагрузку на рентгенологов при анализе снимков, получаемых методами медицинской визуализации. Одним из этапов решения диагностических задач, связанных с оценкой состава тела по снимкам компьютерной томографии (КТ), является выбор снимков, соответствующих определённым позвонкам. Для оценки скелетно-мышечного индекса (СМИ) чаще всего используется срез, соответствующий центру тела позвонка L3. Однако снимки КТ уровня L3 не всегда доступны. В связи с этим авторы [1] исследуют, в какой степени снимки КТ на уровне других позвонков могут быть полезны для оценки состояния мышечной ткани, и делают вывод, что для СМИ значительны корреляции между уровнями C1–C2, C7–T5, T7–L5 и L3 ($r=0,70-0,93$) соответственно. Авторы подчеркивают, что при отсутствии срезов на уровне позвонка L3 срезы на уровне позвонков грудного и поясничного отделов (T1–T5, T7–L5) можно использовать для оценки количественных и качественных показателей мышечной ткани у пациентов с раком головы и шеи, пищевода, легких и меланомой, однако использование срезов на уровне позвонков шейного отдела (C1–C2, C7) обеспечивает более низкую надежность. Авторы [2] показывают, что оценки СМИ, получаемые с использованием

площади мышечной ткани среза КТ на уровне T4, сопоставимы с оценками на базе измерений на уровне L3 у пациентов с раком прямой кишки. В статье [3] подчёркивается, что в случаях проведения КТ грудной клетки снимки уровня позвонка L3 могут быть недоступны. В связи с этим авторы исследования демонстрируют, что для проведения оценки состава тела у пациентов с раком лёгкого также возможно использование снимка КТ на уровне позвонка L1. Авторы [4] оценивают корреляции индексов массы тела, полученных через измерение для всего тела и через оценку по одному или нескольким срезам КТ, и приходят к выводу о том, что усреднение по нескольким срезам на уровнях L1–L3 при вычислении оценок может повысить значение корреляции.

Для упрощения одного из этапов анализа снимков КТ с целью оценки состава тела актуальной является разработка подхода к решению задачи локализации позвонков шейного, грудного и поясничного отделов позвоночника, делающее возможной автоматизацию выбора аксиальных срезов КТ, соответствующих уровням определенных позвонков. С учётом того, что авторами настоящей работы ранее обучены модели сегментации мышечной ткани на снимках КТ, соответствующих позвонкам грудного и поясничного отделов [5], подход позволит автоматизировать этап отбора снимков, которые могут быть обработаны этими моделями.

2. Обзор существующих подходов

В рамках исследования был проведён обзор существующих подходов к решению задачи автоматической сегментации позвонков, которую необходимо выполнять для их точной локализации. В подходе, описанном в статье [6], аксиальные срезы КТ делятся на неперекрывающиеся патчи, которые передаются модели на базе многоуровневого разреженного автокодировщика (SSAE). Выполняется операция скользящего окна и для каждого патча определяется, относится ли он к областям позвонков или нет. Авторы [7] предлагают полуавтоматический подход для сегментации позвоночника по данным КТ, основанный на геометрическом потоке с предварительным применением анизотропной диффузной фильтрации. Для сегментации отдельного позвонка пользователю необходимо вручную выбирать исходную точку. В статье [8] предлагается подход на базе двухэтапной Dense-U-Net, последовательно выполняющей локализацию (с помощью 2D-Dense-U-Net) и сегментацию (с использованием 3D-Dense-U-Net) позвонков. Авторами [9] предложена модель, работающая в два этапа: сначала с помощью CHASPPRAU-Net (модификация архитектуры U-Net) сегментируется позвоночник, после чего при помощи MRU-Net (мобильная остаточная U-Net) распознаются отдельные позвонки. В [10] предлагается гибридный подход к сегментации позвонков на аксиальных срезах КТ, сочетающий в себе свёрточную нейронную сеть (CNN) и полностью свёрточную сеть (FCN). Предлагаемая в [11] модель глубокого доверия на основе патчей (PaDBN) автоматически выбирает признаки из патчей изображений КТ, после чего измеряет межклассовые различия. Полученные дискриминативные признаки подаются на вход классификатора, который оценивает вероятность принадлежности к области позвонка. Авторы [12] предлагают метод автоматической сегментации поясничных позвонков, состоящий из двух частей: локализации и сегментации позвонков поясничного отдела. За первый этап отвечает U-Net, за второй – трёхмерная XUnet. CNN для сегментации позвонков Verte-Box, описанная в статье [13], использует механизм контроля для улучшения признаков и подавления информации о помехах, а также включает в себя многомасштабный блок свёртки, обеспечивающий возможность комбинирования размеров рецептивных полей.

Все рассмотренные подходы объединяет то, что они решают задачу трёхмерной сегментации позвонков, подразумевающую необходимость значительных вычислительных затрат для локализации определённого позвонка. В рамках настоящего исследования задача локализации и сегментации позвонков рассматривается как промежуточный этап, необходимый для нахождения срезов, представляющих интерес для дальнейшей обработки.

Одним из требований к реализуемому подходу, который выступает в качестве основы модуля, расширяющего функционал разработанного ранее средства просмотра файлов DICOM [14], является возможность его выполнения как на персональных компьютерах, так и в облачной среде. При первом варианте необходимо учитывать возможные ограничения, накладываемые вычислительным устройством пользователя (например, такие как отсутствие графического процессора). При втором варианте стоит придерживаться подхода, в рамках которого минимизируется объём данных, передаваемых по сети. С учётом этих ограничений в данном исследовании предлагается подход, анализирующий только одно изображение, соответствующее срединному сагитальному срезу КТ. Преимущество такого подхода в том, что он сочетает в себе высокую точность, малые вычислительные и временные затраты, необходимые для предсказания, а в случае облачного развёртывания позволяет распределять нагрузку между серверной и клиентской частями, кратно снижая сетевой трафик (при формировании срединного сагитального среза на стороне клиента нет необходимости передачи на сервер полного набора аксиальных срезов исследования КТ). Повышение точности обеспечивается за счёт применения предлагаемых методов обработки снимков. Первый из них основан на формировании трёхканальных изображений, каналы которых представлены снимками в исходном виде, с адаптивным выравниванием гистограмм с ограниченным усилением контраста (CLAHE, Contrast Limited Adaptive Histogram Equalization) и с применением анизотропной диффузной фильтрации (ADF, Anisotropic Diffusion Filtering). Второй метод применяется для расширения обучающей выборки, основан на пропуске аксиальных срезов при формировании изображений, соответствующих сагитальным срезам.

3. Постановка задачи

При проведении диагностики, которая подразумевает выполнение количественной оценки состава тела человека по отдельным снимкам КТ на уровнях тел определенных позвонков, врачу-рентгенологу сначала необходимо вручную выбирать срезы, соответствующие этим уровням, чтобы дальше выполнять оконтуривание представляющих интерес тканей в ручном режиме или при помощи программного обеспечения для автоматической сегментации. Задача заключается в разработке программного модуля, который извлекает из DICOM файлов со срезами исследования КТ изображение, содержащее срединный сагитальный срез, после чего выполняет сегментацию позвонков на этом изображении с целью их локализации и определения индексов аксиальных срезов, соответствующих выбираемому пользователем позвонкам. В рамках данной задачи могут быть выделены следующие подзадачи:

- поиск данных для обучения моделей сегментации позвонков и разработки алгоритмов (снимки КТ с мультиклассовыми масками);
- разработка алгоритмов извлечения и предварительной обработки изображений, содержащих срединные сагитальные срезы КТ;
- построение, обучение и тестирование моделей мультиклассовой сегментации позвонков;
- разработка алгоритмов постобработки масок сегментации (необходимо убрать с них отростки позвонков, оставить только тела) и определения индекса искомого аксиального среза;
- разработка алгоритма визуализации полученных масок позвонков для возможности осуществления пользователем визуального контроля корректности выбора среза;
- интеграция разработанного модуля с программным средством для визуализации и анализа DICOM файлов.

4. Описание предлагаемого подхода

Обработка исследования КТ, состоящего из набора файлов DICOM, в рамках предлагаемого подхода реализуется в три этапа.

4.1 Этап 1: Извлечение и предобработка срединного сагиттального среза

Так как результат КТ представляет собой набор аксиальных срезов, хранящихся в отдельных файлах DICOM, для получения массива пикселей, соответствующего срединному сагиттальному срезу, необходимо сначала осуществить циклический обход аксиальных срезов. Массивы пикселей, соответствующие аксиальным срезам, сортируются по значению элемента данных DICOM, идентифицируемого тегом Slice Location (0020, 1041). Формируется трёхмерный массив пикселей, из которого вырезается двумерный массив, соответствующий срединному сагиттальному срезу. Далее это изображение масштабируется по высоте с использованием коэффициента, рассчитанного на базе значений элементов данных DICOM, идентифицируемых тегами Pixel Spacing (0028, 0030) и Slice Thickness (0018, 0050). В зависимости от протяжённости области тела, охватываемой при проведении КТ, соотношения высоты и ширины срединного сагиттального среза могут значительно варьироваться, поэтому для обработки снимков выбрана стратегия заполнения. Пусть изображение имеет размер $W \times H$, при этом $W < H$. Чтобы получить квадратное изображение (будет подаваться на вход моделей сегментации) с применением стратегии заполнения, необходимо создать массив чёрных пикселей размера $H \times H$, после чего разместить в нём исходное изображение, отступив от левого края $0,5 \cdot (H - W)$ пикселей.

Описанная последовательность действий позволяет получить одноканальные изображения в исходном виде. Если обучение или предсказание выполняется на трёхканальных снимках, то в эту последовательность добавляется ещё один шаг – формирование изображений для трёх каналов. Первое из них – исходное, второе – с применением CLAHE с пределом отсечения 2, третье – обработанное методом шумоподавления на базе ADF. Применение шумоподавления на базе ADF позволяет сохранить значимые для решаемой задачи детали исходного изображения, так как этот метод не размывает края [15].

На рис. 1 показан пример тройки изображений, формирующих снимок в трёхканальном виде. Пример сагиттального среза до и после приведения к квадратной форме показан на рис. 2 (для его формирования использовано КТ исследование из набора данных The Cancer Genome Atlas [16]).

Для повышения точности сегментации также предлагается подход к расширению обучающей выборки, учитывающий свойство КТ, влияющее на формирование сагиттальных срезов. Это свойство обуславливает три возможных варианта взаимного расположения соседних аксиальных срезов: они могут перекрываться, располагаться вплотную друг к другу или на расстоянии (всё зависит от значений толщины среза и расстояния между срезами). Для более устойчивой сегментации сагиттальных срезов, полученных для серий КТ, в которых между аксиальными срезами имеются зазоры, применяется аугментация, суть которой заключается в намеренном пропуске ряда аксиальных срезов при формировании дополнительных изображений, соответствующих сагиттальным срезам. Для более наглядной демонстрации различий между изображениями сагиттальных срезов, сформированными на базе одного и того же КТ исследования без пропуска аксиальных срезов и с применением этого подхода, на рис. 3 показаны области этих изображений, ограниченные одинаковыми координатами.

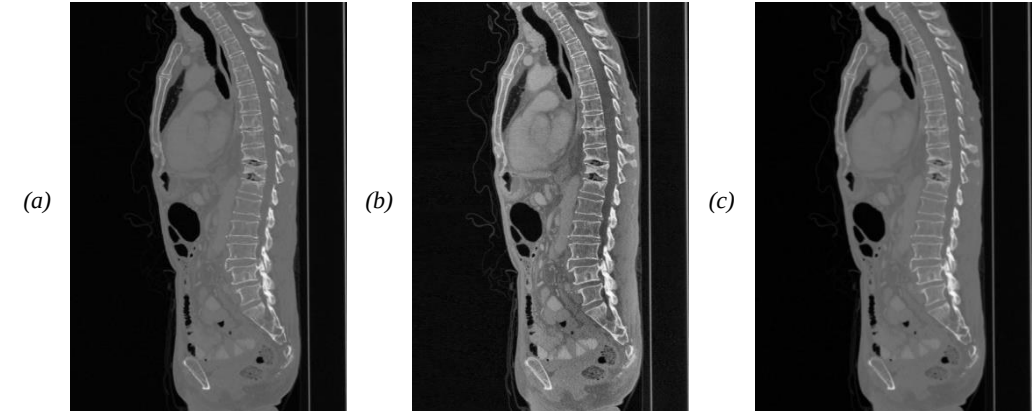


Рис. 1. Составляющие трёхканального изображения: снимок в исходном виде (a), с применением CLAHE (b), с применением ADF (c).
 Fig. 1. Components of a three-channel image: the image in its original form (a), using CLAHE (b), and using ADF (c).

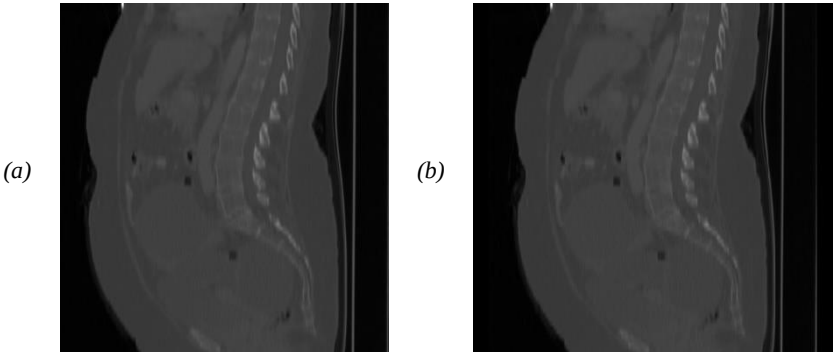


Рис. 2. Срединный сагитальный срез компьютерной томографии в исходном виде (a) и после преобразования к квадратной форме (b).
 Fig. 2. Midsagittal computed tomography image in original form (a) and after transformation to a square shape (b).

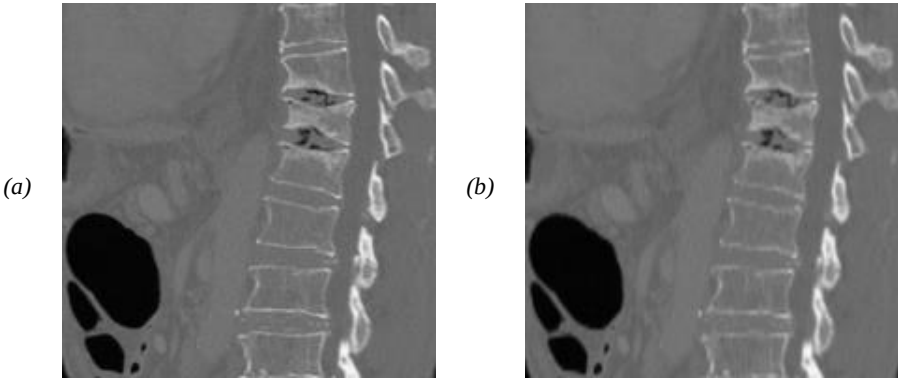


Рис. 3. Сравнение фрагментов изображений срединных сагитальных срезов, сформированных на базе аксиальных срезов одного и того же исследования компьютерной томографии без пропуска (a) и с пропуском (b).
 Fig. 3. Comparison of midsagittal slice image fragments generated from axial slices of the same computed tomography scan without skipping (a) and with skipping (b).

4.2 Этап 2: Сегментация позвонков на срединном сагиттальном срезе

Полученное на первом этапе изображение, содержащее срединный сагиттальный срез, подаётся на вход обученной модели сегментации, имеющей одну из архитектур U-Net [17], Attention U-Net [18], ResU-Net [19], Inception U-Net [20], Dense U-Net [21]. Предсказание модели представляет собой маску, которая содержит метки классов, соответствующих 24 позвонкам, сегментации которых обучена сеть. В силу того, что маски для обучения моделей взяты из набора для полной сегментации позвонков по трёхмерным массивам пикселей, ненулевыми в масках, выдаваемых моделями, являются пиксели, соответствующие не только телам позвонков, но и отросткам. Однако для решения задачи определения положения тела позвонка необходимо найти аксиальный срез, соответствующий центру тела позвонка, поэтому получаемые маски сначала обрабатываются таким образом, чтобы пиксели, соответствующие отросткам, стали нулевыми (чёрными). Первый шаг алгоритма постобработки заключается в переводе маски в бинарный вид и применении к ней эрозии – морфологической операции, которая основана на вычислении локального минимума для заданного ядра и позволяет сузить яркие области изображения.

Далее производится разбиение маски на прямоугольные сегменты, ширина которых равна ширине самой маски, а высота определяется размерами позвонков путём предварительного анализа размеров ограничивающих их контуров. После этого через оценку положений и размеров контуров, попадающих в выделенные сегменты, производится обнуление тех пикселей, которые относятся к отросткам позвонков. На рис. 4 показана маска в исходном виде и после маскирования отростков на ней.

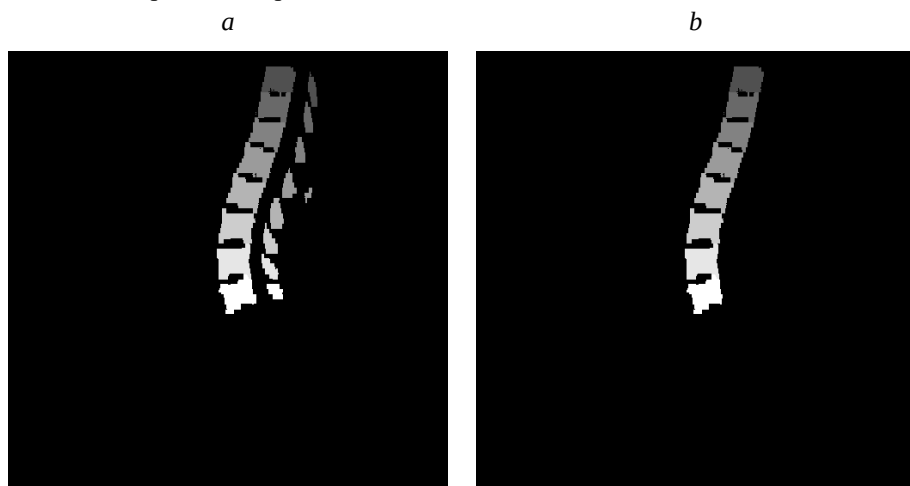


Рис. 4. Маска мультиклассовой сегментации позвонков в исходном виде (a) и после обработки (b).

Fig. 4. Multiclass segmentation mask of vertebrae in the original form (a) and after processing (b).

4.3 Этап 3: Определение положения выбранного позвонка и поиск соответствующего среза

После формирования маски, на которой присутствуют только тела позвонков, выделяется список меток классов позвонков, присутствующих на обработанном срезе. Производится циклический обход списка, при этом на каждой итерации фиксируется положение центра контура, ограничивающего пиксели, которые соответствуют текущей метке. Таким образом, получается список, содержащий метки позвонков и координаты их центров. Далее производится преобразование этих координат с учетом коэффициента, обратного применённому к изображению на этапе 1, и определение индекса аксиального среза,

соответствующего телу центра выбранного позвонка, в массиве. После нахождения нужного аксиального среза формируется изображение, содержащее срединный сагиттальный срез с нанесёнными на него масками, именованиями позвонков, а также горизонтальной линией, соответствующей выбранному аксиальному срезу. Оно может применяться для осуществления визуального контроля корректности выбора среза. Пример такого изображения показан на рис. 5.

Логика функционирования модуля, который разработан на базе описанного подхода, отражена на схеме, показанной на рис. 6.

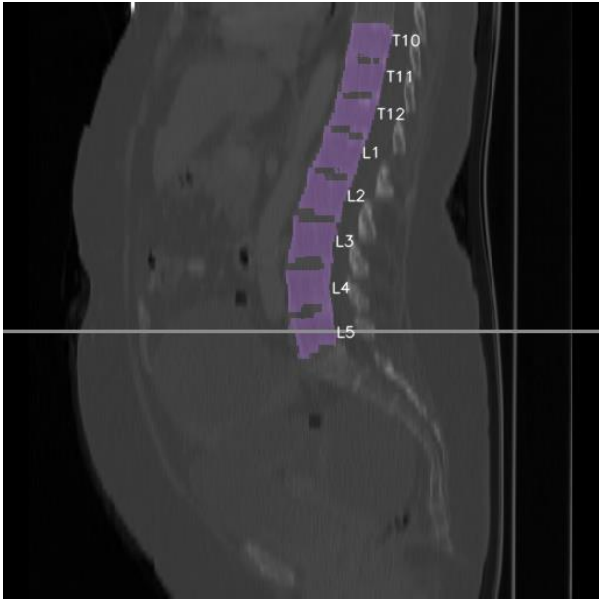


Рис. 5. Пример изображения с нанесенными масками позвонков и линией, соответствующей выбранному аксиальному срезу.
Fig. 5. Example of an image with applied vertebral masks and a line corresponding to the selected axial slice

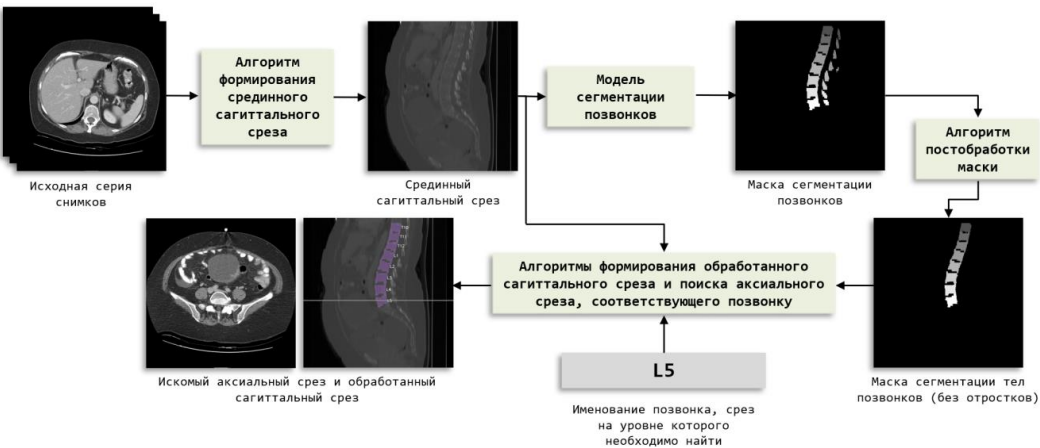


Рис. 6. Логика функционирования разработанного модуля.
Fig. 6. The logic of the functioning of the developed module.

5. Полученные результаты

Программный код разработанного модуля написан на языке Python. В рамках настоящего исследования с использованием библиотеки TensorFlow обучено 20 моделей мультиклассовой сегментации позвонков по срединным сагиттальным срезам КТ, имеющих архитектуры U-Net, Attention U-Net, ResU-Net, Inception U-Net и Dense U-Net. Для каждой из архитектур варьировался подход к формированию обучающих данных, всего их четыре:

- исходные снимки (одноканальный набор);
- трёхканальные снимки (трёхканальный набор);
- исходные снимки с добавлением имитирующих наличие зазора между срезами (одноканальный расширенный набор);
- трёхканальные снимки с добавлением имитирующих наличие зазора между срезами (трёхканальный расширенный набор).

Наборы данных для обучения сформированы на базе набора VerSe [22–24], который содержит 374 исследования 355 пациентов, возраст которых составляет от 42 до 76 лет (использованы данные как VerSe'19, так и VerSe'20). Из каждого исследования выбрано по 15 срезов. Для отбора применялась следующая логика: выбирался срединный сагиттальный срез и по 7 ближайших к нему парасагиттальных срезов с каждой стороны. В тестовую выборку (10% от общего количества срезов) отобраны как срезы в базовом виде, так и срезы с применением подхода к расширению выборки за счёт пропуска аксиальных срезов (исследования, из которых извлекаются срезы, для одноканальных и трёхканальных вариантов тестовых снимков идентичны).

Для оценки моделей сегментации применялся коэффициент схожести Дайса (Dice Similarity Coefficient, DSC), для расчёта которого используется следующая формула:

$$DSC = 2 \cdot \frac{|pred \cap ground_truth|}{|pred| + |ground_truth|}$$

где `pred` обозначает маску, предсказанную моделью, а `ground_truth` – истинную маску. В случае мультиклассовой сегментации маски представляют собой одноканальные массивы, значения которых распределены от 0 до 24 (0 соответствует фону, остальные значения – 24 позвонкам). Для расчёта DSC эти маски предварительно преобразуются таким образом, чтобы каждому классу соответствовал отдельный канал, а значения лежали в пределах от 0 до 1 (One Hot Encoding). Полученные значения DSC приведены в табл. 1.

Среди моделей с архитектурой U-Net (самой лёгкой из рассмотренных) максимальное значение DSC (0,7855) для мультиклассовой сегментации продемонстрировала та, которая обучена на трёхканальном наборе без применения предложенного подхода к аугментации. При этом сравнение двух одноканальных моделей с архитектурой U-Net показывает, что расширение выборки за счёт имитации зазоров между аксиальными срезами повышает DSC. Из четырёх моделей с архитектурой Attention U-Net лучшее значение DSC для мультиклассовой сегментации (0,8857) у версии, обученной на трёхканальном наборе без расширения. Заметно, что подход к расширению выборки не оказал положительного влияния на точность мультиклассовой сегментации моделей с архитектурой Attention U-Net при обучении как на одноканальном, так и на трёхканальном наборах. Для моделей с архитектурами ResU-Net и Dense U-Net наиболее точная мультиклассовая сегментация обеспечивается при обучении на данных расширенного трёхканального набора (DSC 0,8162 и 0,8858). Для моделей Inception U-Net лучшее значение DSC мультиклассовой сегментации (0,8127) достигается при обучении на одноканальном наборе без расширения. Для расчёта метрик бинарной сегментации истинные и предсказанные маски мультиклассовой сегментации были предварительно преобразованы в бинарные. Можно заметить, что для всех пяти рассмотренных архитектур наиболее высокие значения DSC в случае бинарной

сегментации достигаются при использовании трёхканальных данных (для Attention U-Net, ResU-Net, Inception U-Net без расширения, U-Net и Dense U-Net – с ним).

Табл. 1. Значения DSC.

Table 1. DSC values.

Типы выходных масок	Архитектуры моделей сегментации				
	U-Net	Attention U-Net	ResU-Net	Inception U-Net	Dense U-Net
Обученные на одноканальном наборе					
мультиклассовые	0,7526	0,8832	0,7879	0,8127	0,8483
бинарные	0,9505	0,9509	0,9386	0,9426	0,9423
Обученные на расширенном одноканальном наборе					
мультиклассовые	0,7588	0,8720	0,7670	0,7183	0,8758
бинарные	0,9570	0,9494	0,9441	0,9404	0,9476
Обученные на трёхканальном наборе					
мультиклассовые	0,7855	0,8857	0,7949	0,7864	0,8503
бинарные	0,9571	0,9553	0,9501	0,9517	0,9492
Обученные на расширенном трёхканальном наборе					
мультиклассовые	0,7224	0,8566	0,8162	0,7478	0,8858
бинарные	0,9602	0,9525	0,9429	0,9412	0,9527

Для проверки того, насколько точно обученные в рамках работы модели способны сегментировать позвонки на срединных сагиттальных срезах исследований не из набора VerSe, дополнительно проведён эксперимент с данными из открытого набора TotalSegmentator [25]. Для этого авторами настоящей работы отобрано 535 исследований КТ из TotalSegmentator, содержащих более 4 позвонков, и выполнены следующие шаги:

1. Из аксиальных снимков исследований КТ сформированы срединные сагиттальные срезы в одноканальном и трёхканальном видах (применялась та же логика, что и для данных VerSe).
2. Из файлов разметки исследований КТ сформированы соответствующие сагиттальным срезам маски позвонков шейного (C1–C7), грудного (Th1–Th12) и поясничного (L1–L5) отделов позвоночника. Позвонкам на полученных масках назначены те же метки, как при работе с VerSe (например, 19 для Th12, 20 для L1 и т. д.). Так как подходы к разметке данных в рассматриваемых наборах имеют различия (для VerSe характерно отсутствие меток для граничных позвонков исследований КТ, тогда как в TotalSegmentator они размечаются), для минимизации их влияния на значения DSC сформированные маски дополнительно обработаны: на них обнулены пиксели, соответствующие граничным позвонкам, которые не сегментируются обученными на данных VerSe моделями.
3. Всеми обученными моделями выполнена сегментация позвонков на сформированных срединных сагиттальных срезах, произведен расчёт значений DSC (приведены в табл. 2).

Максимальные значения DSC мультиклассовой сегментации для четырёх архитектур из пяти рассмотренных достигаются моделями, обученными на трёхканальных данных (для U-Net, ResU-Net, Dense U-Net – с расширением, для Attention U-Net – без расширения). Из четырёх моделей с архитектурой Inception U-Net лучшее значение DSC мультиклассовой сегментации (0,7978) достигается той, которая обучена на одноканальном наборе без расширения. Самые высокие значения DSC мультиклассовой сегментации продемонстрированы моделями с архитектурами Attention U-Net (0,8554) и Dense U-Net (0,8580). Что касается максимальных значений DSC для бинарной сегментации, для трёх архитектур из пяти рассмотренных они достигаются моделями, обученными на трёхканальных данных (для U-Net – с расширением,

для ResU-Net и Dense U-Net – без расширения), а для двух оставшихся (Attention U-Net и Inception U-Net) – моделями, обученными на одноканальном наборе без расширения.

Табл. 2. Значения DSC на данных TotalSegmentator.

Table 2. DSC values on TotalSegmentator data.

Типы выходных масок	Архитектуры моделей сегментации				
	U-Net	Attention U-Net	ResU-Net	Inception U-Net	Dense U-Net
Обученные на одноканальном наборе					
мультиклассовые	0,7487	0,8396	0,7857	0,7978	0,8396
бинарные	0,9120	0,9133	0,9056	0,9065	0,9072
Обученные на расширенном одноканальном наборе					
мультиклассовые	0,7153	0,8412	0,7955	0,7591	0,8248
бинарные	0,9108	0,9120	0,9062	0,9057	0,9082
Обученные на трёхканальном наборе					
мультиклассовые	0,7439	0,8554	0,7732	0,7741	0,8284
бинарные	0,9114	0,9090	0,9097	0,9003	0,9118
Обученные на расширенном трёхканальном наборе					
мультиклассовые	0,7519	0,8272	0,8006	0,7745	0,8580
бинарные	0,9124	0,9069	0,9040	0,9020	0,9083

Оценка временных затрат на мультиклассовую сегментацию позвонков на срединном сагиттальном срезе исследования КТ, реализуемую моделями каждой из рассмотренных архитектур, была проведена на персональном компьютере с 32 ГБ оперативной памяти, центральным процессором (ЦП) Intel Core i7-11800H и графическим процессором (ГП) NVIDIA GeForce RTX 3080 (16 ГБ видеопамати) для двух конфигураций: вычисления на обоих процессорах и только на центральном процессоре. Полученные оценки приведены в табл. 3. Стоит отметить, что всеми обученными моделями обеспечивается малое время предсказания даже в режиме с использованием только ЦП.

Табл. 3. Среднее время сегментации позвонков на срединном сагиттальном срезе, мс.

Table 3. Average time of vertebral segmentation on the midsagittal slice, ms.

Архитектуры моделей сегментации	Конфигурации персонального компьютера	
	ЦП и ГП	ЦП
U-Net	38,7	65
Attention U-Net	52,3	325
ResU-Net	43,2	130
Inception U-Net	57,1	459
Dense U-Net	55,8	226

Что касается экономии, обеспечиваемой при применении для поиска нужного аксиального среза только срединного сагиттального среза, а не полного массива срезов, то она зависит от того, сколько аксиальных срезов содержит исследование и какой подход используется для их анализа. Для примера дополнительно проведён следующий эксперимент: обучена модель бинарной сегментации позвонков с архитектурой Dense U-Net на аксиальных срезах (размер изображения 256×256), реализована сегментация тестовых исследований. Среднее время сегментации отдельного исследования такой моделью составило 22 с при использовании ЦП и ГП (минимальное – 9 с, максимальное – 43 с), в режиме использования только ЦП – 106 с (минимальное – 41 с, максимальное – 216 с). При этом значение DSC, рассчитанное для масок сегментации срединных сагиттальных и ближайших к ним парасагиттальных срезов, извлеченных из полученных трёхмерных массивов масок исследований тестовой выборки, составило 0,9457. Сравнение соизмеримых по сложности моделей сегментации аксиальных

срезов и срединного сагиттального среза показало, что предпочтительным вариантом для решения задачи поиска нужного позвонка выглядит использование моделей, сегментирующих именно сагиттальные срезы.

Модели с архитектурами Dense U-Net и Attention U-Net, обученные на наборах трёхканальных снимков (у Dense U-Net это версия набора с применением предложенного подхода к расширению, у Attention U-Net – без его применения), показали близкие значения DSC для мультиклассовой сегментации, при этом значение DSC для бинарной сегментации выше у Attention U-Net. Это позволяет сделать вывод о том, что обе эти модели могут быть выбраны в качестве лучших из обученных в рамках настоящей работы. С учётом того, что модель с архитектурой Dense U-Net, обученная на наборе трёхканальных снимков с применением предложенного подхода к расширению, обеспечивает более быструю сегментацию при отсутствии ГП, чем модель с архитектурой Attention U-Net, она выбрана для сегментации позвонков на срединном сагиттальном срезе в разработанном модуле (Этап 2). В связи с этим в процедуру формирования данного среза (Этап 1) добавляется шаг, необходимый для преобразования снимка к трёхканальному виду.

Разработанный модуль интегрирован с программой для визуализации и анализа медицинских файлов DICOM. Через интерфейс программы пользователь открывает директорию с файлами DICOM исследования КТ, выбирает из выпадающего списка позвонки, аксиальный срез на уровне которого необходимо найти, после чего нажимает на кнопку «Найти». Модуль выполняет обработку, на интерфейс выводятся искомый аксиальный срез и обработанный срединный сагиттальный срез. Найденный аксиальный срез также может быть обработан сервисом сегментации мышечной ткани, интегрированным с программой визуализации. Модели, применяемые этим сервисом, обучены в программе для обучения моделей сегментации мышечной ткани по снимкам компьютерной томографии [26] и могут сегментировать мышечную ткань на срезах, соответствующих уровням позвонков Th1–Th12, L1–L5 [5]. Отображение результатов поиска нужного среза и его сегментации на интерфейсе программы-просмотрщика показано на рис. 7.

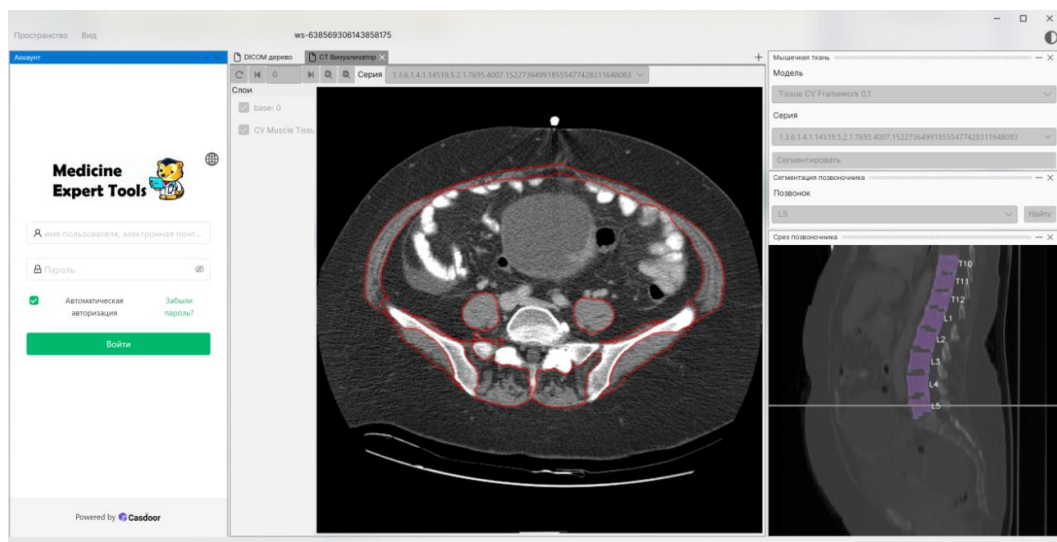


Рис. 7. Пример отображения результатов работы модуля в интерфейсе программы для визуализации и анализа медицинских файлов DICOM.

Fig. 7. An example of displaying the results of the module in the interface of the program for visualization and analysis of DICOM medical files.

6. Заключение

Разработан подход к решению задачи локализации позвонков на срединных сагиттальных срезах компьютерной томографии. На базе разработанного подхода реализован программный модуль, который анализирует серию снимков КТ, выполняет мультиклассовую сегментацию позвонков по формируемому срединному сагиттальному срезу, определяет аксиальный срез, соответствующий уровню тела выбранного пользователем позвонка. Модуль позволяет уменьшить рутинную нагрузку на рентгенологов во время отбора снимков для дальнейшего анализа. Применение предложенного трёхканального подхода к формированию входных данных для моделей мультиклассовой сегментации позвонков позволило повысить точность сегментации для моделей четырёх архитектур из пяти рассмотренных. Предложенный метод расширения обучающей выборки за счёт пропуска аксиальных срезов оказал положительное влияние на точность моделей с архитектурами ResU-Net и Dense U-Net. Для сегментации позвонков на срединном сагиттальном срезе в разработанном модуле выбрана модель с архитектурой Dense U-Net, обученная на наборе трёхканальных снимков с применением предложенного подхода к расширению выборки. Маски сегментации позвонков, выдаваемые моделью, обрабатываются, после чего используются для поиска нужных аксиальных срезов, а также для формирования обработанных изображений, содержащих срединные сагиттальные срезы. Такие изображения передаются на интерфейс программы для визуализации и анализа медицинских файлов DICOM, с которой интегрирован разработанный модуль, и могут использоваться для контроля результатов его работы. Также на интерфейс программы визуализации выводятся искомые аксиальные срезы, которые могут обрабатываться другими модулями, интегрированными с ней. В рамках дальнейшего развития исследования планируется разрабатывать программные модули для количественного и качественного анализа различных тканей и органов по снимкам КТ, интегрируемые с программой-просмотрщиком.

Список литературы / References

- [1]. Van den Broeck J., Sealy M.J., Brussaard C., Kooijman J., Jager-Wittenaar H., Scafoglieri A. The correlation of muscle quantity and quality between all vertebra levels and level L3, measured with CT: An exploratory study. *Frontiers in Nutrition*, 2023, vol. 10, pp. 1148809. DOI: 10.3389/fnut.2023.1148809.
- [2]. Arayne A.A., Gartrell R., Qiao J., Baird P.N., Yeung J.M. Comparison of CT derived body composition at the thoracic T4 and T12 with lumbar L3 vertebral levels and their utility in patients with rectal cancer. *BMC Cancer*, 2023, vol. 23, no. 1, pp. 56. DOI: 10.1186/s12885-023-10522-0.
- [3]. Recio-Boiles A., Galeas J., Goldwasser B., Sánchez K., Man L., Gentzler R., Gildersleeve J., Hollen P., Gralla R. Enhancing evaluation of sarcopenia in patients with non-small cell lung cancer (NSCLC) by assessing skeletal muscle index (SMI) at the first lumbar (L1) level on routine chest computed tomography (CT). *Supportive Care in Cancer*, 2018, vol. 26, no. 7, pp. 2353-2359. DOI: 10.1007/s00520-018-4051-2.
- [4]. Hong J.H., Hong H., Choi Y.R., Kim D.H., Kim J.Y., Yoon J.H., Yoon S.H. CT analysis of thoracolumbar body composition for estimating whole-body composition. *Insights Imaging*, 2023, vol. 14, no. 1, pp. 69. DOI: 10.1186/s13244-023-01402-z.
- [5]. Теплякова А.Р., Шершнев Р.В., Старков С.О. Метод сегментации мышечной ткани на снимках компьютерной томографии на базе предобработанных трехканальных изображений. *Научно-технический вестник информационных технологий, механики и оптики*, том 24, вып. 4, 2024 г., стр. 661-664. DOI: 10.17586/2226-1494-2024-24-4-661-664. / Teplyakova A.R., Shershnev R.V., Starkov S.O. Method of muscle tissue segmentation in computed tomography images based on preprocessed three-channel images. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 4, pp. 661-664 (In Russian). DOI: 10.17586/2226-1494-2024-24-4-661-664.
- [6]. Qadri S.F., Lin H., Shen L., Ahmad M., Qadri S., Khan S., Khan M., Zareen S.S., Akbar M.A., Heyat M.B.B., Qamar S. CT-Based Automatic Spine Segmentation Using Patch-Based Deep Learning. *International Journal of Intelligent Systems*, 2023, vol. 2023, pp. 2345835. DOI: 10.1155/2023/2345835.
- [7]. Khandelwal P., Collins L., Siddiqi K. Spine and Individual Vertebrae Segmentation in Computed Tomography Images Using Geometric Flows and Shape Priors. *Frontiers in Computer Science*, 2021, vol. 3. DOI: 10.3389/fcomp.2021.592296.

- [8]. Cheng P., Yang Y., Yu H., He Y. Automatic vertebrae localization and segmentation in CT with a two-stage Dense-U-Net. *Scientific Reports*, 2021, vol. 11, no. 1. pp. 22156. DOI: 10.1038/s41598-021-01296-18.
- [9]. Saeed M.U., Dikaos N., Dastgir A., Ali G., Hamid M., Hajje F. An Automated Deep Learning Approach for Spine Segmentation and Vertebrae Recognition Using Computed Tomography Images. *Diagnostics*, 2023, vol. 13, no. 1, pp. 2658. DOI: 10.3390/diagnostics13162658.
- [10]. Vania M., Mureja D., Lee D. Automatic spine segmentation from CT images using Convolutional Neural Network via redundant generation of class labels. *Journal of Computational Design and Engineering*, 2019, vol. 6, no. 2, pp. 224-232. DOI: 10.1016/j.jcde.2018.05.002.
- [11]. Qadri S.F., Ai D., Hu G., Ahmad M., Huang Y., Wang Y., Yang J. Automatic Deep Feature Learning via Patch-Based Deep Belief Network for Vertebrae Segmentation in CT Images. *Applied Sciences*, 2019, vol. 9, no. 1, pp. 69. DOI: 10.3390/app9010069.
- [12]. Lu H., Li M., Yu K., Zhang Y., Yu L. Lumbar spine segmentation method based on deep learning. *Journal of Applied Clinical Medical Physics*, 2023, vol. 24, no. 1, pp. e13996. DOI: 10.1002/acm2.13996.
- [13]. Li B., Liu C., Wu S., Li G. Verte-Box: A Novel Convolutional Neural Network for Fully Automatic Segmentation of Vertebrae in CT Image. *Tomography*, 2022, vol. 8, no. 1, pp. 45-58. DOI: 10.3390/tomography8010005.
- [14]. Свидетельство о государственной регистрации программы для ЭВМ. Программа для визуализации и анализа медицинских файлов DICOM, Шершнеv Р. В. (RU), Теплякова А. Р. (RU). № RU2024617422; заявл. 23.03.2024; опубли. 02.04.2024, Бюл. №4. / Shershnev R.V., Teplyakova A.R. Program for visualization and analysis of DICOM medical files. Certificate of state registration of a computer program RU2024617422, registered 23.03.2024, published 02.04.2024 (in Russian).
- [15]. Kaur R., Juneja M., Mandal A. K. A comprehensive review of denoising techniques for abdominal CT images. *Multimedia Tools and Applications*, 2018, vol. 77, no. 17, pp. 22735-22770. DOI: 10.1007/s11042-017-5500-5.
- [16]. Tomczak K., Czerwińska P., Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemporary Oncology*, 2015, vol. 19, no. 1A, pp. A68-A77. DOI: 10.5114/wo.2014.47136.
- [17]. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, vol. 9351, pp. 234-241. DOI: 10.1007/978-3-319-24574-4_28.
- [18]. Oktay O., Schlemper J., Folgoc L.L., Lee M.J., Heinrich M.P., Misawa K., Mori K., McDonagh S.G., Hammerla N.Y., Kainz B., Glocker B., Rueckert D. Attention U-Net: Learning Where to Look for the Pancreas. *arXiv preprint*, 2018. DOI: 10.48550/arXiv.1804.03999.
- [19]. Diakogiannis F.I., Waldner F., Caccetta P., Wu C. ResUNet-a: a deep learning framework for semantic segmentation of remotely sensed data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2020, vol. 162, pp. 94-114. DOI: 10.1016/j.isprsjprs.2020.01.013.
- [20]. Pun N. S., Agarwal S. Inception U-Net Architecture for Semantic Segmentation to Identify Nuclei in Microscopy Cell Images. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 2020, vol. 16, no. 1, pp. 12. DOI: 10.1145/3376922.
- [21]. Cai S., Tian Y., Lui H., Zeng H., Wu Y., Chen G. Dense-UNet: a novel multiphoton in vivo cellular image segmentation model based on a convolutional neural network. *Quantitative Imaging in Medicine and Surgery*, 2020, vol. 10, no. 6, pp. 1275-1285. DOI: 10.21037/qims-19-1090.
- [22]. Löffler M.T., Sekuboyina A., Jacob A., Grau A.-L., Scharf A., Hussein M.E., Kallweit M., Zimmer C., Baum T., Kirschke J.S. A Vertebral Segmentation Dataset with Fracture Grading. *Radiology: Artificial Intelligence*, 2020, vol. 2, no. 4. pp. e190138. DOI: 10.1148/ryai.2020190138.
- [23]. Liebl H., Schinz D., Sekuboyina A., Malagutti L., Löffler M.T., Bayat A., El Hussein M., Tetteh G., Grau K., Niederreiter E., Baum T., Wiestler B., Menze B., Braren R., Zimmer C., Kirschke J.S. A computed tomography vertebral segmentation dataset with anatomical variations and multi-vendor scanner data. *Scientific Data*, 2021, vol. 8, no. 1, pp. 284. DOI: 10.1038/s41597-021-01060-0.
- [24]. VerSe: A Vertebrae labelling and segmentation benchmark for multi-detector CT images / A. Sekuboyina, M.E. Hussein, A. Bayat et al. // *Med Image Anal*, 2021, vol. 73, pp. 102166. DOI: 10.1016/j.media.2021.102166.
- [25]. Dataset with segmentations of 117 important anatomical structures in 1228 CT images (Version 2.0.1) [Data set] / J. Wasserthal // *Zenodo*. 2023. DOI: 10.5281/zenodo.10047292.
- [26]. Свидетельство о государственной регистрации программы для ЭВМ. Программа для обучения моделей сегментации мышечной ткани по снимкам компьютерной томографии, Теплякова А. Р.

(RU), Шершнев Р. В. (RU). № RU2024612322; заявл. 18.01.2024; опубл. 31.01.2024, Бюл. №2. / Teplyakova A.R., Shershnev R.V. Program for training muscle tissue segmentation models from computed tomography images. Certificate of state registration of a computer program RU2024612322, registered 18.01.2024, published 31.01.2024 (in Russian).

Информация об авторах / Information about authors

Анастасия Романовна ТЕПЛЯКОВА – старший преподаватель отделения интеллектуальных кибернетических систем Обнинского института атомной энергетики НИЯУ МИФИ. Сфера научных интересов: компьютерное зрение, глубокое обучение, медицинская визуализация.

Anastasia Romanovna TEPLYAKOVA – Senior Lecturer, Obninsk Institute for Nuclear Power Engineering. Research interests: computer vision, deep learning, medical visualization.

Роман Владимирович ШЕРШНЕВ – старший преподаватель отделения интеллектуальных кибернетических систем Обнинского института атомной энергетики НИЯУ МИФИ. Сфера научных интересов: архитектура программного обеспечения, системы поддержки принятия решений, групповой анализ решений.

Roman Vladimirovich SHERSHNEV – Senior Lecturer, Obninsk Institute for Nuclear Power Engineering. Research interests: software architecture, decision support systems, group decision analysis.

