

# СЕГМЕНТАЦИЯ ТРЕХМЕРНЫХ ОКТ-ИЗОБРАЖЕНИЙ КОЖИ ЧЕЛОВЕКА С ПОМОЩЬЮ НЕЙРОННЫХ СЕТЕЙ С АРХИТЕКТУРОЙ U-NET

DOI: 10.17691/stm2025.17.1.01

УДК 616.591–073.756.8:004.89

Поступила 20.11.2024 г.



**В.А. Шишкова**, младший научный сотрудник<sup>1</sup>;

**Н.В. Громов**, младший научный сотрудник<sup>2</sup>;

**А.М. Миронычева**, младший научный сотрудник<sup>1</sup>; ассистент кафедры кожно-венерических заболеваний<sup>3</sup>;

**М.Ю. Кириллин**, к.ф.-м.н., старший научный сотрудник<sup>1</sup>

<sup>1</sup>Федеральный исследовательский центр Институт прикладной физики им. А.В. Гапонова-Грехова РАН, ул. Ульянова, 46, Н. Новгород, 603950;

<sup>2</sup>Национальный исследовательский Нижегородский государственный университет им. Н.И. Лобачевского, проспект Гагарина, 23, Н. Новгород, 603022;

<sup>3</sup>Приволжский исследовательский медицинский университет, пл. Минина и Пожарского, 10/1, Н. Новгород, 603005

**Цель исследования** — сравнительный анализ алгоритмов сегментации трехмерных ОКТ-изображений кожи человека с использованием нейронных сетей с архитектурой U-Net при обучении модели на двухмерных и трехмерных данных.

**Материалы и методы.** В работе предложены две архитектуры сетей на основе U-Net для сегментации трехмерных ОКТ-изображений кожи, в качестве входных данных для которых выступают двух- и трехмерные блоки трехмерного изображения. Обучение проводили на ОКТ-изображениях толстой кожи, полученных на 7 здоровых волонтерах. Для обучения изображения были размечены в полуавтоматическом режиме специалистами в области ОКТ и дерматологии. Для оценки качества сегментации использовали коэффициент Серенсена–Дайса, который вычисляли по результатам сегментации изображений, не участвовавших в обучении сетей. Дополнительное тестирование возможностей сетей в определении толщин слоев кожи было проведено на независимом наборе данных от 8 здоровых волонтеров.

**Результаты.** При оценке качества сегментации значения коэффициента Серенсена–Дайса для верхних слоев рогового слоя, упорядоченного рогового слоя, клеточного слоя эпидермиса и дермы составили 0,90; 0,94; 0,89 и 0,99 соответственно — для обучения на двухмерных данных и 0,89; 0,94; 0,87 и 0,98 — для обучения на трехмерных данных. Значения, полученные для дермы, хорошо согласуются с результатами других работ, использующих сети на архитектуре U-Net. Толщины упорядоченного рогового слоя и клеточного слоя эпидермиса составили  $153 \pm 24$  и  $137 \pm 17$  мкм соответственно — при обучении сети на двухмерных данных и  $163 \pm 19$  и  $137 \pm 20$  мкм — при обучении на трехмерных данных.

**Заключение.** Сети на основе архитектуры U-Net позволяют сегментировать слои кожи на ОКТ-изображении с высокой точностью, что делает эти сети перспективными для получения значимой диагностической информации в дерматологии и косметологии, например для оценки толщины слоев кожи.

**Ключевые слова:** оптическая когерентная томография; толстая кожа; роговой слой; эпидермис; сверточные нейронные сети; архитектура U-Net.

**Как цитировать:** Shishkova V.A., Gromov N.V., Mironycheva A.M., Kirillin M.Yu. Segmentation of 3D OCT images of human skin using neural networks with U-Net architecture. *Sovremennye tehnologii v medicine* 2025; 17(1): 6, <https://doi.org/10.17691/stm2025.17.1.01>

Для контактов: Кириллин Михаил Юрьевич, e-mail: mkirillin@yandex.ru

English

## Segmentation of 3D OCT Images of Human Skin Using Neural Networks with U-Net Architecture

V.A. Shishkova, Junior Researcher<sup>1</sup>;

N.V. Gromov, Junior Researcher<sup>2</sup>;

A.M. Mironycheva, Junior Researcher<sup>1</sup>; Assistant, Department of Skin and Venereal Diseases<sup>3</sup>;

M.Yu. Kirillin, PhD, Senior Researcher<sup>1</sup>

<sup>1</sup>A.V. Gaponov-Grekhov Institute of Applied Physics of the Russian Academy of Sciences, 46 Ulyanov St., Nizhny Novgorod, 603950;

<sup>2</sup>National Research Lobachevsky State University of Nizhny Novgorod, 23 Prospekt Gagarina, Nizhny Novgorod, 603022, Russia;

<sup>3</sup>Privolzhsky Research Medical University, 10/1 Minin and Pozharsky Square, Nizhny Novgorod, 603005, Russia

**The aim of the study** is a comparative analysis of algorithms for segmentation of three-dimensional OCT images of human skin using neural networks based on U-Net architecture when training the model on two-dimensional and three-dimensional data.

**Materials and Methods.** Two U-Net-based network architectures for segmentation of 3D OCT skin images are proposed in this work, in which 2D and 3D blocks of 3D images serve as input data. Training was performed on thick skin OCT images acquired from 7 healthy volunteers. For training, the OCT images were semi-automatically segmented by experts in OCT and dermatology. The Sørensen–Dice coefficient, which was calculated from the segmentation results of images that did not participate in the training of the networks, was used to assess the quality of segmentation. Additional testing of the networks' capabilities in determining skin layer thicknesses was performed on an independent dataset from 8 healthy volunteers.

**Results.** In evaluating the segmentation quality, the values of the Sørensen–Dice coefficient for the upper stratum corneum, ordered stratum corneum, epidermal cellular layer, and dermis were 0.90, 0.94, 0.89, and 0.99, respectively, for training on two-dimensional data and 0.89, 0.94, 0.87, and 0.98 for training on three-dimensional data. The values obtained for the dermis are in good agreement with the results of other works using networks based on the U-Net architecture. The thicknesses of the ordered stratum corneum and epidermal cellular layer were  $153 \pm 24$  and  $137 \pm 17$   $\mu\text{m}$ , respectively, when the network was trained on two-dimensional data and  $163 \pm 19$  and  $137 \pm 20$   $\mu\text{m}$  when trained on three-dimensional data.

**Conclusion.** Neural networks based on U-Net architecture allow segmentation of skin layers on OCT images with high accuracy, which makes these networks promising for obtaining valuable diagnostic information in dermatology and cosmetology, e.g., for estimating the thickness of skin layers.

**Key words:** optical coherence tomography; thick skin; stratum corneum; epidermis; convolutional neural networks; U-Net architecture.

### Введение

Оптическая когерентная томография (ОКТ) представляет собой метод оптической неинвазивной визуализации биологических тканей, основанный на принципах низкокогерентной интерферометрии и обладающий пространственным разрешением до единиц микрон [1, 2]. ОКТ получила широкое клиническое применение в офтальмологических задачах, однако обладает потенциалом при неинвазивной визуализации внутренней структуры кожных покровов и слизистых оболочек. Преимущество ОКТ в офтальмологии обусловлено низким рассеянием света в тканях глаза, в то время как более сильное рассеяние в коже и слизистых ограничивает практическую глубину зондирования до величин порядка 1–2 мм [3]. Тем не менее такая глубина зондирования позволяет анализировать основные характеристики структурных слоев кожи: рогового слоя, эпидермиса, дермы.

Подходы к сегментации офтальмологических ОКТ-изображений были предложены ранее (см., например, [4]), в том числе и для трехмерных диагностических изображений [5]. Слабое рассеяние зондирующего излучения в структурных слоях тканей глаза обеспечивает довольно высокий контраст границ слоев на ОКТ-изображениях, что делает задачу сегментации слоев относительно простой. В сильно-рассеивающих средах, к которым относится и кожа, можно выделить несколько характерных особенностей, таких как выраженное затухание ОКТ-сигнала с глубиной зондирования и относительно низкий контраст структурных слоев, сравнимый в ряде случаев с контрастом спекл-структуры ОКТ-изображения [6]. Ручная сегментация двухмерных ОКТ-изображений является ресурсозатратной, а системная сегментация трехмерных ОКТ-изображений без привлечения инструментов, автоматизирующих этот процесс, и вообще выглядит весьма затруднительной. Таким образом, эффективное применение ОКТ в дерматологии

требует разработки инструментов, способных быстро и качественно извлекать необходимую диагностическую информацию из полученных трехмерных изображений кожи. Это будет способствовать более широкому использованию ОКТ в клинической дерматологии.

Попытки автоматизировать сегментацию изображений кожи, полученных с помощью ОКТ, предпринимались давно. Например, в статье 2006 г. [7] группа ученых проводила сегментацию трехмерных изображений кожи с целью определения верхней границы кожи, а также отделения эпидермиса от дермы; затем на изображениях выделялись волосяные луковицы. Предложенный алгоритм позволял оценивать среднюю толщину эпидермиса. В алгоритме использовалась последовательность медианных фильтров, применяемых к поверхностям пиков интенсивностей, с последующей аппроксимацией границы слоя полиномиальной функцией. Таким образом, этот алгоритм задействует параметры, подобранные вручную, эмпирическим путем. В статье [8] был предложен метод сегментации, основанный на SVM-классификации по статистическому распределению спектров. Другим подходом, который применяли несколько исследовательских групп, является использование графов. Например, в работе [9] предложен метод определения верхней границы кожи и нижней границы эпидермиса, который включает в себя несколько этапов: предварительную обработку на основе метода взвешенных наименьших квадратов, детектирование верхней границы кожи с помощью графа и обнаружение границы между эпидермисом и дермой на основе локальной интегральной проекции. Теория графов применялась также для автоматического определения поверхности кожи и границы между эпидермисом и дермой на ОКТ-изображениях кожи в исследованиях [10, 11]. Отметим, что вышеупомянутые работы были основаны на классических методах обработки изображений.

В 2015 г. группой ученых Фрайбургского университета специально для сегментации медицинских изображений была разработана архитектура U-Net [12]. Она была создана с учетом того, что размеры выборки обучающих изображений для медицинских задач могут быть существенно ограничены, а границы выделяемых областей не всегда очевидны из-за возможных шумов на изображениях, полученных с помощью специализированной медицинской техники.

Архитектура U-Net и другие архитектуры сверточных нейронных сетей впоследствии стали активно использоваться в офтальмологии и дерматологии. В статьях [4, 13] рассмотрено применение U-Net для сегментации ОКТ-изображений сетчатки. Однако в прозрачных средах контраст границ значительно выше по сравнению с сильно рассеивающими средами, поэтому эти методы требуют доработ-

ки при их адаптации к задачам сегментации ОКТ-изображений кожи.

В работе [14] 2018 г. сверточную сеть U-Net применяли для определения границы между эпидермисом и дермой. В 2019 г. модифицированная U-Net (добавлена проходная связь в слой энкодера) была использована для анализа ОКТ-изображений тканей лабораторных животных, а именно: для сегментации кожи, подкожно-жирового слоя, фасциально-мышечного слоя и татуировок, используемых в качестве опорных меток [15]. Для сегментации ОКТ-изображений тканей лабораторных животных также применяются и другие архитектуры сверточных сетей, таких как ResNet18 [16] или CE-Net [17], сочетающая в себе ResNet и U-Net.

Следует выделить ряд исследований по применению архитектуры U-Net при разработке алгоритмов сегментации диагностических изображений кожи. В работе [18] представлен подход к сегментации эпидермального слоя вместе с фолликулярными структурами на ОКТ-изображениях кожи здоровых волонтеров с помощью сверточной нейронной сети, основанной на архитектуре U-Net с постобработкой, которая заключается в фильтрации изображений. В статье [19] представлен подход к сегментации изображений кожи, полученных с помощью высокочастотного ультразвука (характеристики изображений, полученных с помощью ОКТ и ультразвуковым методом, схожи) с предварительной предобработкой и последующим применением U-Net. В работе [20] с помощью архитектуры U-Net на ОКТ-изображениях кожи человека выделялись роговой слой, эпидермис и дерма, причем обучение проводилось только на изображениях здоровых участков кожи, а алгоритм применялся также и для обработки изображений кожи с повреждениями, такими как рубцовая ткань от лазерной обработки и опухоль. В рамках эксперимента на лабораторных мышах [21] архитектура U-Net также использовалась для сегментации ОКТ-изображений участков кожи, поврежденных лазерным излучением. В работе [22] была предложена предобученная на ОКТ-изображениях кожи грызунов модель сегментации на основе U-Net для дополнительного обучения на данных для кожи человека. При этом авторы утверждают, что одного двумерного сегментированного изображения из трехмерного объема достаточно для точной сегментации всего трехмерного изображения для одного пациента.

Таким образом, в недавних работах по сегментации ОКТ-изображений акцент делается на использовании архитектуры U-Net или подобных сверточных сетей, что позволяет обосновать предпочтение такой архитектуры при выборе модели для сегментации. Следует отметить, что практически во всех приведенных статьях обучение моделей проводилось на основании разметки двумерных данных (Б-сканов), в том числе и в случаях, когда впоследствии алгоритмы применялись для сегментации трехмерных

данных. Такой анализ выполнялся преимущественно для каждого отдельного Б-скана из всего массива данных. Однако существуют исследования, которые показывают, что использование объемной информации может улучшить результаты предсказаний сегментации [23].

**Цель данной работы** — сравнительный анализ алгоритмов сегментации трехмерных ОКТ-изображений кожи человека с использованием архитектуры U-Net при обучении модели на двумерных и трехмерных данных. Исследование проводилось на выборке трехмерных изображений толстой кожи (локализация — пальцы рук). Оба подхода сравниваются по качеству сегментации изображений в задаче выделения границ четырех структурных слоев: верхние слои рогового слоя, упорядоченный роговой слой, клеточный слой эпидермиса и дерма.

## Материалы и методы

**Система для оптической когерентной томографии.** Для получения изображений кожи человека в исследовании использовали установку ОКТ-1300Е (ИПФ РАН, ООО «Биомедтех», Россия) с центральной длиной волны 1300 нм, оснащенную контактным волоконно-оптическим щупом. Установка позволяет получать трехмерные ОКТ-изображения с аксиальным (по глубине) пространственным разрешением 15 мкм. Выходные данные  $I(x, y, z)$  представляют собой массив  $256 \times 512 \times 512$  элементов (рис. 1), каждый из которых соответствует ОКТ-сигналу из соответствующего вокселя в относительных единицах. Физические размеры визуализируемого объема составляют  $1,2 \times 3,0 \times 3,0$  мм. Типичное изображение толстой кожи человека, полученное с помощью ОКТ-1300Е, представлено на рис. 1.

**Набор данных.** В нашей работе для обучения сверточных сетей использовался набор размеченных ОКТ-изображений толстой кожи, полученных на 7 здоровых волонтерах (возраст от 21 до 45 лет; 3 мужчины, 4 женщины). Все изображения сняты с подушечки дистальной фаланги указательного пальца. Для тестирования возможностей сетей в определении толщин структурных слоев кожи дополнительно был использован набор из 16 изображений толстой кожи, полученных на 8 здоровых волонтерах (возраст от 21 до 45 лет; 3 мужчины, 5 женщин). Исследование одобрено локальным этическим комитетом ПИМУ (протокол №17 от 11 октября 2019 г.). Из-за особенностей устройства оптоволоконного зонда прибора ОКТ текстура изображения в пограничных областях (левая и правая границы Б-скана) может быть искажена по сравнению с центральной частью изображения, поэтому для исследования из каждого трехмер-

ного набора данных в каждом массиве была оставлена центральная часть размером  $256 \times 512 \times 256$  вокселей.

**Первичная сегментация ОКТ-изображений.** Наиболее наглядно слои толстой кожи могут быть определены на двумерном ОКТ-изображении (Б-скане). Типичный ОКТ Б-скан толстой кожи человека представлен на рис. 2. Левая часть — оригинальное ОКТ-изображение, в то время как правая сторона — размеченное изображение, на котором выделены четыре структурных слоя. Особенностью толстой кожи человека является толстый роговой слой, который можно подразделить на два слоя [24] — тонкий верхний слой неупорядоченных чешуек (аналогичный роговому слою тонкой кожи, на ОКТ-изображении выглядит как тонкий слой с высоким уровнем сигнала) и толстый слой упорядоченных чешуек (на ОКТ-изображении выглядит как слой с пониженным уровнем сигнала). Под роговым слоем располагается клеточный слой эпидермиса, для которого типичен более высокий сигнал по сравнению с упорядоченными чешуйками эпидермиса; ниже эпидермиса находится дерма, характеризующаяся более низким уровнем ОКТ-сигнала по сравнению с клеточным слоем эпидермиса. Нижнюю границу дермы на ОКТ-изображении детектировать невозможно, поскольку глубина визуализации используемой ОКТ-системы меньше полной толщины толстой кожи.

Первичная разметка данных для обучения сетей на архитектуре U-Net осуществлялась полуавтоматическим методом, после чего верифицировалась специалистами в области ОКТ и дерматологами. Полуавтоматическая сегментация выполнялась методом, принципиально схожим с предложенным в работе [12], который основан на определении среднего уровня сигнала (mean) и его стандартного отклонения (std) в каждом из слоев  $i=\{2, \dots, 5\}$  ( $i=1$  соответствует пространству внутри щупа над поверхностью кожи на ОКТ-изображении). Для этого в каждом слое предварительно выбиралась область  $M_i$  в форме

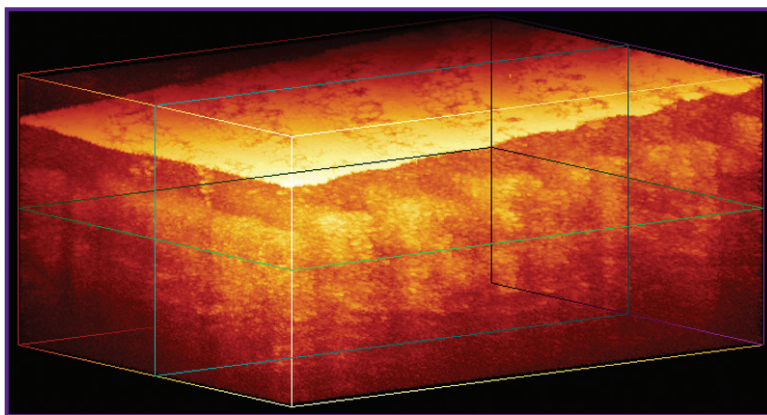
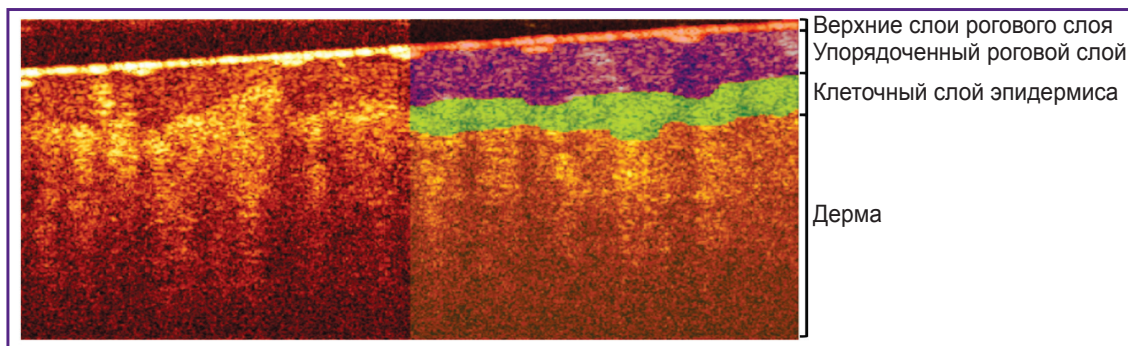


Рис. 1. Визуализация трехмерного ОКТ-изображения кожи человека, полученного с помощью ОКТ-1300Е



**Рис. 2. Пример сегментированного двухмерного ОКТ-изображения толстой кожи**  
Размер изображения составляет 3,0×1,2 мм

кубоида, которая гарантированно принадлежала данному слою, т.е. визуально не включала в себя пограничные воксели. Затем для каждого из слоев на основании статистики распределения сигналов в выбранной области  $M_i$  определялся уровень ОКТ-сигнала, предположительно соответствующий верхней границе слоя:

$$I_i^t = \text{mean}_{M_i}(I) + \alpha_i \text{std}_{M_i}(I),$$

где  $\alpha_i$  — эмпирически подобранный параметр. На основании величины детектировалась поверхность верхней границы определенного слоя  $i$ , к продольной координате которой ( $z$ ) затем применялся медианный фильтр с размером окна  $w_i=11$  вокселей. Полученная поверхность визуально оценивалась специалистами на соответствие реальной границе. В случае неудовлетворительного результата параметры  $\alpha_i$  и  $w_i$  подбирались заново, после чего процедура повторялась. После нахождения поверхности, соответствующей верхней границе слоя  $i=2$  (верхний слой рогового слоя), к области изображения, находящейся под этой границей, был применен медианный фильтр по уровню сигнала для уменьшения спекл-шума. Далее процедура определения границ слоев проходила для нижележащих слоев ( $i>2$ ).

Данный метод имеет значительный недостаток, поскольку требует эмпирического подбора параметров на каждом шаге и визуального контроля качества разметки специалистом. Однако следует отметить, что он существенно выигрывает в скорости по сравнению с полностью ручной разметкой данных при разметке трехмерных массивов данных.

**Применение архитектуры U-Net для сегментации ОКТ-изображений кожи.** Основной задачей нейронной сети в рассматриваемой задаче является сопоставление каждому вокселю ОКТ-изображения метки класса. В рамках постановки выбор ограничивается пятью классами: фон (пространство над поверхностью кожи), верхние слои рогового слоя, упорядоченный роговой слой, клеточный слой эпидермиса и дерма. Соответственно, входными данными для нейронной сети являются трехмерный массив

значений ОКТ-сигнала  $I(x, y, z)$  и соответствующий ему трехмерный массив меток класса  $K(x, y, z)$ , где каждому вокселю сопоставлено целое значение от 1 до 5. В рамках данного исследования проводится сравнение применения двух сверточных нейронных сетей, построенных на архитектуре U-Net и использующих при обучении двумерные или трехмерные данные.

Основная идея архитектуры сети состоит в том, чтобы дополнить обычную сужающуюся часть (contracting path; левую часть буквы “U” в принципиальной схеме сети — энкодер) расширяющейся частью (правая часть буквы “U” — декодер). Эти слои увеличивают разрешение вывода. Для сохранения информации о локализации слои сужающейся и расширяющейся частей сети связаны между собой. При решении задачи малого количества данных в оригинальной статье [11] используется аугментация данных, применяются упругие деформации к имеющимся размеченным изображениям. Это позволяет сети научиться инвариантности к таким деформациям без необходимости видеть данные преобразования в аннотированном наборе изображений.

Для обеспечения достаточного объема обучающей выборки полученные трехмерные массивы данных делятся на блоки, множество которых используется для обучения нейронной сети. В рамках данного исследования сравниваются два вида разделения на блоки. В одном случае в качестве подвыборки выступают Б-сканы размером 256×512×1 воксель (двухмерные данные), в другом — трехмерные блоки размером 256×64×64 вокселя. Выбор элементарных блоков для двух рассматриваемых случаев показан на рис. 3.

**Архитектура сети для обучения на двумерных данных (2D U-Net)** повторяет стандартную архитектуру U-Net, которая описана в оригинальной статье [11], с небольшими изменениями (рис. 4). Сеть состоит из двух частей: энкодера (сужающаяся часть, левая часть) и декодера (расширяющаяся часть, правая часть). Ветвь энкодера имеет 5 ступеней и отвечает за извлечение многомасштаб-

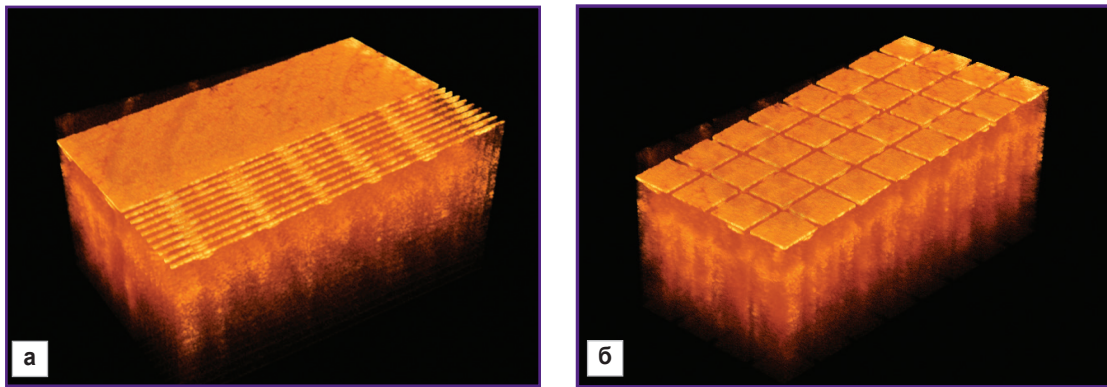


Рис. 3. Разбиение трехмерного ОКТ-изображения на блоки для обучения: (а) двухмерные и (б) трехмерные

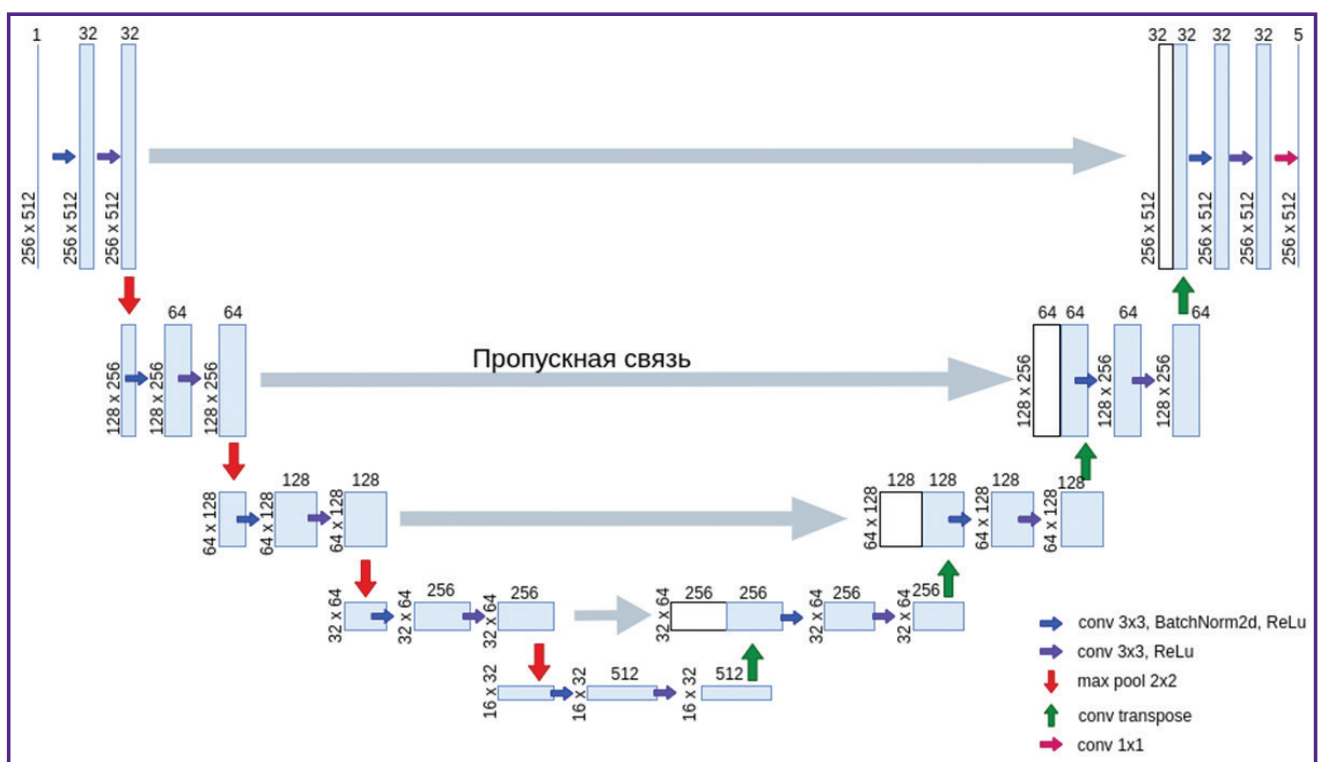


Рис. 4. Архитектуры сети для сегментации ОКТ-изображений с обучением на двухмерных данных (2D U-Net)

ных особенностей входного изображения. Ветвь декодера также включает 5 ступеней и нужна для повышения дискретизации карты признаков, полученных после энкодера, а благодаря пропускным связям в этой ветви удастся восстановить точную локализацию полученных в энкодере признаков. Каждая ступень энкодера состоит из свертки размером  $3 \times 3$  с дополнением исходного изображения (padding) в 1 воксель, за которой следует батч-нормализация и нелинейная функция активации ReLu. Затем выполняется еще одна аналогичная свертка и ReLu-активация. После этого к полученным картам признаков применяется оператор maxpool раз-

мером  $2 \times 2$ , который уменьшает пространственные размеры, тем самым сжимая информацию и позволяя увеличивать число карт признаков. 1, 2, 3, 4 и 5-й этапы кодирования генерируют 32, 64, 128, 256, 512 признаков соответственно. В ветви декодера каждая ступень включает в себя соединение с картами из симметричного уровня энкодера, последующую свертку размером  $3 \times 3$ , батч-нормализацию, нелинейную функцию активации ReLu, еще одну свертку размером  $3 \times 3$  и ReLu-активацию. Для перехода на новую ступень декодера к картам признаков применяется транспонированный сверточный слой с целью повысить дискретизацию карт признаков.

Последний блок декодера состоит из сверточного слоя с ядром размером  $1 \times 1$ . Таким образом, каждому вокселю ставятся в соответствие пять степеней уверенности принадлежности к каждому из классов, после чего вокселю присваивается класс, соответствующий наибольшей степени уверенности.

При использовании модели 2D U-Net для сегментации трехмерного изображения оно разделялось на блоки размерами  $256 \times 512 \times 1$  воксель. Для каждого такого блока проводилась сегментация, после чего полученные данные соединялись в трехмерный массив.

**Архитектура сети для обучения на трехмерных данных (3D U-Net)** повторяет архитектуру, описанную выше, с той разницей, что на вход подается блок размером  $256 \times 64 \times 64$  вокселя (рис. 5) и двумерные операции заменены их трехмерными аналогами (например, двумерные свертки заменены трехмерными). Для сегментирования полного трехмерного ОКТ-изображения модели 3D U-Net изображение разбивалось на блоки с пересечениями размерами  $256 \times 64 \times 64$  вокселя, для каждого из которых независимо проводилось вычисление пяти трехмерных карт степеней уверенности принадлежности к каждому из классов. Для вокселей, принадлежащих области пересечения блоков, степень уверенности принадлежности к каждому из классов определялась как сумма значений для каждого из блоков, взвешенных Гауссовой функцией с радиусом 16 вокселей в

зависимости от расстояния до центра блока. После этой процедуры происходило присвоение каждому вокселю значения класса в соответствии с максимальным значением степени уверенности по аналогии с применением 2D U-Net. Это позволило избавиться от особенностей, возникающих на границах блоков.

**Обучение сетей с архитектурой U-Net.** Доступный набор данных состоял из 7 трехмерных изображений размером  $256 \times 512 \times 256$  вокселей от 7 разных волонтеров. Для того, чтобы избежать переобучения и некорректных значений при оценке качества на зависимых данных, два трехмерных массива из семи были оставлены для тестовой выборки и не использовались в процессе обучения. Далее применялся метод кросс-валидации, а именно leave-one-out кросс-валидация, где объектом считается одно трехмерное изображение, полученное от конкретного пациента. Из построенных пяти моделей была выбрана та, которая при тестировании показывала лучшие результаты. Эта модель уже применялась к тестовой выборке.

В сеть 2D U-Net на вход подавались двумерные изображения размером  $256 \times 512 \times 1$ . Такие изображения были получены из трехмерных изображений как срезы вдоль оси с размерностью 512 вокселей, что обеспечивало 256 блоков для обучения для каждого ОКТ-изображения. В сеть 3D U-Net на вход подавались трехмерные изображения размерами

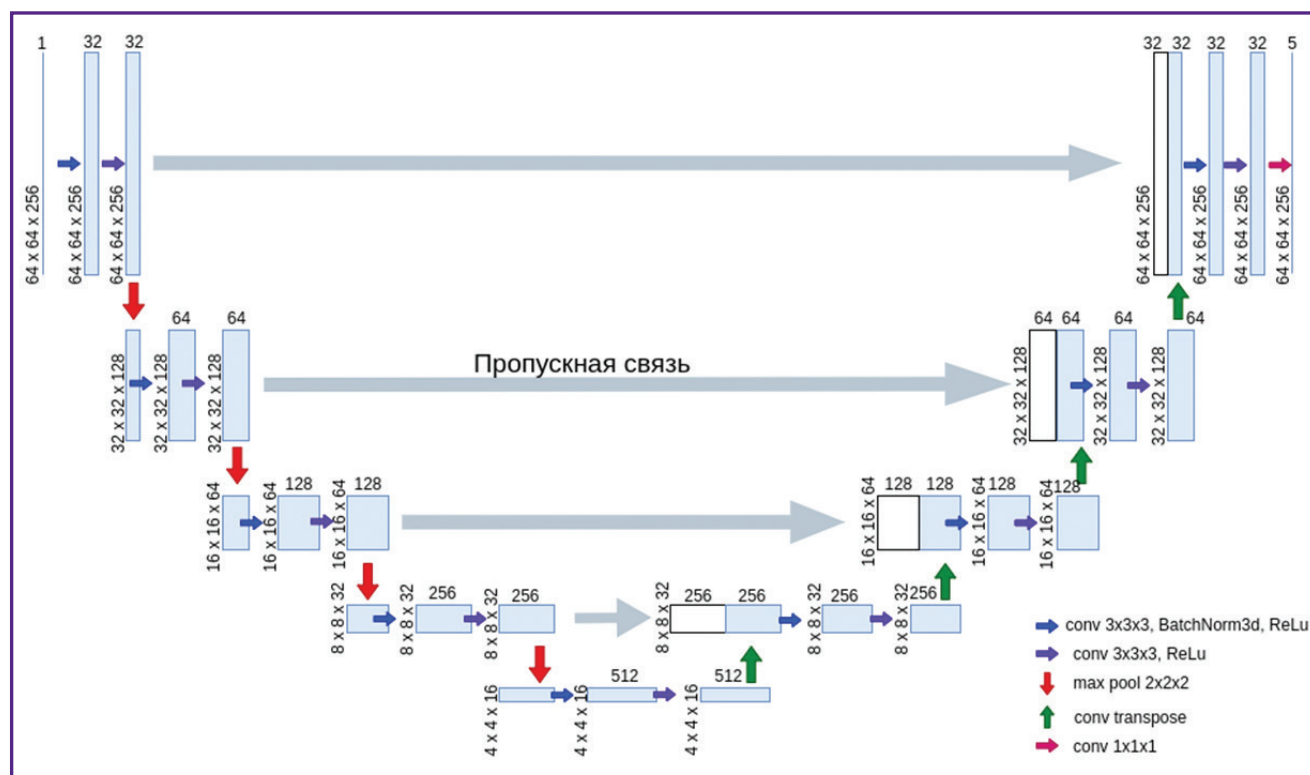


Рис. 5. Архитектуры сети для сегментации ОКТ-изображений с обучением на трехмерных данных (3D U-Net)

256×64×64 вокселя. Для их получения трехмерное изображение разбивалось без пересечений на такие блоки, что обеспечивало 32 блока для обучения для каждого трехмерного ОКТ-изображения.

При обучении использовалась функция потерь — кросс-энтропия, оптимайзер — ADAM (Adaptive Moment Estimation). Это наиболее хорошо зарекомендовавшие себя в данных задачах функция потерь и оптимайзер. Коэффициент скорости обучения был равен 0,001. Коэффициенты, используемые для вычисления скользящих средних градиента и его квадрата, установлены как  $\text{betas}=(0,9; 0,98)$ .

**Процесс тестирования.** При оценке качества использовался коэффициент Серенсена–Дайса (DSC). Пусть  $K^{\text{true}}(x, y, z)$  — истинный массив меток класса, а  $K^{\text{segm}}(x, y, z)$  — массив меток класса, полученных в результате применения нейронной сети. Тогда коэффициент Серенсена–Дайса представляет собой отношение удвоенного количества вокселей определенного класса, совпавших в массивах  $K^{\text{true}}(x, y, z)$  и  $K^{\text{segm}}(x, y, z)$ , к сумме количества вокселей этих классов в каждом из массивов:

$$DSC_i = 2 \frac{n(K^{\text{true}}(x, y, z) = K^{\text{segm}}(x, y, z) = i)}{n(K^{\text{true}}(x, y, z) = i) + n(K^{\text{segm}}(x, y, z) = i)},$$

где  $i$  — значение индекса класса; функция  $n(\dots)$  возвращает количество вокселей, для которых выполняется условие, стоящее в скобках. Коэффициент DSC принимает значения от 0 до 1, где значение 1 соот-

ветствует случаю полного совпадения размеченной и предсказанной масок. Это метрика, которая широко используется для оценки качества работы алгоритмов сегментации. Метрика подсчитывалась для трехмерных масок для обеих моделей.

## Результаты

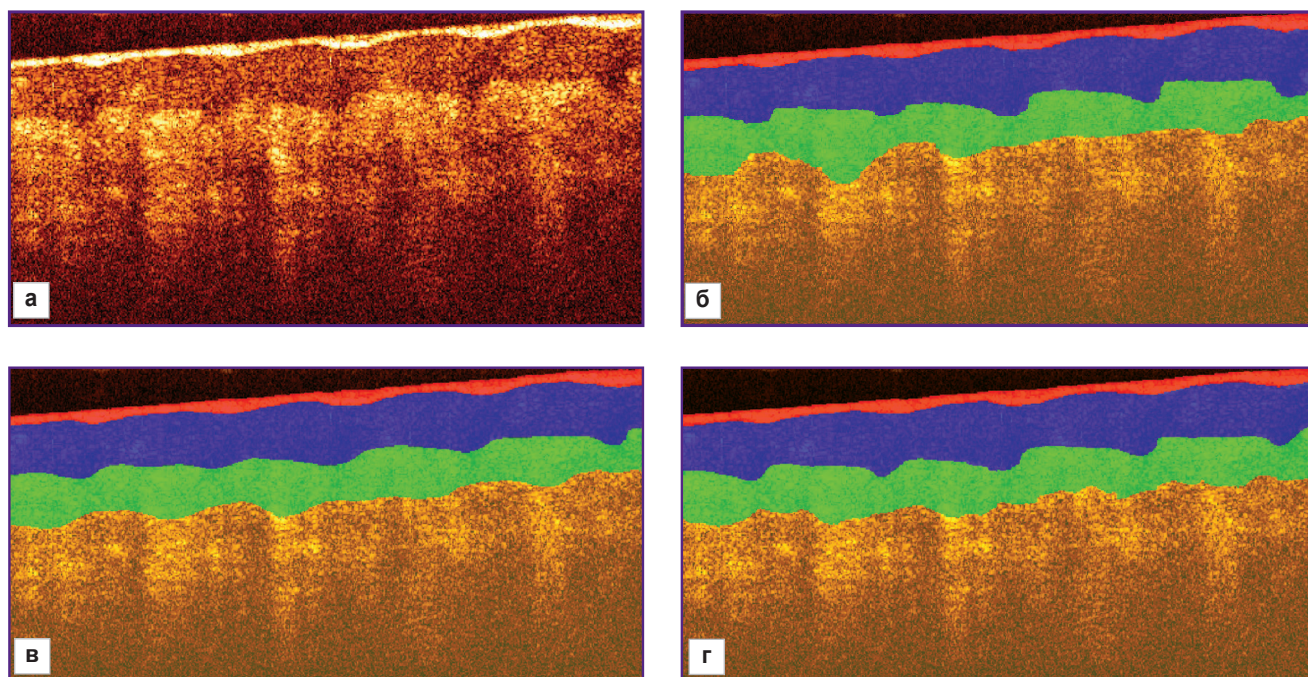
Пример сегментации структурных слоев на двухмерном ОКТ-изображении приведен на рис. 6. На рис. 6, а представлено исходное ОКТ-изображение, на рис. 6, б — сегментация, полученная в полуавтоматическом режиме под контролем специалистов. Рис. 6, в демонстрирует результаты сегментации с помощью 2D U-Net, а рис. 6, г — результат обработки с помощью 3D U-Net.

На рис. 7 представлена визуализация результатов применения 2D U-Net и 3D U-Net ко всему трехмерному массиву.

Для численной характеристики качества сегментации трехмерных ОКТ-изображений сетями 2D U-Net и 3D U-Net она была проведена для двух размеченных ОКТ-изображений, не входивших в обучающую выборку. На рис. 8 показана диаграмма соответствия значений коэффициента DSC для различных слоев изображений из тестовой выборки.

В табл. 1 приведены усредненные значения коэффициентов Серенсена–Дайса для каждого слоя по трехмерным изображениям тестовой выборки.

Дальнейшее тестирование разработанных алгоритмов сегментации происходило на дополнительном



**Рис. 6.** Сегментация двумерного ОКТ-изображения толстой кожи:

а — оригинал; б — полуавтоматическая разметка; в — размечено моделью 2D U-Net; г — размечено моделью 3D U-Net



Рис. 7. Результат сегментации трехмерного ОКТ-изображения с помощью 2D U-Net (а) и 3D U-Net (б)

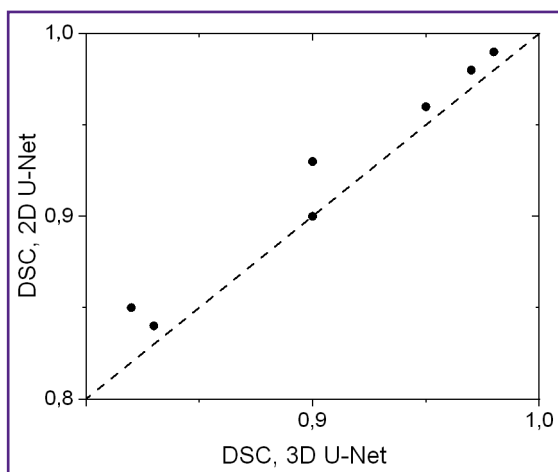
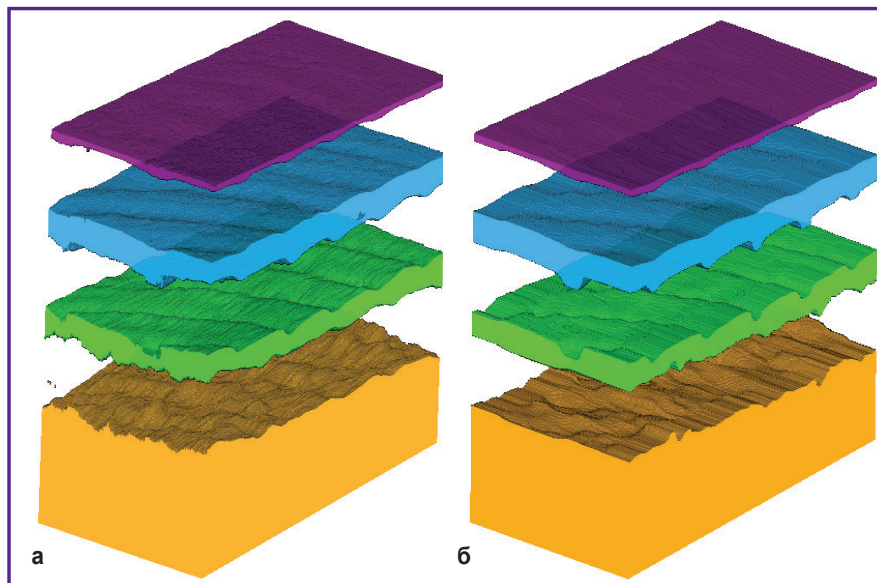


Рис. 8. Сравнение коэффициентов DSC для моделей 2D U-Net и 3D U-Net при сегментации слоев на трехмерных ОКТ-изображениях из тестовой выборки

Таблица 1

Усредненные значения коэффициентов DSC для моделей 2D U-Net и 3D U-Net

Название слоя	2D U-Net	3D U-Net
Верхний слой рогового слоя	0,90	0,89
Упорядоченный роговой слой	0,94	0,94
Клеточный слой эпидермиса	0,89	0,87
Дерма	0,99	0,98

наборе из 16 изображений толстой кожи человека (от 8 волонтеров), не пересекающемся с обучающей и тестовой выборками. Для всех изображений этой

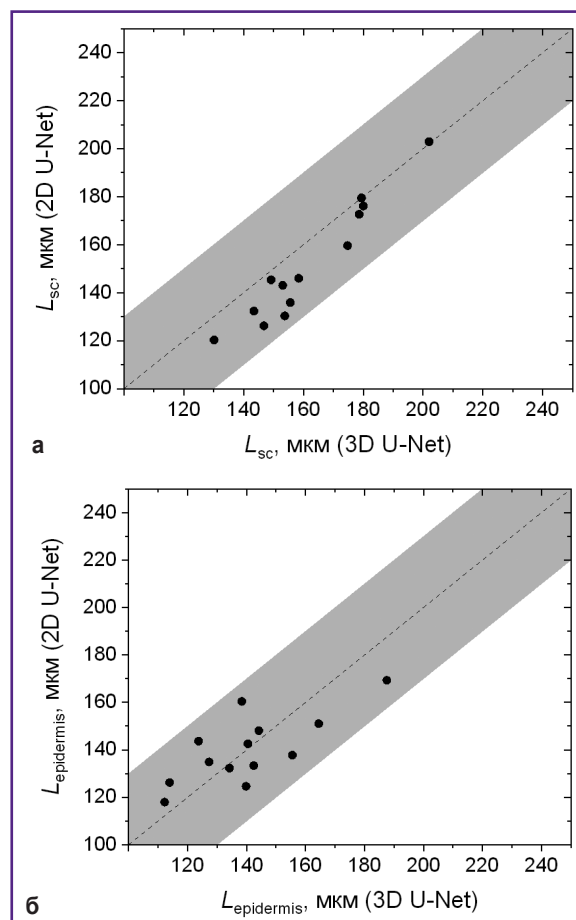


Рис. 9. Сравнение результатов определения толщин упорядоченного рогового слоя (а) и клеточного слоя эпидермиса (б) для моделей 2D U-Net и 3D U-Net при сегментации слоев на трехмерных ОКТ-изображениях из тестовой выборки

Серой полосой показана возможная погрешность определения положения границ при оценке, связанная с продольным пространственным разрешением ОКТ-системы

Таблица 2

Толщины морфологических слоев толстой кожи, полученные из данных сегментации, в сравнении с результатами работы [24]

Название слоя	Толщина, мкм (2D U-Net)	Толщина, мкм (3D U-Net)	Толщина, мкм [24]
Упорядоченный роговой слой	153±24	163±19	175 [147; 210]
Эпидермис	137±17	137±20	119 [112; 126]

Примечание. Данные представлены как  $M \pm SD$  и  $Me [Q1; Q3]$ .

выборки была проведена сегментация, на основе которой определены толщины слоев в предположении, что средний показатель преломления кожи составляет 1,4. Результаты сопоставления оценок толщин упорядоченного рогового слоя и эпидермиса представлены на рис. 9.

Усредненные значения толщин упорядоченного рогового слоя и эпидермиса, полученные на основе результатов сегментации ОКТ-изображений, представлены в табл. 2 в сравнении со значениями, полученными ранее в работе [24].

## Обсуждение

Результаты применения алгоритма сегментации для одного Б-скана (см. рис. 6) показывают, что обе сети обеспечивают сегментацию, согласующуюся с исходной разметкой, и могут использоваться для оценки средней толщины слоев. Однако визуальное применение сети 3D-U-Net (см. рис. 6, *в*) на данном примере выглядит более точным: эта сеть точнее передает особенности границы между роговым слоем и эпидермисом, отражающей наличие папиллярных узоров, в то время как результаты сегментации с помощью 2D U-Net (см. рис. 6, *б*) демонстрируют менее выраженный рельеф границы, ассоциированный с папиллярными узорами. Следует отметить, что результаты, полученные для верхнего слоя рогового слоя, не могут быть интерпретированы однозначно. Граница ткани имеет наиболее высокий уровень сигнала благодаря рассеянию от поверхности, определяемому высоким скачком показателя преломления, а также случайной ориентацией поверхностных чешуек рогового слоя. Более того, соприкосновение поверхности контактного ОКТ-щупа с поверхностью кожи не везде плотное, что ведет к образованию воздушных лакун, границы которых также дают высокий уровень сигнала. Поскольку продольное разрешение ОКТ-системы составляет 15 мкм, эта величина представляет собой погрешность определения положения границы. Однако ее величина сравнима с толщиной верхних слоев рогового слоя, которые на изображении прилегают к поверхности контактного щупа.

Оценка точности сегментации с помощью коэффициента Серенсена–Дайса (см. рис. 8) показала,

что все полученные DSC-значения превышают 0,8, причем значения для сети 2D U-Net для одних и тех же слоев не ниже тех, которые получены в случае использования трехмерных данных при обучении (2D U-Net). Усредненные значения DSC (см. табл. 1) демонстрируют величину не ниже 0,87, причем точность 2D U-Net для всех слоев оказывается не ниже точности 3D U-Net. Таким образом, исходное предположение о том, что трехмерная информация, содержащаяся в трехмерных обучающих блоках, может повысить точность сегментации, не подтвердилось. Предположительно ситуация будет иной в случае наличия объемных макронеоднородностей кожи (опухолей, волосяных фолликулов и т.д.), и применение трехмерной модели улучшит результат предсказания.

При сравнении качества работы разработанного алгоритма с качеством разметки алгоритмов, приведенных в других исследованиях с ОКТ-изображениями, возникает проблема корректного сопоставления результатов, которая вызвана тем, что в каждом исследовании выделяются разные структурные слои на участках кожи разных локализаций. Поскольку морфологически наиболее похожим слоем толстой и тонкой кожи является дерма, мы сравнили полученное DSC-значение для дермы с аналогичным значением других работ, в которых исследовалась тонкая кожа. В работе [20] DSC для дермы составил 0,96, что практически совпадает с результатами работы [18], где DSC для дермы равен  $0,96 \pm 0,01$ . Следует отметить, что обе эти работы использовали нейронные сети, основанные на архитектуре U-Net. Применение разработанного в настоящей работе алгоритма обеспечивает значения DSC для дермы 0,99 и 0,98 для моделей 2D U-Net и 3D U-Net соответственно, что сопоставимо с результатами других работ.

Анализ оценок толщин структурных слоев кожи показал (см. рис. 9), что сеть 3D U-Net дает более высокие оценки толщины упорядоченного рогового слоя (см. рис. 9, *а*) по сравнению с 2D U-Net. Рассогласование между результатами применения двух моделей для этого слоя не превышает погрешности, обусловленной пространственным разрешением ОКТ-прибора. Аналогичные результаты для клеточного слоя эпидермиса представлены на

рис. 9, б, из которого видно, что сеть 3D U-Net дает более высокие оценки толщины слоя для случаев большей толщины эпидермиса, тогда как для меньших значений толщины наблюдается обратный тренд. Следует отметить, что практически для всех случаев рассогласование между двумя моделями также не превышает приборной погрешности установки, обусловленной продольным пространственным разрешением системы.

Анализ усредненных по выборке толщин структурных слоев (см. табл. 2) показал, что модель 2D U-Net дает более низкие значения и для средних значений по сравнению с моделью 3D U-Net, однако их разница не превышает ни стандартного отклонения в группе, ни погрешности, обеспечиваемой пространственным разрешением ОКТ-системы. Полученные значения также хорошо согласуются со значениями, продемонстрированными ранее в работе [24]: диапазоны значений толщин упорядоченного рогового слоя, вычисленные на основе данных сегментации, полностью укладываются в диапазоны, опубликованные в этой работе на основании анализа большой выборки, а для клеточного слоя эпидермиса интервалы существенно перекрываются. Это говорит о перспективности предложенного подхода в извлечении морфологической информации из массивов результатов ОКТ-исследования кожи.

Помимо диагностических задач, данные, полученные при автоматической сегментации ОКТ-изображений, могут быть использованы для построения моделей распространения света в биологических тканях и формирования сигналов в системах оптической диагностики [25], а также для задач дозиметрии в фотодинамической терапии [26].

## Заключение

В работе продемонстрированы возможности сверточных нейронных сетей, основанных на архитектуре U-Net, в задаче сегментации трехмерных ОКТ-изображений толстой кожи человека. Основной целью исследования было сравнение различных подходов к выбору блоков изображения для обучения нейросети, что в конечном итоге определяет ее структуру. По имеющимся данным, ранее такие исследования не проводились.

Было показано, что модели при сегментации показывают схожие результаты: DSC для модели 2D U-Net составляет 0,90; 0,94; 0,89; 0,99 для верхнего слоя рогового слоя, рогового слоя, эпидермиса и дермы соответственно. Значения для модели 3D U-Net следующие: 0,90; 0,95; 0,88; 0,98. Следовательно, целесообразно использовать ту модель, для которой легче будет собрать данные для обучения. Например, для сегментации тонкой кожи может быть проще собрать набор размеченных двумерных изображений. В рассмотренном случае применялся полуавтоматический метод первичной разметки с ис-

пользованием априорной информации о конкретном трехмерном массиве, что позволило обучить трехмерную модель.

В большинстве известных работ, посвященных сегментации ОКТ-изображений кожи, в качестве объекта исследования выступала тонкая кожа, тогда как в настоящей работе в качестве объекта выступала толстая кожа, которая морфологически отличается от тонкой кожи. Полученные оценки толщин упорядоченного рогового слоя и клеточного слоя эпидермиса составили  $153 \pm 24$  и  $137 \pm 17$  мкм соответственно — при использовании данных 2D U-Net и  $163 \pm 19$  и  $137 \pm 20$  мкм соответственно — для данных 3D U-Net.

Программное обеспечение, в которое встроены предложенные модели, может стать важным дополнением к системе ОКТ и применяться непосредственно в клинической практике.

**Благодарности.** Авторы благодарны коллективу Молодежной лаборатории искусственного интеллекта и обработки больших массивов данных ННГУ им. Н.И. Лобачевского за полезные обсуждения.

**Финансирование.** Исследование выполнено за счет гранта Российского научного фонда №24-15-00175, <https://rscf.ru/project/24-15-00175/>.

**Конфликт интересов** отсутствует.

## Литература/References

1. *Руководство по оптической когерентной томографии.* Под ред. Gladkoy N.D., Shakhovoy N.M., Sergeeva A.M. М: Физматлит; 2007.

*Rukovodstvo po opticheskoy kogerentnoy tomografii* [Handbook of optical coherence tomography]. Pod red. Gladkovoy N.D., Shakhovoy N.M., Sergeeva A.M. [Gladkova N.D., Shakhova N.M., Sergeev A.M. (editors)]. Moscow: Fizmatlit; 2007.

2. *Мультимодальная оптическая когерентная томография в клинической медицине.* Под ред. Gladkoy N.D., Gelikonova G.V., Kiselevoy E.B. М: Физматлит; 2022.

*Multimodal'naya opticheskaya kogerentnaya tomografiya v klinicheskoy meditsine* [Multimodal optical coherence tomography in clinical medicine]. Pod red. Gladkovoy N.D., Gelikonova G.V., Kiselevoy E.B. [Gladkova N.D., Gelikonov G.V., Kiseleva E.B. (editors)]. Moscow: Fizmatlit; 2022.

3. Wan B., Ganier C., Du-Harpur X., Harun N., Watt F.M., Patalay R., Lynch M.D. Applications and future directions for optical coherence tomography in dermatology. *Br J Dermatol* 2021; 184(6): 1014–1022, <https://doi.org/10.1111/bjd.19553>.

4. Venhuizen F.G., van Ginneken B., Liefers B., van Grinsven M.J.J.P., Fauser S., Hoyng C., Theelen T., Sánchez C.I. Robust total retina thickness segmentation in optical coherence tomography images using convolutional neural networks. *Biomed Opt Express* 2017; 8(7): 3292–3316, <https://doi.org/10.1364/BOE.8.003292>.

5. Zhang X., Yousefi S., An L., Wang R.K. Automated segmentation of intramacular layers in Fourier domain optical coherence tomography structural images from normal

- subjects. *J Biomed Opt* 2012; 17(4): 046011, <https://doi.org/10.1117/1.JBO.17.4.046011>.
6. Schmitt J.M., Xiang S.H., Yung K.M. Speckle in optical coherence tomography. *J Biomed Opt* 1999; 4(1): 95–105, <https://doi.org/10.1117/1.429925>.
7. Hori Y., Yasuno Y., Sakai S., Matsumoto M., Sugawara T., Madjarova V., Yamanari M., Makita S., Yasui T., Araki T., Itoh M., Yatagai T. Automatic characterization and segmentation of human skin using three-dimensional optical coherence tomography. *Opt Express* 2006; 14(5): 1862–1877, <https://doi.org/10.1364/oe.14.001862>.
8. Ali M., Hadj B. Segmentation of OCT skin images by classification of speckle statistical parameters. *IEEE International Conference on Image Processing* 2010; p. 613–616, <https://doi.org/10.1109/icip.2010.5653019>.
9. Li A., Cheng J., Yow A.P., Wall C., Wong D.W., Tey H.L., Liu J. Epidermal segmentation in high-definition optical coherence tomography. *Annu Int Conf IEEE Eng Med Biol Soc* 2015; 2015: 3045–3048, <https://doi.org/10.1109/EMBC.2015.7319034>.
10. Taghavikhailbad A., Adabi S., Clayton A., Soltanizadeh H., Mehregan D., Avanaki M.R.N. Semi-automated localization of dermal epidermal junction in optical coherence tomography images of skin. *Appl Opt* 2017; 56(11): 3116–3121, <https://doi.org/10.1364/AO.56.003116>.
11. Srivastava R., Yow A.P., Cheng J., Wong D.W.K., Tey H.L. Three-dimensional graph-based skin layer segmentation in optical coherence tomography images for roughness estimation. *Biomed Opt Express* 2018; 9(8): 3590–3606, <https://doi.org/10.1364/BOE.9.003590>.
12. Ronneberger O., Fischer P., Brox T. U-Net: convolutional networks for biomedical image segmentation. *Medical image computing and computer-assisted intervention — MICCAI 2015*. 2015; p. 234–241, [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28).
13. Roy A.G., Conjeti S., Karri S.P.K., Sheet D., Katouzian A., Wachinger C., Navab N. ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks. *Biomed Opt Express* 2017; 8(8): 3627–3642, <https://doi.org/10.1364/BOE.8.003627>.
14. Calderon-Delgado M., Tjui J.W., Lin M.Y., Huang S.L. High resolution human skin image segmentation by means of fully convolutional neural networks. *2018 International Conference on Numerical Simulation of Optoelectronic Devices (NUSOD)*. 2018; p. 31–32, <https://doi.org/10.1109/nusod.2018.8570241>.
15. Kepp T., Droigk C., Casper M., Evers M., Hüttmann G., Salma N., Manstein D., Heinrich M.P., Handels H. Segmentation of mouse skin layers in optical coherence tomography image data using deep convolutional neural networks. *Biomed Opt Express* 2019; 10(7): 3484–3496, <https://doi.org/10.1364/BOE.10.003484>.
16. Chueh K.M., Kao H.L., Chen H.H., Shun C.T., Calderon-Delgado M., Huang S.L. Deep feature learning for contour segmentation of aorta's intima by using sub-micron-resolution OCT. *2019 IEEE International Conference on BioPhotonics (BioPhotonics)*. 2019; p. 1–2, <https://doi.org/10.1109/biophotonics.2019.8896753>.
17. Gu Z., Cheng J., Fu H., Zhou K., Hao H., Zhao Y., Zhang T., Gao S., Liu J. CE-Net: context encoder network for 2D medical image segmentation. *IEEE Trans Med Imaging* 2019; 38(10): 2281–2292, <https://doi.org/10.1109/TMI.2019.2903562>.
18. Del Amor R., Morales S., Colomer A., Mogensen M., Jensen M., Israelsen N.M., Bang O., Naranjo V. Automatic segmentation of epidermis and hair follicles in optical coherence tomography images of normal skin by convolutional neural networks. *Front Med (Lausanne)* 2020; 7: 220, <https://doi.org/10.3389/fmed.2020.00220>.
19. Czajkowska J., Badura P., Korzekwa S., Płatkowska-Szczerek A. Deep learning approach to skin layers segmentation in inflammatory dermatoses. *Ultrasonics* 2021; 114: 106412, <https://doi.org/10.1016/j.ultras.2021.106412>.
20. Liu X., Chuchvara N., Liu Y., Rao B. Real-time deep learning assisted skin layer delineation in dermal optical coherence tomography. *OSA Contin* 2021; 4(7): 2008–2023, <https://doi.org/10.1364/osac.426962>.
21. Gao T., Liu S., Gao E., Wang A., Tang X., Fan Y. Automatic segmentation of laser-induced injury OCT images based on a deep neural network model. *Int J Mol Sci* 2022; 23(19): 11079, <https://doi.org/10.3390/ijms231911079>.
22. Ji Y., Yang S., Zhou K., Lu J., Wang R., Roccliffe H.R., Pellicoro A., Cash J.L., Li C., Huang Z. Semisupervised representative learning for measuring epidermal thickness in human subjects in optical coherence tomography by leveraging datasets from rodent models. *J Biomed Opt* 2022; 27(8): 085002, <https://doi.org/10.1117/1.JBO.27.8.085002>.
23. Zhang Y., Liao Q., Ding L., Zhang J. Bridging 2D and 3D segmentation networks for computation efficient volumetric medical image segmentation: an empirical study of 2.5D solutions. *arXiv* 2022, <https://doi.org/10.48550/arXiv.2010.06163>.
24. Shlivko I.L., Kirillin M.Y., Donchenko E.V., Ellinsky D.O., Garanina O.E., Neznakhina M.S., Agrba P.D., Kamensky V.A. Identification of layers in optical coherence tomography of skin: comparative analysis of experimental and Monte Carlo simulated images. *Skin Res Technol* 2015; 21(4): 419–425, <https://doi.org/10.1111/srt.12209>.
25. Kirillin M., Meglinski I., Kuzmin V., Sergeeva E., Myllylä R. Simulation of optical coherence tomography images by Monte Carlo modeling based on polarization vector approach. *Opt Express* 2010; 18(21): 21714–21724, <https://doi.org/10.1364/OE.18.021714>.
26. Kurakina D., Sergeeva E., Khilov A., Kirillin M. Light dose and fluorescence imaging depth in dual-wavelength PDT: a numerical study for various photosensitizer distributions in a layered biotissue. *Journal of Biomedical Photonics & Engineering* 2024; 10(4): 040318, <https://doi.org/10.18287/jbpe24.10.040318>.