
ТЕОРЕТИЧЕСКИЕ ОСНОВЫ ИНФОРМАТИКИ

THEORETICAL FOUNDATIONS OF COMPUTER SCIENCE

УДК 004.93

СРАВНИТЕЛЬНЫЙ АНАЛИЗ НЕЙРОННЫХ СЕТЕЙ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ СЕГМЕНТАЦИИ ЯДЕР КЛЕТОК НА ИММУНОГИСТОХИМИЧЕСКИХ ФЛУОРЕСЦЕНТНЫХ ИЗОБРАЖЕНИЯХ РАКОВЫХ ТКАНЕЙ

СЮЙ СЫЛУНЬ¹⁾, В. В. СКАКУН¹⁾

¹⁾Белорусский государственный университет, пр. Независимости, 4, 220030, г. Минск, Беларусь

Аннотация. Анализ гистологических и иммуногистохимических изображений составляет основу диагностики многих видов раковых заболеваний. Процессу автоматизации анализа цифровых изображений, в частности сегментации ядер клеток на них, сегодня уделяется особое внимание. Благодаря отличной производительности нейронных сетей глубокого обучения и сравнительно высокому уровню достоверности получаемых результатов появляется возможность сочетать ручную и автоматизированную обработку изображений. К настоящему времени создано множество архитектур нейронных сетей для сегментации объектов на изображениях. Однако большая вариабельность изображений раковых клеток не позволяет создать универсальный алгоритм для сегментации

Образец цитирования:

Сюй Сылунь, Скакун ВВ. Сравнительный анализ нейронных сетей глубокого обучения для сегментации ядер клеток на иммуногистохимических флуоресцентных изображениях раковых тканей. *Журнал Белорусского государственного университета. Математика. Информатика.* 2024;1:59–70. EDN: TOOSJI

For citation:

Xu Silun, Skakun VV. Comparative analysis of deep learning neural networks for the segmentation of cancer cell nuclei on immunohistochemical fluorescent images. *Journal of the Belarusian State University. Mathematics and Informatics.* 2024;1: 59–70. Russian. EDN: TOOSJI

Авторы:

Сюй Сылунь – соискатель кафедры системного анализа и компьютерного моделирования факультета радиопизики и компьютерных технологий. Научный руководитель – В. В. Скакун.

Виктор Васильевич Скакун – кандидат физико-математических наук, доцент; доцент кафедры системного анализа и компьютерного моделирования факультета радиопизики и компьютерных технологий.

Authors:

Xu Silun, competitor at the department of system analysis and computer simulation, faculty of radiophysics and computer technologies.

xusilun@hotmail.com

<https://orcid.org/0000-0002-1356-0897>

Viktor V. Skakun, PhD (physics and mathematics), docent; associate professor at the department of system analysis and computer simulation, faculty of radiophysics and computer technologies.

skakun.victor@gmail.com

<https://orcid.org/0000-0003-0880-4188>

ядер клеток на изображениях разного вида тканей, полученных с помощью различных методик. В работе проведен сравнительный анализ архитектур нейронных сетей глубокого обучения для сегментации ядер клеток на иммуногистохимических флуоресцентных изображениях срезов раковой ткани молочной железы. Установлено, что сети, основанные на архитектуре U-Net, дают стабильно хорошие результаты. Наилучшее качество сегментации продемонстрировала архитектура UNet 3+.

Ключевые слова: иммуногистохимические изображения; изображения раковых клеток; сегментация ядер; нейронные сети; глубокое обучение; U-Net.

COMPARATIVE ANALYSIS OF DEEP LEARNING NEURAL NETWORKS FOR THE SEGMENTATION OF CANCER CELL NUCLEI ON IMMUNOHISTOCHEMICAL FLUORESCENT IMAGES

XU SILUN^a, V. V. SKAKUN^a

^aBelarusian State University, 4 Niezaliezhnasci Avenue, Minsk 220030, Belarus

Corresponding author: Xu Silun (xusilun@hotmail.com)

Abstract. The analysis of histological and immunohistochemical images forms the basis for the diagnosis of many types of cancer. The process of automating the analysis of digital images, in particular, the segmentation of cell nuclei on them, is of great attention recently. Due to the excellent performance of deep learning neural networks and the relatively high level of reliability of the obtained results, it becomes possible to combine manual and automated image processing. To date, many neural network architectures have been created for segmenting objects in images. However, the high variability of images of cancer cells does not allow creating an universal algorithm for segmenting the cells nuclei on images of different types of tissues obtained using different techniques. In this paper, a comparative analysis of the architectures of deep learning neural networks for segmentation of cancer cell nuclei on immunohistochemical fluorescent images of breast cancer was carried out. It was established that networks based on the U-Net architecture give consistently good results. The UNet 3+ architecture showed the best segmentation quality.

Keywords: immunohistochemical images; cancer cells images; nuclear segmentation; neural networks; deep learning; U-Net.

Введение

В настоящее время цифровая обработка изображений стала неотъемлемой частью медицинской диагностики и оказалась востребованной для очень широкого круга ее приложений. Анализ гистологических и иммуногистохимических изображений составляет основу диагностики многих видов раковых заболеваний [1–3]. Важным этапом в процессе анализа данного типа изображений является их сегментация. В медицинских исследованиях сегментация изображений позволяет разделять различные виды тканей, а также органы, части скелета и иные объекты путем выделения и классификации их признаков. Например, цель многих исследований патологии раковых опухолей состоит в том, чтобы сегментировать раковые клетки, строму и прочие интересующие исследователей области на изображениях срезов ткани [4].

В диагностике раковых заболеваний очень востребованным является выделение ядер как раковых, так и здоровых клеток, позволяющее количественно оценить степень развития патологии [5]. Задача сегментации ядер клеток сводится к задаче семантической сегментации, когда необходимо определить, что каждый пиксел изображения принадлежит либо ядру клетки, либо другим ее частям. Данная задача усложняется тем, что срез исследуемой ткани имеет толщину, значительно превосходящую размеры клетки. Соответственно, приходится работать с двумерным изображением трехмерной структуры, при котором неизбежны наложение объектов по глубине изображения и неоднородность в интенсивности пикселей, формирующих эти объекты в полученном изображении. Значительно повысить качество сегментации ядер клеток можно, воспользовавшись методикой получения флуоресцентных изображений, где представляющие интерес объекты окрашиваются с помощью флуоресцентных соединений. Комбинация методов флуоресцентной спектроскопии и методов иммуногистохимии позволяет существенно улучшить результаты сегментации ядер клеток и их последующую классификацию на раковые и здоровые [1; 5].

Анализ изображений в медицинской практике сталкивается с одной и той же проблемой, а именно с необходимостью привлечения высококвалифицированных специалистов для ручной обработки большого количества изображений. Качество анализа биомедицинских изображений и результат диагностики не всегда достигают стабильно высокого уровня. Проведение скрининговых исследований большого количества изображений в сжатые сроки является сложной задачей.

С начала 2010-х гг. благодаря созданию инновационных архитектур сверточных нейронных сетей (*convolution neural network*, CNN) появилась возможность эффективно решать задачи классификации и семантической сегментации изображений [3; 6]. Отличная производительность нейронных сетей глубокого обучения и сравнительно высокий уровень достоверности получаемых результатов позволяют сочетать ручной анализ биомедицинских изображений с автоматизированной их обработкой методами глубокого обучения, чтобы добиться наилучшего согласования между точностью и стабильностью диагностических результатов и себестоимостью выполнения работ.

Успешное применение методов глубокого обучения для классификации и семантической сегментации изображений привело к бурному росту как числа архитектур, так и их сложности. Архитектура Visual Geometry Group (VGG) [7], опубликованная в 2014 г., возглавила тенденцию к увеличению глубины CNN. Ее разработчики стремились заменить ядро свертки большого размера ядром малого размера в целях оптимизации качества классификации и повышения вычислительной производительности. В 2016 г. компанией *Microsoft Research* была предложена архитектура Residual Network (ResNet) [8], которая восполняет потери информации, обусловленные изменением разрешения изображения с высокого на низкое при прохождении его через слои нейронной сети. Разработанная в тот же период полностью сверточная нейронная сеть (*fully convolutional network*, FCN) [9] дала следующий импульс к развитию CNN в этой области. В течение года одна за другой были предложены две архитектуры – U-Net и SegNet, совершенствующие архитектуру FCN и значительно повышающие скорость и качество обучения нейронных сетей. Архитектура U-Net [10] специально разработана для семантической сегментации биомедицинских изображений, а архитектура SegNet [11] ориентирована на упрощение структуры сети и повышение скорости обучения. При решении задачи сегментации биомедицинских изображений сквозная (*end-to-end*) структура U-Net, реализующая архитектуру кодер – декодер, продемонстрировала наилучшие характеристики, что предопределило бурный рост числа архитектур, развивающих и совершенствующих ее топологию. В 2016 г. и позже были предложены еще несколько архитектур, таких как FC-DenseNet [12], UNet++ [13], MultiResUNet [14], Res-Seg-net [15], UNet 3+ [16], TransUNet [17] и др., направленных на совершенствование архитектуры нейронной сети и улучшение характеристик сегментации.

Тем не менее из-за высокой вариабельности изображений раковых клеток очень тяжело достичь стабильно хорошего результата. Несмотря на наличие сравнительно большого числа публикаций, посвященных данной тематике, и готовых решений, доступных, например, на веб-сервисе *GitHub*, часто трудно найти действительно рабочую модель или алгоритм для сегментации определенного типа объектов на изображениях срезов разного вида раковых тканей, полученных с помощью различных методик. Кроме того, высокая вариабельность изображений раковых клеток может создавать трудности при получении размеченной экспертами обучающей выборки достаточно большого объема. Ручное выделение границ сотен и тысяч объектов на изображениях, многие из которых едва просматриваются, перекрываются и т. д., требует как усидчивости и значительных временных затрат, так и высочайшей квалификации и опыта экспертов.

Цель настоящей работы состоит в сравнительном анализе различных архитектур нейронных сетей глубокого обучения для сегментации ядер клеток на иммуногистохимических флуоресцентных изображениях срезов раковой ткани молочной железы. Существенную сложность при проведении исследования представляет ограниченность объема обучающей и тестовой выборок. У авторов имеется всего 9 размеченных экспертами изображений с разрешением 2048×2048 пк, что далеко от традиционно требуемых объемов данных для обучения CNN. Ранее авторами были проанализированы различные методы предварительной обработки (подготовки) данных, позволившие даже на основе такого небольшого количества изображений получить сравнительно неплохие результаты сегментации, используя ставшую уже классической архитектуру U-Net [18]. Соответственно, возник вопрос о выборе наиболее эффективной архитектуры CNN и ее параметров для сегментации ядер клеток на иммуногистохимических флуоресцентных изображениях срезов раковой ткани молочной железы.

Сравнительный анализ архитектур нейронных сетей глубокого обучения

В данном исследовании ограничимся анализом наиболее часто используемых при сегментации биомедицинских изображений архитектур нейронных сетей. Все рассмотренные структуры реализуют архитектуру кодер – декодер, поскольку при семантической сегментации изображений требуется попиксельное сопоставление результатов сегментации с исходным изображением.

FCN. Архитектура FCN явилась первой успешной архитектурой в области семантической сегментации [9]. Изначально данная сеть была ориентирована на решение задач классификации изображений, но после соответствующей доработки стала применяться к задачам сегментации. Первая часть сети (кодер) выделяет признаки объектов с помощью сверточных слоев, понижающих размерность с использованием операции субдискретизации (*pooling*). Вторая часть сети (декодер) также с применением операций свертки увеличивает

размерность полученной карты признаков до размера изначального изображения (up sampling). Кроме того, в архитектуре FCN было предложено соединять (суммировать) карты признаков, сформированные на разных уровнях кодера, с картами признаков декодера той же размерности для передачи детализированной информации, имеющейся на верхних уровнях кодера и потерянной при последующей субдискретизации, на соответствующие им уровни декодера. Передача карт признаков кодера на декодер получила устойчивое название skip-соединения (skip connection). На рис. 1 представлена структурная схема архитектуры FCN-8.

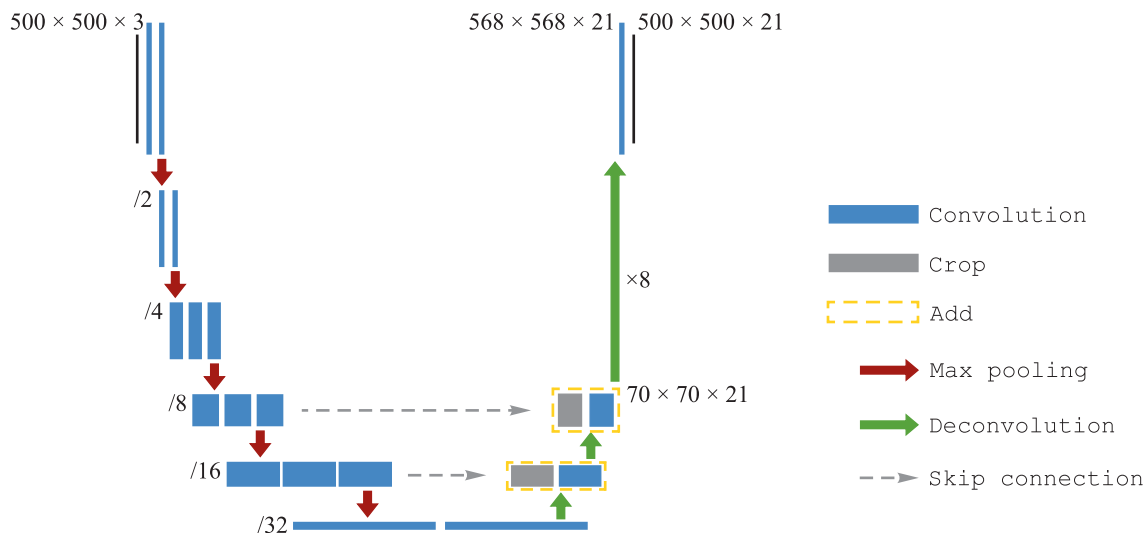


Рис. 1. Структурная схема архитектуры FCN-8
 Fig. 1. FCN-8 architecture structural scheme

U-Net. Архитектура U-Net, предложенная в 2015 г. [10], показала отличные результаты при сегментации медицинских изображений. По сравнению с моделью FCN наиболее важными отличительными особенностями модели U-Net являются полностью симметричная архитектура кодер – декодер, удвоение размерности карт признаков на каждом уровне и введение соединений карт признаков на всех уровнях с помощью операции конкатенации. На рис. 2 отображена стандартная архитектура U-Net, имеющая 23 сверточных слоя.

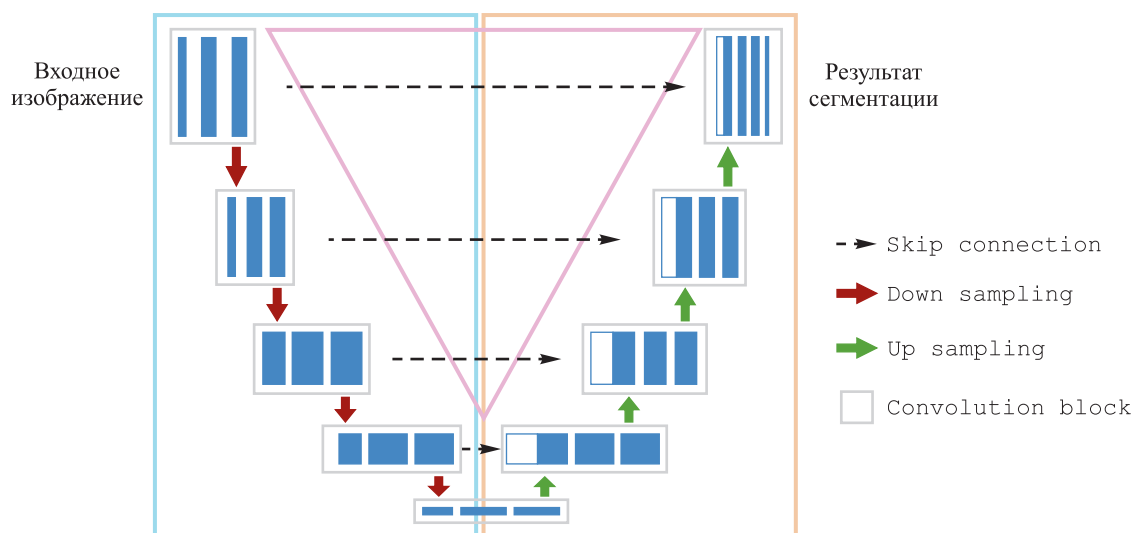


Рис. 2. Структурная схема архитектуры U-Net
 Fig. 2. U-Net architecture structural scheme

Архитектура схематично разделена на три части: голубая рамка соответствует кодеру, оранжевая рамка – декодеру, а розовый треугольник – соединениям сверточных блоков для передачи карт признаков с кодера на декодер (см. рис. 2). На данной архитектуре основано большинство современных моделей сегментации изображений, получивших широкое распространение в медицине в последние

годы. В целом практически все нововведения в моделях, разработанных после появления архитектуры U-Net, реализуют следующие варианты оптимизации ее структуры: 1) комбинацию идей разных моделей для улучшения характеристик как кодера, так и декодера; 2) замену сверточного блока в кодере или декодере новым функциональным модулем, способным улучшить возможность абстрагирования признаков более высоких уровней; 3) изменение топологии соединений карт признаков, полученных на разных уровнях кодера и декодера.

SegNet. Главной особенностью архитектуры SegNet является попытка оптимизации соединений блоков кодера и декодера (отмечены розовым треугольником на рис. 2). В целях повышения эффективности передаются не полные карты признаков, а индексные таблицы соответствующих субдискретизирующих слоев, которые хранят индексы активированных пикселей. Таким образом, слои повышающей дискретизации сети получают от соответствующих субдискретизирующих слоев необходимую информацию о том, как повысить размерность и восстановить сжатые (и, следовательно, утерянные) данные. Модель кодера SegNet имеет топологию VGG-16 [7] без полносвязных слоев. Удаление полносвязных слоев обеспечивает кодеру SegNet меньшие затраты на обучение, чем у многих других архитектур. Уровни сети декодера соответствуют структуре сети кодера.

FC-DenseNet. Модель FC-DenseNet [19] расширяет структуру сети U-Net и также реализует архитектуру кодер – декодер. Кодер FC-DenseNet представляет собой классификационную сеть DenseNet [12] без полносвязных слоев. В отличие от других сетей в архитектуре FC-DenseNet реализованы так называемые плотные сверточные блоки (dense block), а также соединение карт признаков не только между кодером и декодером, но и между отдельными уровнями кодера. В плотных сверточных блоках осуществляется перенос как низкоуровневых признаков из предыдущих слоев, так и высокоуровневых признаков из более поздних слоев, что обеспечивает эффективное повторное использование этих признаков и учет информации различного уровня при повышающей дискретизации и формировании карт сегментации. Внедрение такого плотного соединения дает определенные преимущества: улучшается коэффициент использования параметров сети, и все слои могут повторно обращаться к ранее рассчитанным картам признаков. На рис. 3 изображены общая структура и три основных функциональных блока архитектуры FC-DenseNet.

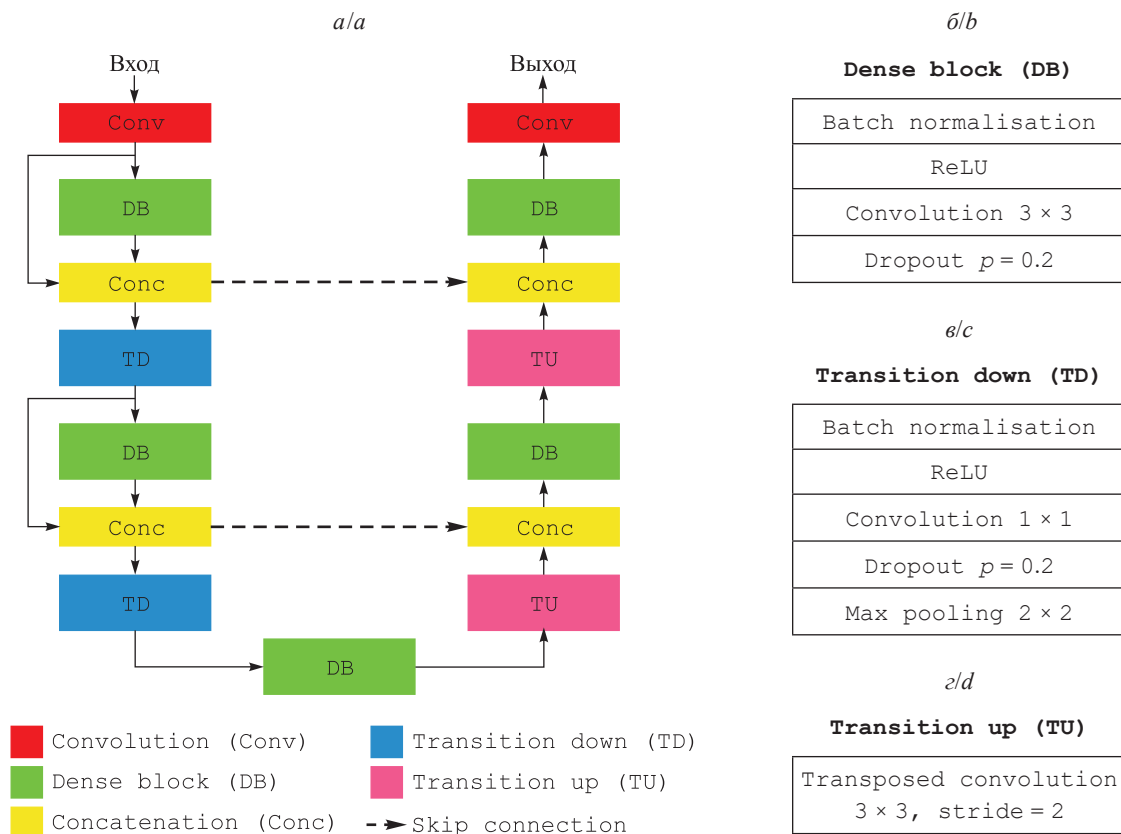


Рис. 3. Общая структура (а) и основные блоки (б – плотный сверточный блок; в – блок понижающей дискретизации; г – блок повышающей дискретизации) архитектуры FC-DenseNet

Fig. 3. General structure (a) and main blocks (b – dense convolutional block; c – down sampling block; d – up sampling block) of the FC-DenseNet architecture

UNet++. Модель UNet++ представляет собой сеть вида кодер – декодер, в которой подсети кодера и декодера соединены серией дополнительных вложенных слоев [13]. В архитектуре U-Net карты признаков кодера напрямую поступают в декодер, а в архитектуре UNet++ они проходят через дополнительный плотный сверточный блок с короткими скип-соединениями [13], количество сверточных слоев которого зависит от уровня пирамиды, схематично изображенной на рис. 4. Добавление вложенных слоев позволяет еще больше сократить семантический разрыв между картами признаков подсетей кодера и декодера.

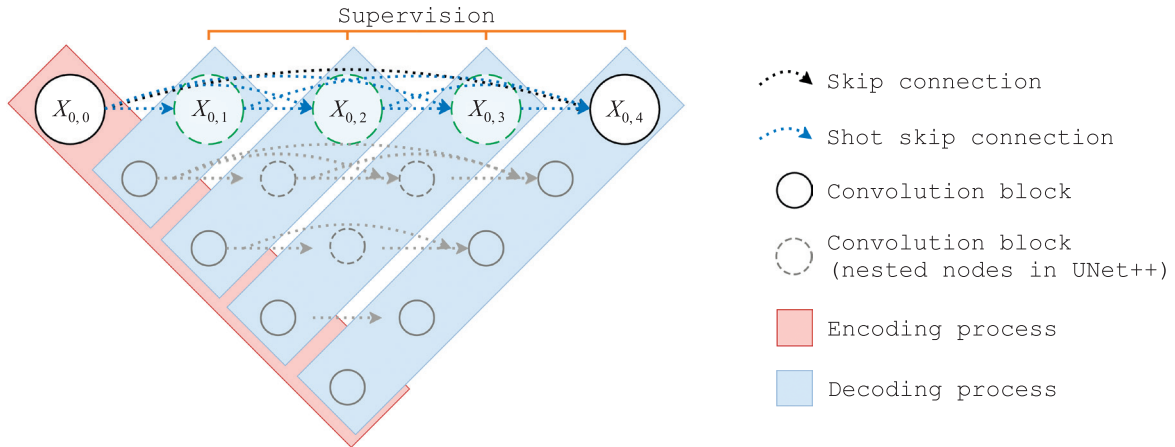


Рис. 4. Структурная схема архитектуры UNet++
 Fig. 4. UNet++ architecture structural scheme

Res-Seg-net. Изначально данная архитектура разрабатывалась для сегментации мультиспектральных изображений дистанционного зондирования Земли [20]. Проблемы сегментации мультиспектральных изображений подобны проблемам сегментации изображений биомедицинского характера (малый объем доступных размеченных данных, возможность переобучения и необходимость в большом количестве параметров для повышения детализации сегментации). Модель Res-Seg-net соответствует типичной архитектуре U-Net и, с одной стороны, использует достоинства модели ResNet [8] (добавляются остаточные соединения (residual connection)) для борьбы с проблемой затухающего градиента, а с другой стороны – достоинства модели SegNet для сокращения количества обучаемых параметров декодера. Это классический случай объединения функций двух моделей для повышения эффективности как кодера, так и декодера.

UNet 3+. В моделях FCN, U-Net, UNet++ карты признаков низких уровней собирают детальную информацию, позволяя более точно выделять границы объектов, а карты признаков верхних уровней содержат обобщенную информацию для определения местоположения объекта. Однако общей проблемой рассмотренных выше моделей является то, что все они не используют одновременно информацию, представленную на разных уровнях. Для преодоления этой проблемы в архитектуре UNet 3+ было предложено соединять блоки декодера не только с блоками кодера одного уровня, но и с блоками кодера всех уровней над этим слоем, что позволяет эффективно и полностью, но с меньшим количеством параметров передавать семантическую информацию со всех уровней кодера на каждый блок декодера (рис. 5) [16]. Кроме того, для получения иерархической информации из агрегированных карт признаков в структуре UNet 3+ применяется полномасштабная контролируемая постобработка (supervision).

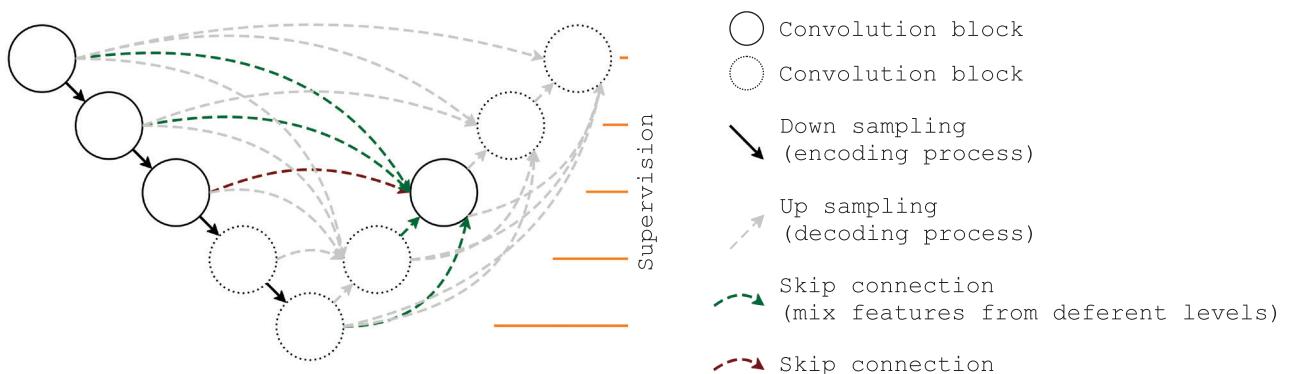


Рис. 5. Структурная схема архитектуры UNet 3+
 Fig. 5. UNet 3+ architecture structural scheme

TransUNet. В архитектуре TransUNet модификации подверглась ветвь кодера путем внедрения блока преобразования Transformer (рис. 6) [17]. Чтобы компенсировать потерю разрешения, вызванную применением блоков преобразования, модель TransUNet реализует гибридную архитектуру CNN-Transformer, которая позволяет использовать как подробную пространственную информацию высокого разрешения, получаемую от сверточной сети, так и глобальный контекст, закодированный в блоках преобразования.

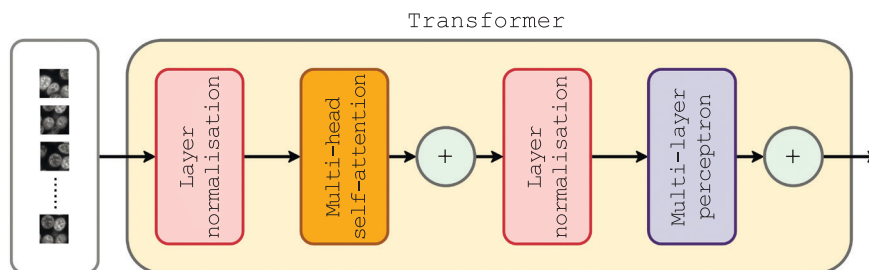


Рис. 6. Структурная схема блока преобразования Transformer, используемого в архитектуре TransUNet

Fig. 6. Structural scheme of the Transformer block used in the TransUNet architecture

Материалы и методы исследования

Материалы. В статье исследуются иммуногистохимические флуоресцентные изображения раковых клеток молочной железы, полученные при 10-кратном увеличении на инвертированном эпифлуоресцентном микроскопе Nikon TE200 (Япония) с CCD-камерой Photometrics серии 300 (Германия) и сохраненные в стандарте RGB. Размер изображений составляет 2048×2048 пк в каждом из трех цветовых каналов, разрешение – 0,2 мкм/пк, или 5 мкм. В качестве индикатора раковых клеток используется рецептор эстрогена, помечаемый цианиновым красителем Cy5 и регистрируемый в красном канале изображения [1]. В противоположность здоровым клеткам в цитоплазме раковых клеток накапливается белок цитокератин, который маркируется цианиновым красителем Cy3 и регистрируется в зеленом канале изображения. Для маркировки всех ядер используется краситель 4,6-диамидино-2-фенилиндолил дигидрохлорид (DAPI), так как он обладает выраженной способностью присоединяться к РНК. Максимум спектра испускания данного флуоресцентного соединения лежит в синем канале. Соответственно, маркерами раковых клеток являются два красителя – Cy5 и Cy3.

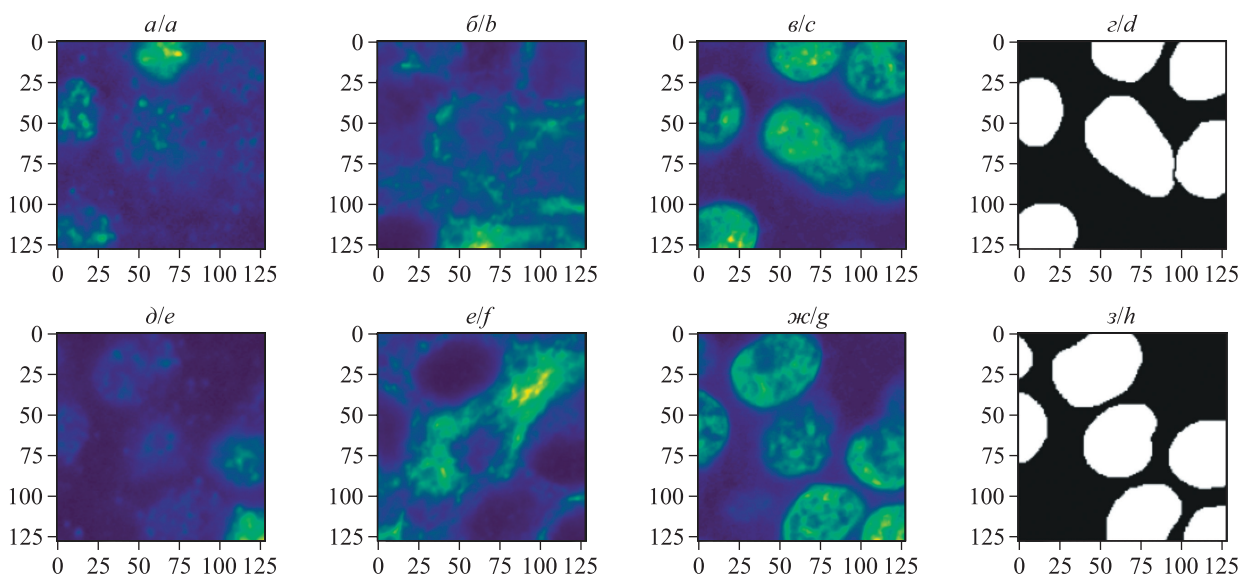


Рис. 7. Пример двух фрагментов иммуногистохимического флуоресцентного изображения клеток раковой ткани молочной железы. Красный канал (а, д) представляет ядра раковых клеток, зеленый канал (б, е) – цитоплазму, а синий канал (в, ж) – все ядра, включая здоровые.

В последнем столбце (з, з) приведены размеченные экспертами маски клеток

Fig. 7. An example of two fragments of an immunohistochemical fluorescent image of breast cancer cells.

The red channel (a, e) represents the nuclei of cancer cells, the green channel (b, f) – the cytoplasm, and the blue channel (c, g) – all nuclei, including healthy ones. The last column (d, h) shows cell masks labeled by experts

Протокол подготовки образцов и методика получения изображений подробно описаны в работе [1]. В общей сложности 9 экспериментальных изображений были вручную размечены экспертами. Размеченные изображения, на которых выделены только маски ядер раковых клеток (ground truth), представляют собой бинарные изображения, где пикселям ядер клеток присвоено значение 1. Несколько случайно отобранных фрагментов экспериментальных изображений размером 256×256 пк показаны на рис. 7. Размеченные маски ядер клеток представлены в четвертом столбце (см. рис. 7, з и з).

Методы. Сегментация изображений выполнялась с помощью искусственных нейронных сетей глубокого обучения, реализующих семантическую сегментацию. Выделяются два класса пикселей: класс пикселей, составляющих ядра раковых клеток, и класс остальных пикселей безотносительно к месту их расположения (ядра здоровых клеток, цитоплазма, межклеточное пространство).

Программная реализация нейронных сетей осуществлялась на языке Python с помощью инфраструктуры TensorFlow (версия 1.14) и высокоуровневой библиотеки Keras (версия 2.2.4). Для проведения исследований использовались среда Google Colab и тарифный план Google Colab Pro+, предоставляющий доступ к вычислительным мощностям графического процессора NVIDIA A100 40 GB. Для аугментации изображений применялась библиотека Albumentations (*albumentations.ai*), а вычисление метрик оценки качества сегментации выполнялось с помощью методов библиотеки Segmentation models (*segmentation-models.readthedocs.io*).

Метрики оценки качества сегментации. Для вычисления функции потерь (Loss) использовался расчет бинарной кросс-энтропии:

$$L = -\frac{1}{N} \sum_{i=1}^N y_i \log(p(y_i)) + (1 - y_i) \log(1 - p(y_i)),$$

где y_i – значение i -го пикселя в размеченной экспертами маске изображения (метка класса; может принимать значения 0 или 1); $p(y_i)$ – вероятность принадлежности i -го пикселя исследуемого изображения метке y_i ; N – количество пикселей. Чем меньше значение функции потерь, тем лучше эффект обучения.

Коэффициент подобия Жаккара, или, иначе, пересечение над объединением (*intersection over union*, IOU), используется для расчета метрики, оценивающей степень совпадения результатов сегментации с размеченной экспертами маской:

$$\text{IOU} = \frac{|A \cap B|}{|A \cup B|},$$

где A и B – прогнозируемый результат и размеченная экспертами маска соответственно. Данный коэффициент вычисляется при пороговом значении вероятности предсказания 0,5. Чем больше коэффициент IOU, тем лучше сегментация (ближе к размеченной экспертами маске).

Результаты и их обсуждение

Все архитектуры нейронных сетей, рассмотренные выше, были программно реализованы на языке Python с использованием библиотек TensorFlow и Keras. Детали реализации приведены в подразделе «Методы». Исходные данные, представляющие собой иммуногистохимические флуоресцентные изображения среза ткани опухоли молочной железы, описаны в подразделе «Материалы». Особенностью настоящей работы является исследование возможности и эффективности сегментации ядер клеток на иммуногистохимических флуоресцентных трехканальных изображениях в условиях очень малой входной выборки (всего 9 размеченных изображений размером 2048×2048 пк).

Чтобы обеспечить объективное сравнение полученных результатов, использовался один и тот же набор аннотированных изображений для обучения нейронной сети. Процедуры предварительной обработки изображений и гиперпараметры обучения нейронных сетей также были одинаковыми. В ходе ранее проведенных авторами исследований [18] было установлено, что наиболее эффективным способом предварительной подготовки набора данных для сегментации ядер клеток на иммуногистохимических флуоресцентных изображениях среза ткани опухоли молочной железы с применением нейронной сети глубокого обучения U-Net является разбиение исходного изображения на фрагменты размером 256×256 пк, а затем выделение в каждом из них 8 патчей размером 128×128 пк в случайном смещении от начала фрагмента. Подобный подход позволяет сформировать обучающую выборку достаточно большого объема. Таким образом, разрешение изображений, которые использовались для обучения и тестирования нейронных сетей глубокого обучения, исследуемых в данной работе, составляло 128×128 пк. Один и тот же алгоритм предварительной обработки применялся как к обучающим, так и к тестовым изображениям. Из обучающего набора были удалены пустые патчи (с количеством ненулевых пикселей менее 150) [18].

Из 9 исходных (2048×2048 пк) размеченных изображений 7 изображений использовались для обучения, 1 изображение – для проверки процесса обучения и 1 изображение – для тестирования. Соответственно, размер обучающей выборки после этапа предварительной обработки составил 12 064 патча ($197\,656\,576$ пк), размер проверочной и тестовой выборок – по 1644 патча. Количество эпох в процессе обучения было выбрано равным 30, размер серии (*batch size*) как для обучения, так и для проверки равнялся 32. В целях экономии времени обучения был применен метод ранней остановки (*early stop*): если значение функции потерь на проверочном наборе не уменьшалось после 10 итераций в процессе обучения, процесс обучения нейронной сети прекращался. Для обеспечения объективности сравнения архитектур их реализации, заимствованные из репозитория веб-сервиса *GitHub*, приводились к единому виду. Во всех случаях по возможности использовались одни и те же наборы гиперпараметров, но количество сверток и их размерность на первом этапе исследований не изменялись. Таким образом, более поздние (производные) модели, такие как Res-Seg-net, UNet 3+ и TransUNet, имели гораздо большее количество обучаемых параметров, чем их прототипы FCN-8, U-Net и FC-DenseNet. Как известно, чем больше параметров обучается в модели, тем более высокой точности можно от них ожидать, но при этом время обучения также значительно возрастает. В данном исследовании архитектура TransUNet имела наибольшее количество параметров и самое продолжительное время обучения, близкое к 1 ч, однако не во всех случаях обеспечивала наивысшую точность сегментации.

Результаты проведенных вычислительных экспериментов сведены в табл. 1. Сравнительный анализ полученных значений метрик оценки качества сегментации показывает, что на нашем наборе данных наилучшее качество сегментации как по значению функции потерь, так и по значению коэффициента IOU демонстрирует модель UNet 3+, а наихудшее – модель SegNet. Примеры лучшего и худшего результатов сегментации, полученных с помощью моделей UNet 3+ и SegNet соответственно, представлены на рис. 8.

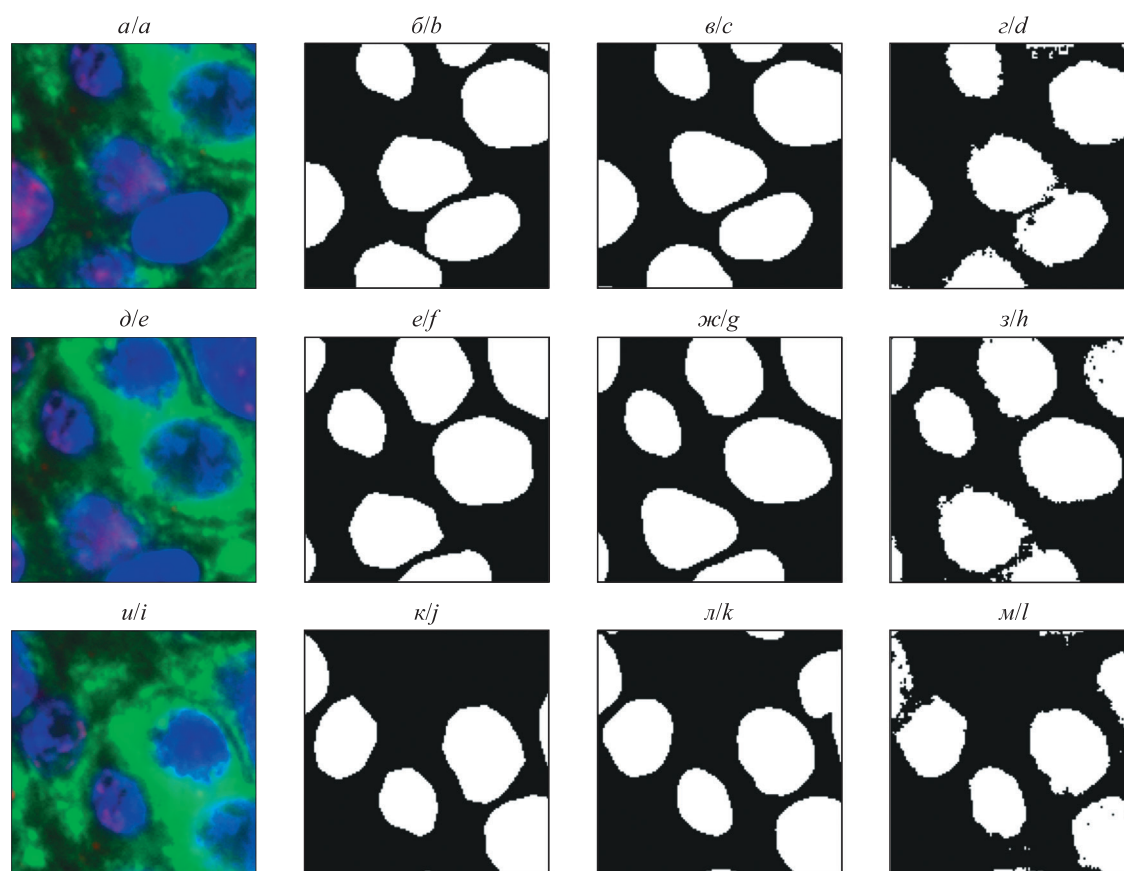


Рис. 8. Примеры лучшего (*в, ж, л*) и худшего (*з, з, м*) результатов сегментации, полученных с применением моделей Unet 3+ и SegNet соответственно (*а, д, и* – исходное изображение; *б, е, к* – маска ядер, размеченная экспертами; *в, з, ж, з, л, м* – результат сегментации)

Fig. 8. Examples of the best (*c, g, k*) and worst (*d, h, l*) segmentation results obtained using the Unet 3+ and SegNet models respectively (*a, e, i* – original image; *b, f, j* – mask of nuclei marked by experts; *c, d, g, h, k, l* – result of segmentation)

Таблица 1

Метрики оценки качества сегментации

Table 1

Evaluation metrics for assessing the quality of segmentation

Архитектура	Количество обучаемых параметров	Значение функции потерь (бинарная кросс-энтропия)	Значение коэффициента IOU
FCN-8	2 248 721	0,1677	0,8260
U-Net	1 941 105	0,1688	0,8273
SegNet	29 443 075	0,2051	0,7938
FC-DenseNet-56	1 374 865	0,1575	0,8321
FC-DenseNet-103	9 319 521	0,2213	0,7983
UNet++	2 261 889	0,1600	0,8363
Res50-Seg-net	34 819 201	0,1845	0,8040
Res101-Seg-net	53 837 441	0,1806	0,8047
UNet 3+	26 072 449	0,1492	0,8433
TransUNet	40 504 473	0,1601	0,8371

Примечание. Полужирным шрифтом выделены лучшие значения метрик.

Как видно из представленных результатов, все исследованные архитектуры нейронных сетей достаточно эффективно выполняют сегментацию ядер клеток на иммуногистохимических флуоресцентных изображениях раковой ткани, что неудивительно, поскольку для исследования были отобраны архитектуры, показавшие наилучшее качество сегментации изображений. Для анализируемого класса изображений наилучшие результаты дают архитектуры, основанные на архитектуре U-Net, специально разработанной для сегментации медицинских изображений. Хотя архитектура U-Net была опубликована еще в 2015 г., модели, реализующие ее базовую структуру кодер – декодер с соединением карт признаков на разных уровнях, по-прежнему играют важную роль в области семантической сегментации медицинских изображений.

Тем не менее имеется существенный разрыв в значениях показателей между лучшим и худшим вариантами, но, учитывая рабочую нагрузку на аппаратные средства, в значительной степени определяемую количеством обучаемых параметров, архитектура нейронной сети, которая дает максимальное значение коэффициента IOU, не всегда является оптимальной. На основании полученных результатов авторы отобрали четыре наилучшие архитектуры и для обеспечения более справедливого их сравнения значительно сократили количество обучаемых параметров в моделях Unet 3+ и TransUNet. Так, в кодере и декодере архитектуры UNet 3+ количество ядер свертки было уменьшено в 4 раза. В архитектуре TransUNet структура блоков преобразования Transformer в кодере осталась неизменной, размерность же остальных сверточных слоев уменьшилась в 4 раза. Процесс обучения модифицированных нейронных сетей был проведен заново. Время обучения существенно снизилось (для модели UNet 3+ оно сократилось с 2164,8 до 538,5 с, т. е. примерно в 4 раза). Результаты второго этапа исследований представлены в табл. 2. Как и следовало ожидать, метрики оценки качества сегментации ухудшились, но незначительно.

Таблица 2

Метрики оценки качества сегментации при сравнимом количестве параметров обучения

Table 2

Metrics for assessing the quality of segmentation with a comparable number of learning parameters

Архитектура	Количество обучаемых параметров	Значение функции потерь (бинарная кросс-энтропия)	Значение коэффициента IOU
U-Net	1 941 105	0,1688	0,8273
UNet++	2 261 889	0,1600	0,8363

Окончание табл. 2
Ending of the table 2

Архитектура	Количество обучаемых параметров	Значение функции потерь (бинарная кросс-энтропия)	Значение коэффициента IOU
UNet 3+	1 631 329	0,1531	0,8406
TransUNet	2 343 217	0,1643	0,8351

Примечание. Полужирным шрифтом выделены лучшие значения метрик.

Модель Unet 3+ продолжает демонстрировать наиболее высокую точность сегментации ядер на изображениях раковой ткани молочной железы. Модели Unet++ и TransUNet показывают незначительно более низкую точность сегментации. Классическая модель U-Net находится на последнем месте по точности сегментации, однако ее отставание от лидеров совсем невелико. Важно отметить, что архитектура U-Net намного проще в реализации, что может быть важным фактором для ее применения в отдельных случаях.

Заключение

Проведен сравнительный анализ наиболее эффективных нейронных сетей глубокого обучения, таких как FCN-8, U-Net, SegNet, FC-DenseNet, UNet++, Res-Seg-net, UNet 3+, TransUNet, показавших хорошие результаты в семантической сегментации изображений, при решении задачи сегментации ядер клеток на иммуногистохимических флуоресцентных изображениях срезов раковой ткани молочной железы. Установлено, что применение CNN, основанных на архитектуре U-Net, позволяет эффективно выполнять сегментацию ядер клеток на флуоресцентных изображениях срезов опухолевой ткани. Среди них наилучшие результаты сегментации показала нейронная сеть глубокого обучения на базе структуры UNet 3+. Основой для повышения качества сегментации в большинстве рассматриваемых случаев является наличие соединений (*skip connections*) карт признаков. Данные соединения (как внутри кодера или декодера, так и между ними) значительно улучшают качество сегментации. В модели UNet 3+ механизм соединений реализован наиболее эффективно, что и обусловило получение наилучшего результата.

Качество сегментации ядер клеток можно существенно повысить, применив постобработку результатов сегментации. Наибольшее количество ошибок связано с получением слившихся ядер после финальной пороговой классификации пикселей (пиксели, имеющие значение больше 0,5, относят к классу «принадлежащие ядру раковой клетки»). Использование, например, метода водораздела позволит улучшить итоговое качество сегментации.

Библиографические ссылки / References

1. Chung GG, Zerkowski MP, Ghosh S, Camp RL, Rimm DL. Quantitative analysis of estrogen receptor heterogeneity in breast cancer. *Laboratory Investigation*. 2007;87(7):662–669. DOI: 10.1038/labinvest.3700543.
2. Camp RL, Chung GG, Rimm DL. Automated subcellular localization and quantification of protein expression in tissue microarrays. *Nature Medicine*. 2002;8(11):1323–1328. DOI: 10.1038/nm791.
3. Todewale NS. Lesion segmentation from mammogram images using a U-Net deep learning network. *International Journal of Engineering Research and Technology*. 2020;9(2):406–411.
4. Lagree A, Mohebpour M, Meti N, Saednia K, Lu Fang-I, Slodkowska E, et al. A review and comparison of breast tumor cell nuclei segmentation performances using deep convolutional neural networks. *Scientific Reports*. 2021;11:8025. DOI: 10.1038/s41598-021-87496-1.
5. Lisitsa YV, Yatskou MM, Apanasovich VV, Apanasovich TV. Algorithm for automatic segmentation of nuclear boundaries in cancer cells in three-channel luminescent images. *Journal of Applied Spectroscopy*. 2015;82(4):634–643. DOI: 10.1007/s10812-015-0156-2.
6. Saood A, Hatem I. COVID-19 lung CT image segmentation using deep learning methods: U-Net versus SegNet. *BMC Medical Imaging*. 2021;21:19. DOI: 10.1186/s12880-020-00529-5.
7. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556v1 [Preprint]. 2014 [cited 2022 March 18]: [10 p.]. Available from: <https://arxiv.org/abs/1409.1556v1>.
8. He Kaiming, Zhang Xiangyu, Ren Shaoqing, Sun Jian. Deep residual learning for image recognition. In: *2016 IEEE conference on computer vision and pattern recognition (CVPR); 2016 June 27–30; Las Vegas, USA*. [S. l.]: IEEE; 2016. p. 770–778. DOI: 10.1109/CVPR.2016.90.
9. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. arXiv:1411.4038v1 [Preprint]. 2014 [cited 2022 March 18]: [10 p.]. Available from: <https://arxiv.org/abs/1411.4038v1>.
10. Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. *Medical image computing and computer-assisted intervention (MICCAI-2015). Proceedings of the 18th International conference; 2015 October 5–9; Munich, Germany. Part 3*. Cham: Springer; 2015. p. 234–241 (Goos G, Hartmanis J, van Leeuwen J, editors. Lecture notes in computer science; volume 9351). DOI: 10.1007/978-3-319-24574-4_28.

11. Badrinarayanan V, Kendall A, Cipolla R. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2017;39(12):2481–2495. DOI: 10.1109/TPAMI.2016.2644615.
12. Jégou S, Drozdal M, Vazquez D, Romero A, Bengio Y. The one hundred layers tiramisù: fully convolutional DenseNets for semantic segmentation. In: *2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW); 2017 July 21–26; Honolulu, USA*. [S. l.]: IEEE; 2017. p. 1175–1183. DOI: 10.1109/CVPRW.2017.156.
13. Zhou Zongwei, Siddiquee MMR, Tajbakhsh N, Liang Jianming. UNet++: a nested U-Net architecture for medical image segmentation. In: Stoyanov D, Taylor Z, Carneiro G, Syeda-Mahmood T, Martel A, Maier-Hein L, et al., editors. *Deep learning in medical image analysis and multimodal learning for clinical decision support. Proceedings of the 4th International workshop DLMIA-2018 and 8th International workshop ML-CDS-2018 held in conjunction with MICCAI-2018; 2018 September 20; Granada, Spain*. Cham: Springer; 2018. p. 3–11 (Goos G, Hartmanis J, van Leeuwen J, editors. Lecture notes in computer science; volume 11045). DOI: 10.1007/978-3-030-00889-5_1.
14. Ibtehaz N, Rahman MS. MultiResUNet: rethinking the U-Net architecture for multimodal biomedical image segmentation. *Neural Networks*. 2020;121:74–87. DOI: 10.1016/j.neunet.2019.08.025.
15. Saxena N, Kishore Babu N, Raman B. Semantic segmentation of multispectral images using Res-Seg-net model. In: *2020 IEEE 14th International conference on semantic computing (ICSC); 2020 February 3–5; San Diego, USA*. [S. l.]: IEEE; 2020. p. 154–157. DOI: 10.1109/ICSC.2020.00030.
16. Huang Huimin, Lin Lanfen, Tong Ruofeng, Hu Hongjie, Zhang Qiaowei, Iwamoto Y, et al. UNet 3+: a full-scale connected U-Net for medical image segmentation. In: *2020 IEEE International conference on acoustics, speech and signal processing (ICASSP); 2020 May 4–8; Barcelona, Spain*. [S. l.]: IEEE; 2020. p. 1055–1059. DOI: 10.1109/ICASSP40776.2020.9053405.
17. Chen Jieneng, Lu Yongyi, Yu Qihang, Luo Xiangde, Adeli E, Wang Yan, et al. TransUNet: transformers make strong encoders for medical image segmentation. arXiv:2102.04306 [Preprint]. 2021 [cited 2022 March 21]: [13 p.]. Available from: <https://arxiv.org/abs/2102.04306>.
18. Xu Silun, Skakun V. Comparison of deep learning preprocessing algorithms of nuclei segmentation on fluorescence immunohistology images of cancer cells. In: Tuzikov AV, Belotserkovsky AM, Lukashovich MM, editors. *Pattern recognition and information processing. Revised selected papers of the 15th International conference PRIP-2021; 2021 September 21–24; Minsk, Belarus*. Cham: Springer; 2022. p. 166–177 (Communications in computer and information science; volume 1562). DOI: 10.1007/978-3-030-98883-8_12.
19. Huang Gao, Liu Zhuang, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In: *2017 IEEE conference on computer vision and pattern recognition (CVPR); 2017 July 21–26; Honolulu, USA*. [S. l.]: IEEE; 2017. p. 2261–2269. DOI: 10.1109/CVPR.2017.243.
20. Pan Bin, Shi Zhenwei, Xu Xia, Shi Tianyang, Zhang Ning, Zhu Xinzhong. CoinNet: copy initialization network for multispectral imagery semantic segmentation. *IEEE Geoscience and Remote Sensing Letters*. 2019;16(5):816–820. DOI: 10.1109/LGRS.2018.2880756.

Получена 03.01.2024 / исправлена 14.02.2024 / принята 14.02.2024.
Received 03.01.2024 / revised 14.02.2024 / accepted 14.02.2024.