



Научная статья

УДК 004.896; 630*96

DOI: 10.37482/0536-1036-2024-1-126-140

Сегментация перекрывающихся изображений деревьев на цифровых снимках лесных массивов

И.В. Петухов[✉], *д-р техн. наук, проф.*; *ResearcherID: [A-9472-2014](#)*,

ORCID: <https://orcid.org/0009-0000-2365-4857>

К.О. Иванов, *канд. техн. наук*; *ResearcherID: [A-6724-2014](#)*,

ORCID: <https://orcid.org/0000-0002-6395-4085>

Д.М. Ворожцов, *канд. техн. наук*; *ResearcherID: [JAN-6772-2023](#)*,

ORCID: <https://orcid.org/0009-0005-8396-2778>

А.А. Роженцов, *д-р техн. наук, проф.*; *ResearcherID: [AAU-8039-2020](#)*,

ORCID: <https://orcid.org/0000-0003-0043-0070>

Н.И. Роженцова, *канд. техн. наук*; *ResearcherID: [JMD-0172-2023](#)*,

ORCID: <https://orcid.org/0009-0006-4024-7886>

Л.А. Стешина, *канд. техн. наук, доц.*; *ResearcherID: [JMC-9977-2023](#)*,

ORCID: <https://orcid.org/0009-0006-1526-991X>

Поволжский государственный технологический университет, пл. Ленина, д. 3, г. Йошкар-Ола, Россия, 424000; PetuhovIV@volgatech.net[✉], IvanovKO@volgatech.net, VorozhtsovDM@volgatech.net, RozhencovAA@volgatech.net, RozhencovaNI@volgatech.net, SteshinaLA@volgatech.net

Поступила в редакцию 11.06.23 / Одобрена после рецензирования 09.09.23 / Принята к печати 13.09.23

Аннотация. Применение систем поддержки принятия решений на основе компьютерного зрения и искусственного интеллекта значительно улучшает условия работы операторов технологических машин лесного комплекса, профессиональная деятельность которых связана с высокой интенсивностью и психоэмоциональными перегрузками. При помощи компьютерного зрения и искусственного интеллекта оператор легко и быстро может получить информацию о состоянии лесосеки и выбрать оптимальные решения для проведения рабочих операций. Это облегчает его труд и позволяет сократить время на поиск и анализ данных о лесосеке. При этом одним из ключевых элементов подобных систем является подсистема автоматической сегментации объектов на изображениях. Исследована возможность сегментации перекрывающихся объектов на изображениях лесных массивов с помощью сверточной нейронной сети на базе архитектуры Mask R-CNN. В отличие от большинства работ по схожим темам, используются цветные изображения, полученные с помощью RGB-камеры, а не лидара. Это создает перспективу снижения стоимости аппаратно-программных комплексов поддержки принятия решений операторами лесосечных машин. В качестве сегментируемых объектов выступают изображения ствола и кроны деревьев хвойных и лиственных пород, перекрывающихся друг друга. С помощью графического редактора GIMP выполнена ручная разметка цветных изображений, содержащих в общей сложности 134 дерева 4 различ-

© Петухов И.В., Иванов К.О., Ворожцов Д.М., Роженцов А.А., Роженцова Н.И., Стешина Л.А., 2024



Статья опубликована в открытом доступе и распространяется на условиях лицензии CC BY 4.0

ных пород: ель, осина, береза и сосна. С использованием разработанной базы данных поставлен эксперимент по дообучению сверточной нейронной сети Mask R-CNN сегментации перекрывающихся частей деревьев на цифровых снимках лесных массивов. При этом нейронная сеть была предварительно обучена с применением набора данных Microsoft COCO dataset, содержащего более 200 000 изображений 80 различных классов объектов, таких как люди, автомобили, животные и различные предметы. В процессе обучения нейронной сети изображения, подаваемые на ее вход, подвергались серии линейных и нелинейных геометрических преобразований, что позволило увеличить объем обучаемых данных в 11 раз. В итоге точность сегментации изображений стволов и крон хвойных и лиственных пород деревьев, перекрывающих друг друга, составила 79 %, что допускает использование нейронных сетей подобной архитектуры в системах поддержки принятия решений для операторов лесосечных машин.

Ключевые слова: глубокое обучение, искусственный интеллект, сверточная нейронная сеть, сегментация перекрывающихся объектов, система поддержки принятия решений для операторов лесосечных машин

Благодарности: Работа была поддержана грантом № 23-19-00568 РФФ.

Для цитирования: Петухов И.В., Иванов К.О., Ворожцов Д.М., Роженцов А.А., Роженцова Н.И., Стешина Л.А. Сегментация перекрывающихся изображений деревьев на цифровых снимках лесных массивов // Изв. вузов. Лесн. журн. 2024. № 1. С. 126–140. <https://doi.org/10.37482/0536-1036-2024-1-126-140>

Original article

Segmentation of Overlapping Tree Images in the Digital Photographs of Forest Areas

Igor V. Petukhov[✉], Doctor of Engineering, Prof.; ResearcherID: [A-9472-2014](https://orcid.org/0009-0000-2365-4857),

ORCID: <https://orcid.org/0009-0000-2365-4857>

Konstantin O. Ivanov, Candidate of Engineering; ResearcherID: [A-6724-2014](https://orcid.org/0000-0002-6395-4085),

ORCID: <https://orcid.org/0000-0002-6395-4085>

Dmitry M. Vorozhtsov, Candidate of Engineering; ResearcherID: [JAN-6772-2023](https://orcid.org/0009-0005-8396-2778),

ORCID: <https://orcid.org/0009-0005-8396-2778>

Alexey A. Rozhentsov, Doctor of Engineering, Prof.; ResearcherID: [AAU-8039-2020](https://orcid.org/0000-0003-0043-0070),

ORCID: <https://orcid.org/0000-0003-0043-0070>

Nataliya I. Rozhentsova, Candidate of Engineering; ResearcherID: [JMD-0172-2023](https://orcid.org/0009-0006-4024-7886),

ORCID: <https://orcid.org/0009-0006-4024-7886>

Ludmila A. Steshina, Candidate of Engineering, Assoc. Prof.;

ResearcherID: [JMC-9977-2023](https://orcid.org/0009-0006-1526-991X), ORCID: <https://orcid.org/0009-0006-1526-991X>

Volga State University of Technology, Lenina pl., 3, Yoshkar-Ola, 424000, Russian Federation;

PetuhovIV@volgatech.net[✉], IvanovKO@volgatech.net, VorozhtsovDM@volgatech.net,

RozhencovAA@volgatech.net, RozhencovaNI@volgatech.net, SteshinaLA@volgatech.net

Received on June 11, 2023 / Approved after reviewing on September 9, 2023 / Accepted on September 13, 2023

Abstract. The use of decision support systems based on computer vision and artificial intelligence significantly improves the working conditions for the operators of technological machines in the timber sector, whose work implies high intensity and psycho-emotional

overload. By means of computer vision and artificial intelligence the operator can quickly and easily obtain the data on the state of the cutting area and adopt the optimal solution for holding the working operation. This facilitates his work and reduces the time spent searching and analyzing the data on the cutting area. Meanwhile, one of the key elements of such a system is a subsystem for automatic segmentation of objects in the photograph. We have explored the possibility of segmenting overlapping objects in the photographs of forest areas using a convolutional neural network based on the Mask R-CNN architecture. Unlike in most works on similar topics, the objects of this study are color photographs taken by an RGB camera rather than a lidar. This creates the prospect for reducing the cost of hardware and software systems used to support decision-making by the operators of logging machines. The images of the stems and crowns of coniferous and deciduous trees overlapping each other are the segmented objects under consideration. Using the GIMP graphic editor, we have manually marked the color photographs depicting a total of 134 trees of 4 different species: spruce, aspen, birch and pine. Utilizing the developed database, we have carried out an experiment to further train the Mask R-CNN convolutional neural network for segmentation of overlapping parts of the trees in the digital photographs of forest areas. The neural network has been pre-trained using the Microsoft COCO dataset containing more than 200,000 images of 80 different classes of objects such as people, cars, animals and various items. While training the neural network, the images supplied to its input were subjected to a series of various linear and nonlinear geometric transformations, which made it possible to increase the volume of training data by 11 times. As a result, the accuracy of segmentation of the images of the stems and crowns of coniferous and deciduous trees overlapping each other has reached 79 %, which allows the use of neural networks based on a similar architecture in decision support systems for logging machine operators.

Keywords: deep learning, artificial intelligence, convolutional neural network, segmentation of overlapping objects, decision support system for logging machine operators

Acknowledgements: This work was supported by the Russian Science Foundation grant no. 23-19-00568.

For citation: Petukhov I.V., Ivanov K.O., Vorozhtsov D.M., Rozhentsov A.A., Rozhentsova N.I., Steshina L.A. Segmentation of Overlapping Tree Images in the Digital Photographs of Forest Areas. *Lesnoy Zhurnal* = Russian Forestry Journal, 2024, no. 1, pp. 126–140. (In Russ.). <https://doi.org/10.37482/0536-1036-2024-1-126-140>

Введение

Работа операторов технологических машин характеризуется высокой интенсивностью, психоэмоциональными перегрузками, необходимостью: соблюдения технологических норм и правил, тщательного контроля окружающей обстановки, достижения целевых показателей трудовой деятельности. В частности, оператору харвестера, особенно при проведении выборочных рубок, рубок ухода и т. п., необходимо определять и учитывать в работе таксационные показатели деревьев, их распределение по лесосеке, знать характеристики используемого оборудования, схему разработки лесосеки. Существенную помощь в этом могут оказывать средства поддержки принятия решений, базирующиеся на системах технического зрения и технологиях искусственного интеллекта [26].

В литературе описаны подходы к сегментации изображений деревьев в лесных массивах, однако данные подходы либо ориентированы на решение специализированных задач, связанных с дистанционным зондированием при использовании соответствующих ракурсов наблюдений (верхнебоковой) [3, 18,

19], либо требуют применения дополнительного оборудования (лидаров) [22, 27–30]. Исследования, посвященные сегментации перекрывающихся изображений деревьев по одиночным снимкам, отсутствуют. В связи с этим актуальной является задача разработки системы помощи операторам лесосечных машин, обеспечивающей возможность выделения единичных деревьев на изображении древостоя, изображений кроны и ствола для конкретного дерева, определения его породы, главных таксационных характеристик. Для решения подобных задач в настоящее время широко используются подходы на основе глубинного обучения.

Цель – разработка элементов подсистемы машинного зрения для системы поддержки принятия решений оператором лесосечных машин.

Объекты и методы исследования

Анализ 2D- и 3D-изображений лесных массивов является основой для выполнения технологическим оборудованием лесохозяйственных операций в автономном режиме [12]. Существующие системы обычно используют входные данные от лидарных датчиков для решения задач навигации, обнаружения деревьев и оценки диаметра, тогда как цифровые камеры в сочетании с алгоритмами глубокого обучения обычно решают задачи классификации видов или обнаружения лесных аномалий [9]. В любом из этих случаев недоступность данных для обучения и разнообразие лесов сдерживают применение методов глубокого обучения в целях создания автономных систем.

В области лесного хозяйства глубокое обучение привлекает большое внимание, поскольку может обеспечить решение таких проблем, как обнаружение деревьев, определение породного состава и оценка параметров древостоя [12]. Однако этот подход требует достаточного количества аннотированных изображений на этапе обучения.

В настоящее время создание обучающих наборов данных ведется в двух направлениях. Первый, наименее трудозатратный, – на основе синтетических изображений. Так, на ресурсе GitHub размещен аннотированный набор данных объемом 43 тыс. синтетических изображений. Как утверждают авторы, модель нейронной сети, обученная на этом наборе данных, достигает точности 90,4 % для обнаружения деревьев, 87,2 % для сегментации деревьев и оценки ключевых точек (см) [18]. Второй метод создания обучающих наборов данных – ручная сегментация реальных цифровых изображений. Метод является трудозатратным, и к созданию наборов данных привлекаются как IT-специалисты, так и специалисты в области лесного хозяйства. Их задача – сегментация изображений, которая заключается в фиксировании отдельно стоящих деревьев и составных частей этих деревьев, таких как ствол и крона. В ходе сегментации выполняется аннотирование изображения, т. е. присвоение каждому объекту определенных меток (названий). Решение задачи сегментации изображений осложняется тем, что некоторые породы трудно дешифрируются, если на изображении видна только часть дерева. Например, ствол осины сложно отличить от ствола березы, когда на изображении не видна крона.

Основное назначение инструментов ручной сегментации – предоставление возможностей для выделения интересующей области изображения и прикрепления к ней описания. Предварительный анализ изображений лесных массивов и

лесотаксационных материалов позволил определить количество классов объектов, которые можно отграничить при сегментации. Таким образом, исходя из возможного породного состава (18 пород) и того, что на изображении можно обозначить 2 части дерева (ствол и крону), количество классов объектов равно 36. Каждому классу объектов был присвоен свой цвет, и составлена цветовая схема для всех потенциально наблюдаемых объектов, представленная в таблице.

Цветовая схема классов объектов для сегментации изображений деревьев

The object class color scheme for segmentation of the tree images

Порода	Часть дерева	RGB цвета			Цвет
		R	G	B	
Сосна (Pinus)	Крона (Crown)	248	144	96	
	Ствол (Trunk)	248	64	0	
Ель (Picea)	Крона	208	96	200	
	Ствол	128	16	136	
Пихта (Abies)	Крона	184	168	232	
	Ствол	64	32	136	
Кедр (Pinus_sibirica)	Крона	248	88	88	
	Ствол	224	0	0	
Лиственница (Larix)	Крона	176	104	104	
	Ствол	128	0	0	
Дуб (Quercus)	Крона	128	128	128	
	Ствол	72	72	72	
Ясень (Fraxinus)	Крона	128	160	192	
	Ствол	0	0	128	
Клен (Acer)	Крона	208	144	160	
	Ствол	184	32	40	
Вяз (Ulmus)	Крона	224	112	0	
	Ствол	128	64	0	
Береза (Betula)	Крона	128	200	248	
	Ствол	24	144	248	
Осина (Populus_tremula)	Крона	136	200	160	
	Ствол	0	136	80	
Ольха серая (Alnus_incana)	Крона	168	240	72	
	Ствол	80	136	0	
Ольха черная (Alnus_glutimosa)	Крона	200	176	216	
	Ствол	112	32	168	
Липа (Tilia)	Крона	240	232	144	
	Ствол	248	248	0	
Ива (Salix)	Крона	248	160	208	
	Ствол	248	48	152	
Черемуха (Padus)	Крона	248	152	56	
	Ствол	200	96	0	
Рябина (Sorbus)	Крона	248	96	88	
	Ствол	184	16	0	
Яблоня (Malus)	Крона	248	152	136	
	Ствол	240	48	8	

Примечание: В скобках приведено название в программе; R – красный; G – зеленый; B – синий.

Для сегментации изображений использовался компьютерный графический редактор Gimp. В редакторе для каждого объекта отдельного дерева создавался слой, в котором обводились контуры данного объекта. Далее выполнялась заливка контура согласно цветовой схеме из таблицы. Так как изображение дерева состоит из объектов двух классов (ствол и крона), то для сегментации одного дерева может потребоваться создание двух слоев. При этом достаточно часто на изображении можно выделить только одну часть дерева: ствол или крону. Каждому создаваемому слою присваивалось имя, которое формировалось по следующему правилу: <порода>-<часть дерева>-<номер дерева>, например: Pinus-Trunk-5. Таким образом сформирован файл проекта, состоящий из набора слоев, количество которых соответствует количеству выделенных фрагментов. Данный подход позволяет видеть на изображении объекты, частично перекрываемые другими. На практике очень часто встречается перекрытие фрагмента ствола кроной того же или рядом растущего дерева, редко – стволом соседнего дерева. Пример создания многослойного файла представлен на рис. 1.









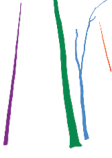




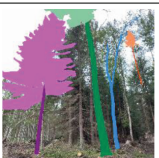
Часть дерева	Сосна	Ель	Береза	Осина	Все породы
Крона					
Ствол					
Дерево целиком					

Рис. 1. Послойное представление результатов сегментации

Fig. 1. The layered representation of segmentation results

В соответствии с описанной методикой была выполнена сегментация 13 изображений. Среднее количество пород на изображении – 2. Всего выделено 134 дерева 4 различных пород. Максимальное количество пород на изображении – 4. Минимальное – 1. На других подобных изображениях представлены чистые еловые насаждения, объем таких изображений составил 25 % от общего количества. Распределение числа деревьев по породам имеет следующий вид: сосна – 51 %; ель – 37 %; береза – 11 % и осина – 1 %. Общее количество выделенных объектов составило 196 шт., из них стволы – 61 %, кроны – 39 %. Полученные в ходе сегментации изображений данные не противоречат материалам лесоустройства по исследуемому участку. Пример сегментированного изображения приведен на рис. 2.



Рис. 2. Результаты ручной сегментации изображения

Fig. 2. The results of manual image segmentation

Исходя из результатов сегментации, возможно допустить применение автоматической сегментации деревьев (автоматического анализа изображений) в ходе проведения государственной инвентаризации лесов при условии, что апробация машинного обучения покажет высокую точность.

Как правило, в качестве результатов сегментации выступают пиксельные маски сегментированных объектов на изображении, цветные пиксели которых соответствуют сегментированным объектам, а черные – фону. В области компьютерного зрения и обработки изображений наибольшее распространение получили семантическая сегментация и сегментация экземпляров. Различия между ними демонстрирует рис. 3. Все 3 объекта на рис. 3, *а* – люди. При семантической сегментации все они будут классифицироваться как единый экземпляр. На рис. 3, *б* также есть 3 объекта (люди), но здесь разные объекты одного и того же класса выделены как разные экземпляры. Это пример сегментации экземпляров. Поскольку в нашей работе требуется, чтобы различные породы деревьев и их элементы отмечались как отдельные объекты, мы будем использовать сегментацию экземпляров.

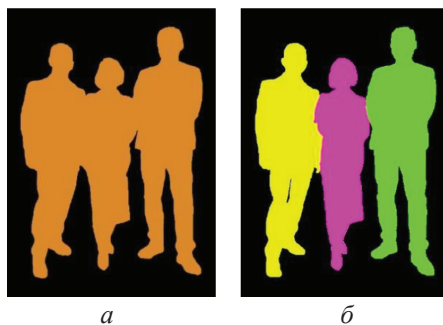


Рис. 3. Различия между семантической сегментацией (*а*) и сегментацией экземпляров (*б*) на изображении

Fig. 3. The differences between semantic segmentation (*a*) and instance segmentation (*b*) in the photograph

Наилучшую точность сегментации в настоящее время показывают различные методы глубокого обучения [11, 16, 20]. Среди преимуществ, которые предоставляют глубокие нейронные сети для сегментации объектов на изображении, следует отметить способность к обработке перекрывающихся объектов [6]. При этом одни из лучших результатов сегментации экземпляров

объектов на изображении показывают нейронные сети на базе современной архитектуры Mask R-CNN [14]. Они могут обнаруживать и сегментировать каждый объект независимо от наличия перекрытия, что делает их эффективными для решения задач сегментации в сложных сценах, поэтому данная архитектура нейронной сети была выбрана нами для сегментации частей деревьев.

Для обучения подобных нейронных сетей используются «большие» базы данных, в частности Microsoft COCO dataset [7], которая содержит более 200 000 изображений, охватывающих 80 различных классов объектов, таких как люди, автомобили, животные и различные предметы. Каждое изображение сопровождается подробной разметкой, включающей в себя координаты ограничивающего прямоугольника для конкретного объекта на изображении, класс объекта и маску сегментации, описывающую пиксели объекта. Такая база данных позволяет осуществлять обучение нейронных сетей сегментации объектов и далее проводить их дообучение на собственной небольшой базе данных изображений, содержащей объекты новых классов. Эта база данных также была использована нами для первичного обучения модели.

Алгоритм сегментации с использованием нейронной сети Mask R-CNN упрощенно можно представить в виде диаграммы, изображенной на рис. 4.



Рис. 4. Алгоритм получения бинарных масок объектов с помощью нейронной сети Mask R-CNN

Fig. 4. The algorithm for obtaining binary masks of the objects using the Mask R-CNN neural network

На шаге 1 для вычисления локальных признаков объектов используется сверточная нейронная сеть. При этом могут применяться нейронные сети различной архитектуры, такие как CNN ResNet-50/101, Feature Pyramid Network (FPN) или VGG16 [21]. Данные нейронные сети состоят из слоев свертки и слоев субдискретизации, работа которых иллюстрируется рис. 5.

В сверточном слое ко входному или к промежуточному многоканальному изображению применяются несколько операций 3-мерной свертки в формате построчной развертки. Как правило, размерность ядра свертки по высоте и ширине выбирается равной 3×3 или 7×7 пикс., при этом количество каналов яркости в ядре совпадает с числом каналов яркости обрабатываемого изображения. Результат каждой свертки сохраняется в отдельный канал изображения [2]. Например, если к цветному RGB-изображению, поданному на вход нейронной сети, применяется 32 операции свертки, то на выходе данного сверточного слоя будет 32-канальное изображение. На начальном этапе работы нейронной сети

значения в узлах ядер свертки выбираются случайным образом и корректируются с целью минимизации ошибки при обучении нейронной сети с помощью метода обратного распространения ошибки [1, 4, 5, 11, 13, 25]. Также следует отметить, что яркость изображения, полученная на выходе сверточного слоя, подвергается преобразованию с помощью нелинейной функции активации, в частности, в архитектуре VGG16 используется функция ReLu [8].

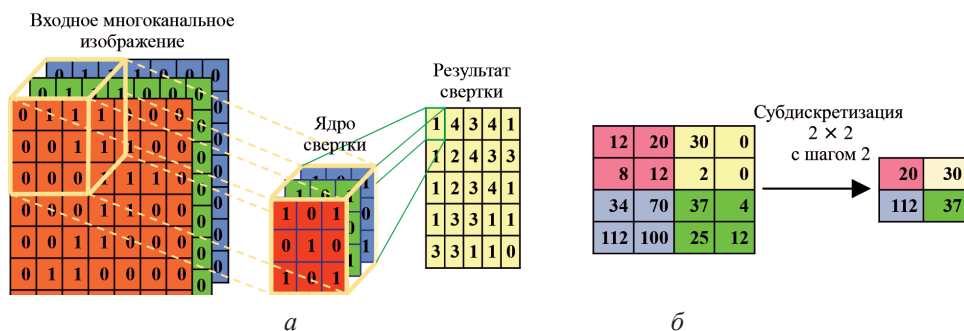


Рис. 5. Пример свертки с 3-мерным ядром (а) и операции субдискретизации изображения (б)
Fig. 5. An example of a 3D convolution (a) and an example of image subsampling operation (б)

Операция субдискретизации заключается в разбиении каждого канала изображения на равные области и выборе максимальной яркости из каждой области (рис. 5, б) [6, 23]. Пересечение областей изображения контролируется шагом субдискретизации: если он равен ширине и высоте области изображения, то перекрытия не происходит. Как правило, разрешение изображения на выходе слоя субдискретизации уменьшается в 2 раза, что видно из рис. 5, б. Схематичное изображение архитектуры нейронной сети VGG16, иллюстрирующее последовательность операций свертки и субдискретизации, представлено на рис. 6.

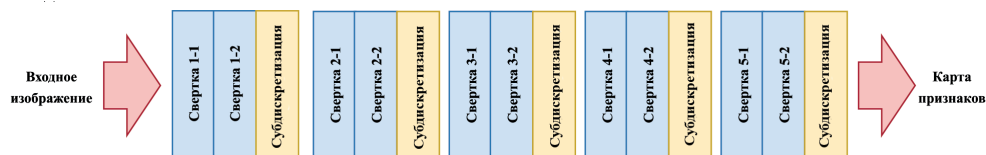


Рис. 6. Архитектура нейронной сети VGG16

Fig. 6. The VGG16 neural network architecture

На следующем шаге работы нейронной сети Mask R-CNN осуществляется поиск минимальных прямоугольных областей на изображении, содержащих объекты, которые надо сегментировать. Для этого используется метод Region Proposal Network (RPN) [10, 24]. В полученной карте признаков с выхода сверточной нейронной сети наращивается рецептивное поле с помощью сверточного слоя с ядрами $3 \times 3 \times 512$ пикс. При этом итоговая матрица не меняется в размерах. Далее с использованием сгенерированной карты признаков происходит поиск регионов, где могут содержаться объекты. В классическом варианте метода RPN для генерации карты признаков используется сверточная нейронная сеть VGG16.

Карта признаков изображения, полученная на выходе нейронной сети VGG16, в 16 раз меньше изначального размера изображения, при этом имеет 512 каналов яркости. Работа метода RPN основана на том, что каждой ячейке карты признаков может соответствовать область на исходном изображении, содержащая какой-либо объект. При определении размеров этой области для каждой ячейки карты признаков принимается $k = 9$ гипотез, что объект на исходном изображении находится в пределах области (якоре) размером 128×128 , 128×256 , 256×128 , 256×256 , 512×256 , 256×512 , 512×512 , 512×1024 , 1024×512 пикс. Обозначим данное множество гипотез как $H = \{h_0, h_1, \dots, h_8\}$, также отметим, что одной ячейке карты признаков на выходе нейронной сети VGG16 соответствует вектор размерности $c = 512$. Далее каждая ячейка карты признаков подается на два отдельных сверточных слоя с ядром свертки 1×1 пикс. и количеством выходных каналов \hat{c} .

Первый слой имеет параметр $\hat{c} = 2k$ и необходим для определения вероятности наличия или отсутствия какого-либо объекта для каждой гипотезы. То есть на выходе слоя для каждой гипотезы из множества H имеем пару значений вероятностей наличия и отсутствия объекта в якоре, размеры которого соответствуют текущей гипотезе h .

Второй слой имеет параметр $\hat{c} = 4k$ и необходим для вычисления точных параметров области, ограничивающей объект на изображении. На выходе данного слоя для каждой гипотезы из множества H имеем 4 значения – x, y, h, w (x, y – координаты центра области, четко ограничивающие объект на изображении по осям абсцисс и ординат соответственно; h, w – высота и ширина области соответственно).

Одна из гипотез из множества H принимается по максимальному значению вероятности наличия объекта внутри якоря, полученного на выходе первого сверточного слоя. При этом используется пороговое значение. Если вероятность наличия объекта меньше 0,7, то считается, что текущей ячейке из карты признаков не соответствует ни одна область с объектом на входном изображении. Точные параметры области, ограничивающей объект на входном изображении, берутся с выхода второго сверточного слоя для гипотезы с максимальной вероятностью. Упрощенная схема работы метода RPN представлена на рис. 7.

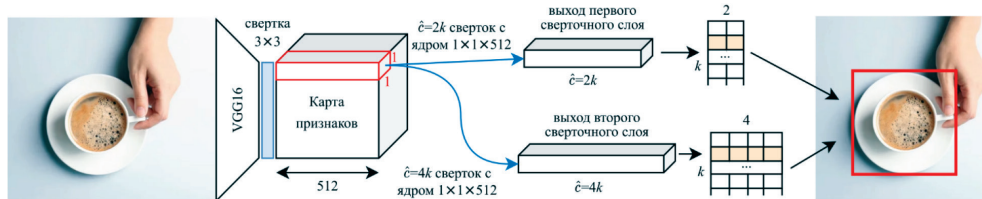


Рис. 7. Упрощенная схема работы метода RPN

Fig. 7. A simplified diagram of the RPN method

В нашем варианте нейронной сети Mask R-CNN для построения карты признаков изображения использована сверточная нейронная сеть FPN с параметрами, как в работе [31]. Карта признаков изображения, полученная на выходе данной нейронной сети, имеет в отличие от VGG16 256 каналов яркости.

Для классификации объектов изображения и выделения их бинарных масок на карте его признаков, полученной с выхода сверточной нейронной сети, обозначаются прямоугольные области, соответствующие прямоугольным областям на исходном изображении. Эти области приводятся к фиксированному размеру пикселей с помощью процедуры RoIAlign [14], в основе которой лежит билинейная интерполяция изображения. Модифицированные области поступают на вход последовательной нейронной сети для классификации объекта и на вход сверточной нейронной сети для генерации маски. Данный процесс иллюстрируется рис. 8.

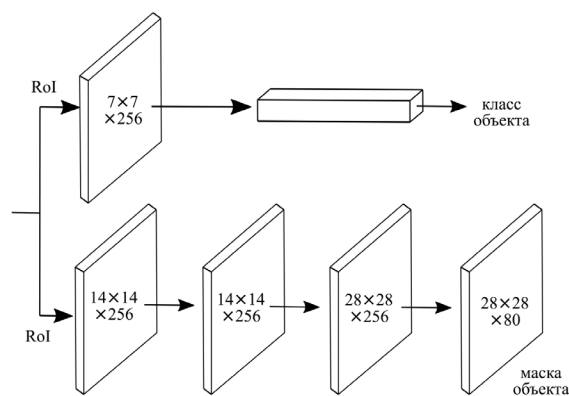


Рис. 8. Схема процесса генерации маски и классификации объекта в нейронной сети Mask R-CNN

Fig. 8. The diagram of the mask generation and object classification process in the Mask R-CNN neural network

Последовательная нейронная сеть содержит один слой, количество нейронов в котором равно числу распознаваемых классов, сверточная нейронная сеть имеет параметры, представленные на рис. 8.

Результаты исследования и их обсуждение

Мы использовали готовые весовые коэффициенты нейронной сети Mask R-CNN для библиотеки TensorFlow, обученной на базе данных Microsoft COCO dataset [7, 17]. К 80 классам объектов, для распознавания которых предназначена данная нейронная сеть, было добавлено еще 4: ствол лиственного дерева, ствол хвойного дерева, крона лиственного дерева, крона хвойного дерева. Дообучение нейронной сети осуществлялось с применением нашей базы данных, содержащей 13 размеченных изображений. При этом за счет различных преобразований изображений этот набор данных был увеличен в 11 раз. Каждое изображение зеркально отражалось по вертикали и горизонтали, поворачивалось на 45° и 90° , масштабировалось с коэффициентом масштабирования 0,5 по вертикали и 1,5 по горизонтали, подвергалось передискретизации с пропуском 1 пикс. по оси абсцисс, переводилось в градации серого, у всех изображений уменьшался оттенок и увеличивалась насыщенность – на 20 [15], случайным образом изменялась яркость в каждом канале в диапазоне от –10 до 10 по сравнению с исходным значением, инвертировались цветочные каналы, повышалась резкость. В итоге было получено 143 изображения, 110 из которых использовались для обучения нейронной сети, а 33 – для проверки. Коэффициенты нейронной сети корректировались всякий раз после подачи 5 изображений, нейронная сеть обучалась на 300 эпохах. Графики обучения нейронной сети представлены на рис. 9.

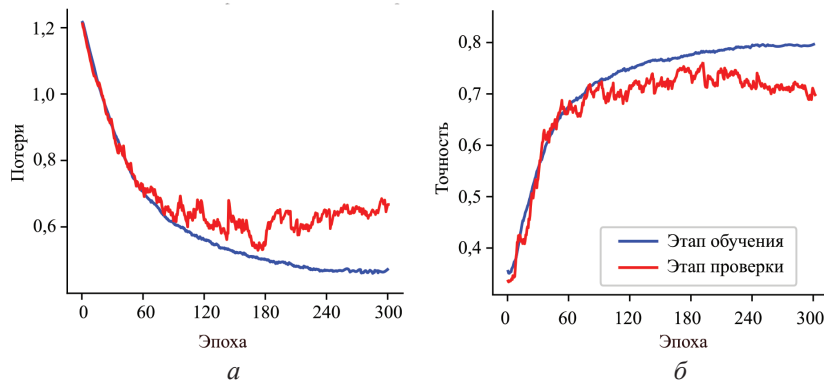


Рис. 9. Потери (а) и точность (б) нейронной сети Mask R-CNN при обучении сегментации перекрывающихся объектов на цифровых снимках лесных массивов

Fig. 9. The losses (a) and accuracy (б) of the Mask R-CNN neural network in training segmentation of overlapping objects in the digital photographs of forest areas

Из рис. 9 видно, что переобучение нейронной сети наступает в районе 240 эпохи, поскольку при дальнейшем обучении точность нейронной сети на этапе проверки не растет, а ошибка начинает увеличиваться. При этом точность сегментации объектов деревьев составила 79 %. Результат сегментации перекрывающихся объектов на цифровых снимках лесных массивов представлен на рис. 10.

Рис. 10. Результат сегментации перекрывающихся объектов на цифровых снимках лесных массивов

Fig. 10. The result of segmentation of overlapping objects in the digital photographs of forest areas



Заключение

Показана возможность использования нейронной сети Mask R-CNN для сегментации изображений элементов деревьев, перекрывающих друг друга, на цифровых снимках лесных массивов. При этом кроме стандартной процеду-

ры процесса сегментации объектов с помощью нейронной сети Mask R-CNN, включающей вычисление признаков изображения, обнаружение объектов, их классификацию и сегментацию, важным этапом является предварительное обучение модели на базе данных большого объема. В нашем случае в качестве подобной базы данных выступил набор изображений Microsoft COCO dataset. Предварительно обученная нейронная сеть Mask R-CNN позволяет осуществлять ее дообучение на собственной небольшой базе данных изображений, содержащей объекты новых классов. Для дообучения нейронной сети Mask R-CNN нами было использовано 143 изображения, при этом точность сегментации изображений стволов и крон хвойных и лиственных пород деревьев, перекрывающих друг друга, составила 79 %.

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

1. Иванов К.О. Классификация фрагментов электроэнцефалограммы по степени патологической значимости с помощью последовательных нейронных сетей // Вестн. ПГТУ. Сер.: Радиотехн. и инфокоммуникац. системы. 2022. № 4(56). С. 24–37.

Ivanov K.O. Classification of Electroencephalogram Segments Based on Pathological Significance Using Sequential Neural Networks. *Vestnik Povolzhskogo gosudarstvennogo tekhnologicheskogo universiteta. Seriya: Radiotekhnicheskie i infokommunikatsionnye sistemy* = Vestnik of Volga State University of Technology. Series: Radio Engineering and Infocommunication Systems, 2022, no. 4(56), pp. 24–37. (In Russ.). <https://doi.org/10.25686/2306-2819.2022.4.24>

2. Николенко С., Кадуринов А., Архангельская Е. Глубокое обучение. Погружение в мир нейронных сетей. СПб.: Питер, 2022. 480 с.

Nikolenko S., Kadurin A., Arkhangelskaya E. *Deep Learning. Dive into the World of Neural Networks*. St. Petersburg, Piter Publ., 2022. 480 p. (In Russ).

3. Портнов А.М., Иванова Н.В., Шашков М.П. Опыт использования нейронной сети DeepForest для детектирования деревьев в широколиственном лесу // Математическая биология и биоинформатика: докл. Междунар. конф. / под ред. В.Д. Лахно. Т. 9. Пушино: ИМПБ РАН, 2022. Ст. № e45.

Portnov A.M., Ivanova N.V., Shashkov M.P. Experience of Using the DeepForest Neural Network to Detect Trees in a Broad-Leaved Forest. *Proceedings of the International Conference "Mathematical Biology and Bioinformatics"*. Ed. V.D. Lakhno. Pushchino, IMPB RAS, 2022, vol. 9, art. no. e45. (In Russ). <https://doi.org/10.17537/icmbb22.12>

4. Тимофеев В.С., Сивак М.А. Робастная нейронная сеть с простой архитектурой // Сиб. журн. индустр. математики. 2021. Т. 24, № 4. С. 126–138.

Timofeev V.S., Sivak M.A. Robust Neural Network with Simple Architecture. *Sibirskiy zhurnal industrialnoy matematiki* = Journal of Applied and Industrial Mathematics, 2021, vol. 24, no. 4, pp. 126–138. (In Russ). <https://doi.org/10.33048/SIBJIM.2021.24.409>

5. Callan R. *The Essence of Neural Network*. Upper Saddle River: Prentice Hall Europe, 1998. 232 p.

6. Chollet F. *Deep Learning with Python: Second Edition*. New York, Manning Publications Co., 2021. 504 p.

7. COCO: Common Objects in Context. Available at: <https://cocodataset.org/#home> (accessed 17.01.23).

8. Daubechies I., DeVore R., Foucart S., Hanin B., Petrova G. Nonlinear Approximation and (Deep) ReLU Networks. *Constructive Approximation*, 2022, vol. 55, pp. 127–172. <https://doi.org/10.1007/s00365-021-09548-z>

9. Diez Y., Kentsch S., Fukuda M., Caceres M.L.L., Moritake K., Cabezas M. Deep Learning in Forestry Using UAV-Acquired RGB Data: A Practical Review. *Remote Sensing*, 2021, vol. 13, no. 14, art. no. 2837. <https://doi.org/10.3390/rs13142837>
10. Girshick R. Fast R-CNN. *2015 IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 1440–1448.
11. Goodfellow I., Benigo Y., Courville A. *Deep Learning*. Cambridge, MIT Press Publ., 2017. 785 p.
12. Grondin V., Fortin J.-M., Pomerleau F., Giguère P. Tree Detection and Diameter Estimation Based on Deep Learning. *Forestry: An International Journal of Forest Research*, 2023, vol. 96, iss. 2, pp. 264–276. <https://doi.org/10.1093/forestry/cpac043>
13. Haykin S. *Neural Networks: A Comprehensive Foundation (3rd Edition)*. Hoboken: Prentice-Hall Inc., 2007. 938 p.
14. He K., Gkioxari G., Dollár P., Girshick R. Mask R-CNN. *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988.
15. *Imgaug*: a Library for Image Augmentation in Machine Learning Experiments. Available at: <https://imgaug.readthedocs.io/en/latest/> (accessed 17.07.23).
16. Ivanov K.O., Kazarinov A.A., Rozhentsov A.A., Dubrovin V.N., Baev A.A., Evdokimov A.O. An Algorithm for Segmentation of Kidney Tissues on CT Images Based on a U-Net Convolutional Neural Network. *Biomedical Engineering*, 2023, vol. 56, pp. 424–428. <https://doi.org/10.1007/s10527-023-10249-z>
17. *Mask R-CNN for Object Detection and Segmentation*. Available at: https://github.com/matterport/Mask_RCNN (accessed 17.07.23).
18. Moussaid A., El Fkihi S., Zennayi Y. Tree Crowns Segmentation and Classification in Overlapping Orchards Based on Satellite Images and Unsupervised Learning Algorithms. *Journal of Imaging*, 2021, vol. 7, no. 241. <https://doi.org/10.3390/jimaging7110241>
19. Ochoa C.G., Caruso P., Ray G., Deboodt T., Jarvis T.W., Guldán S.J. Ecohydrologic Connections in Semiarid Watershed Systems of Central Oregon USA. *Water*, 2018, vol. 2, no. 10, art. no. 181. <https://doi.org/10.3390/w10020181>
20. Pointer I. *Programming PyTorch for Deep Learning: Creating and Deploying Deep Learning Applications*. O'Reilly Media Inc., 2019. 217 p.
21. Qassim H., Verma A., Feinzimer D. Compressed Residual-VGG16 CNN Model for Big Data Places Image Recognition. *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)*, 2018, pp. 169–175.
22. Qianwei L., Weifeng M., Jianpeng Z., Yicheng L., Dongfan X., Jinliang W. Point-Cloud Segmentation of Individual Trees in Complex Natural Forest Scenes Based on a Trunk Growth Method. *Journal of Forestry Research*, 2021, vol. 32, pp. 2403–2414. <https://doi.org/10.1007/s11676-021-01303-1>
23. Rao D., McMahan B. *Natural Language Processing with PyTorch: Build Intelligent Language Applications Using Deep Learning*. O'Reilly Media Inc., 2019. 254 p.
24. Ren S., Kaiming H., Ross G., Jian S. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *NIPS'15: Proceedings of the 28th International Conference on Neural Information Processing Systems*. Great Britain, Oxford, December, 2015, vol. 1, pp. 91–99.
25. Rojas R. The Backpropagation Algorithm. *Neural Networks*, 1996, pp. 149–182. https://doi.org/10.1007/978-3-642-61068-4_7
26. Tarasova A., Strielkowski W., Larionova N. Digital Relations and Innovative Ecosystem for Sustainable Engineering Education: A Case of Russia. *Increasing Supply Chain Performance in Digital Society*. IGI Global, 2022, pp. 304–318.
27. Wang D., Liang X., Mofack II G., Martin-Ducup O. Individual Tree Extraction from Terrestrial Laser Scanning Data via Graph Pathing. *Forest Ecosystems*, 2021, vol. 8, no. 67. <https://doi.org/10.1186/s40663-021-00340-w>

28. Weinstein B.G., Graves S.J., Marconi S., Singh A., Zare A., Stewart D. A Benchmark Dataset for Canopy Crown Detection and Delineation in Co-Registered Airborne RGB, LiDAR and Hyperspectral Imagery from the National Ecological Observation Network. *PLoS Computational Biology*, 2021, vol. 7, no. 17. <https://doi.org/10.1371/journal.pcbi.1009180>

29. Windrim L., Bryson M. Detection, Segmentation, and Model Fitting of Individual Tree Stems from Airborne Laser Scanning of Forests Using Deep Learning. *Remote Sensing*, 2020, vol. 12, no. 9, art. no. 1469. <https://doi.org/10.3390/rs12091469>

30. Xu X., Iuricich F., Floriani L.D. A Topology-Based Approach to Individual Tree Segmentation from Airborne LiDAR Data. *GeoInformatica*, 2023, vol. 27, pp. 759–788. <https://doi.org/10.1007/s10707-023-00487-4>

31. Yang Z., Dong R., Xu H., Gu J. Instance Segmentation Method Based on Improved Mask R-CNN for the Stacked Electronic Components. *Electronics*, 2020, vol. 9, no. 6, art. no. 886. <https://doi.org/10.3390/electronics9060886>

Конфликт интересов: Авторы заявляют об отсутствии конфликта интересов
Conflict of interest: The authors declare that there is no conflict of interest

Вклад авторов: Все авторы в равной доле участвовали в написании статьи
Authors' Contribution: All authors contributed equally to the writing of the article