
ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

DOI:10.25743/ICT.2024.29.5.008

Семантическая сегментация осиновых и березовых древостоев на RGB-изображениях с БПЛА при помощи сверточных нейронных сетей

И. А. ПЕСТУНОВ^{1,2,*}, Р. А. КАЛАШНИКОВ¹, С. А. РЫЛОВ^{1,3}

¹Федеральный исследовательский центр информационных и вычислительных технологий,
630090, Новосибирск, Россия

²Никитский ботанический сад — Национальный научный центр РАН, 298648, Ялта, Россия

³Хакасский государственный университет им. Н.Ф. Катанова, 655017, Абакан, Республика
Хакасия, Россия

*Контактный автор: Пестунов Игорь Алексеевич, e-mail: pestunov@ict.nsc.ru

Поступила 14 декабря 2023 г., доработана 05 февраля 2024 г., принята в печать 12 февраля 2024 г.

Исследуется возможность автоматического выделения и классификации осиновых и березовых насаждений на RGB-изображениях сверхвысокого пространственного разрешения, полученных с беспилотного летательного аппарата. Для решения задачи семантической сегментации проведено сравнение работы сверточных нейронных сетей на основе различных архитектур: U-Net, FPN, PSPNet, Linknet, DeepLabV3, DeepLabV3+. Результаты показали, что архитектуры DeepLabV3+ и модификации U-Net с Inception-блоками позволяют получить наилучшие результаты семантической сегментации, достигая наиболее высоких значений метрик качества IoU (~0.83) и F-score (~0.91).

Ключевые слова: обработка изображений, беспилотная съемка, семантическая сегментация, классификация, БПЛА, выделение древостоев, сверточная нейронная сеть.

Цитирование: Пестунов И.А., Калашников Р.А., Рылов С.А. Семантическая сегментация осиновых и березовых древостоев на RGB-изображениях с БПЛА при помощи сверточных нейронных сетей. Вычислительные технологии. 2024; 29(5):100–112. DOI:10.25743/ICT.2024.29.5.008.

Введение

Проблема определения видового состава лесных насаждений по данным дистанционного зондирования уже на протяжении нескольких десятилетий привлекает внимание многих ученых из разных стран [1], и в последние годы интерес к этой теме неуклонно растет [2–10]. Актуальность этой проблемы обусловлена прежде всего постоянным совершенствованием и развитием средств и технологий дистанционного зондирования, включая интенсивное распространение в последние годы технологий беспилотной съемки.

Результаты, которые представлены в работах [7, 11–16], убедительно показывают, что высокая точность определения породного состава лесных насаждений может быть достигнута при использовании лидарных данных совместно с данными гиперспектральной съемки высокого пространственного разрешения. Однако системы, построенные на основе гиперспектральных данных, имеют лишь ограниченное практическое примене-

ние по следующим основным причинам. Во-первых, высокая стоимость гиперспектральных камер делает практическое использование таких систем экономически неэффективным для большинства малых и средних предприятий, занимающихся лесоустройством или лесозаготовками. Во-вторых, такие системы не позволяют обеспечить устойчивую надежность результатов, потому что они полагаются на информацию о спектральной отражательной способности крон деревьев, которая определяется не только породой дерева, но и целым рядом самых разнообразных факторов, например временем года, регионом, воздействием теней, зависит от погоды, освещенности, состояния почвы и подстилки, плотности листьев и др. Поэтому в последние годы проводятся исследования [2, 17, 18], результаты которых позволили бы создать системы определения видового состава древостоя на основе RGB-изображений сверхвысокого пространственного разрешения (2–10 см/пиксель), полученных с беспилотных летательных аппаратов (БПЛА) с помощью недорогих цифровых фотокамер.

Дистанционное зондирование на базе малых беспилотных летательных аппаратов является быстро развивающейся технологией [19]. С помощью БПЛА можно получать гибкое временное разрешение и чрезвычайно высокое (до нескольких сантиметров) пространственное разрешение изображений, на которых можно увидеть особенности дерева на уровне ветвей и даже листьев. Поэтому при распознавании пород деревьев на изображениях сверхвысокого пространственного разрешения значительную роль играют не только спектральные признаки, но и пространственные (текстурные и геометрические). Совместное использование спектральных и пространственных признаков в рамках традиционных подходов к анализу и распознаванию изображений является сложной и к настоящему времени не до конца решенной задачей [20–23].

Наиболее эффективный подход к анализу аэрокосмических изображений высокого пространственного разрешения заключается в применении методов глубокого обучения на основе сверточных нейронных сетей (СНС) [24]. Это связано с тем, что сверточные нейронные сети специально разработаны для анализа пространственных закономерностей и при их использовании не требуется “ручного” извлечения пространственных признаков. В отличие от традиционных алгоритмов сегментации и распознавания изображений, СНС обеспечивают возможность совместного анализа спектральных и пространственных свойств объектов на изображении. Более того, СНС позволяют решать задачи семантической сегментации изображений, т. е. одновременно производятся выделение и классификация объектов [25].

Цель данной работы состоит в экспериментальном исследовании наиболее подходящих моделей сверточных нейронных сетей для семантической сегментации близких по спектральным характеристикам осиновых и березовых древостоев на RGB-изображениях, получаемых с помощью беспилотных летательных аппаратов. Результаты этих исследований могут найти практическое применение, например, при определении размера арендной платы за лесные участки, используемые для заготовки древесины. Оценка площади насаждений осуществляется специалистами путем непосредственного выезда на место для осмотра лесного участка. Автоматизация этих работ позволит сократить затраты и повысить объективность получаемых оценок.

1. Объект исследования и исходные данные

В качестве объекта исследования выбран участок смешанного леса в районе п. Аламбай Алтайского края площадью около 4 км², на котором главными лесооб-



Рис. 1. Пример исходного RGB-изображения размером 6000×4000 пикселей
Fig. 1. An example of the original RGB image with the size of 6000×4000 pixels

разующими породами являются береза (*betula pendula*, *betula alba*) и осина (*populus tremula*).

Исходными данными для обработки служили 68 RGB-изображений размером 6000×4000 пикселей с пространственным разрешением порядка 5 см, полученных с помощью БПЛА. Съемка производилась с высоты 250 м. Пример исходного RGB-изображения представлен на рис. 1. Съемка выполнена ООО “Беспилотные технологии” в апреле 2021 г. Месяц апрель для съемки был выбран не случайно. В этот период на ветвях деревьев нет ни листьев, ни снега, ни изморози и осиновые и березовые насаждения можно легко дешифрировать. На RGB-снимках, полученных в летний вегетационный период, выделить и распознать близкие по спектральным характеристикам березовые и осиновые древостоя представляется более сложной задачей.

В настоящей работе семантическая сегментация выполнялась на три класса: “березовые насаждения”, “осиновые насаждения” и “фон”.

2. Архитектуры сверточных нейронных сетей

Для семантической сегментации используется целый ряд различных архитектур сверточных нейронных сетей [24]. В данной работе для исследований использованы следующие шесть архитектур: U-Net, DeepLabV3, DeepLabV3+, FPN, LinkNet, PSPNet, а также предлагаемая нами модификация U-Net с Inception-блоками.

В последнее время для решения различных прикладных задач широко применяется трансферное обучение (transfer learning). Предобученные модели позволяют использовать готовые карты признаков, сформированные при решении других задач. В данной работе используется семейство предобученных сетей ResNet. Они представляют собой тип глубоких нейронных сетей, которые разработаны для решения проблемы затухающего градиента. Эта проблема часто возникает при создании глубоких нейронных сетей с большим числом сверток.

Существует несколько модификаций архитектур ResNet: ResNet-18, ResNet-34, ResNet-50, ResNet-101, ResNet-152. Проведенные исследования показали, что использование достаточно глубоких энкодеров применительно к данной задаче приводит к пе-

реобучению сети и ухудшению качества получаемых результатов. Поэтому в дальнейшем в качестве энкодера во всех шести исследуемых нами архитектурах использовалась сверточная сеть ResNet-50, которая обеспечивает приемлемое качество сегментации при относительно небольших вычислительных затратах.

Использование в исследовании архитектуры U-Net обусловлено тем, что она хорошо зарекомендовала себя при решении задач семантической сегментации лесных насаждений [4, 5]. Изначально U-Net разработана для сегментации биомедицинских изображений [25], где продемонстрировала высокое качество результатов. Такая архитектура позволяет комбинировать как низко-, так и высокоуровневые признаки.

DeepLabV3 — семейство моделей сегментации изображений, разработанных компанией Google. Отличительной особенностью этих архитектур является наличие расширенных сверток (atrous or dilated convolutions), которые помогают увеличить поле зрения сети и учитывать объекты разных размеров без увеличения вычислительных затрат и потери качества изображения. Использование механизма ASPP (atrous spatial pyramid pooling) позволяет агрегировать признаки на разных масштабах и, соответственно, лучше учитывать контекст объектов на изображениях.

DeepLabV3+ является расширением и улучшением архитектуры DeepLabV3, которое внедряет энкодер-декодерную структуру. Энкодер формирует карты признаков и уменьшает размер изображения, а декодер восстанавливает исходное разрешение входных данных.

Архитектуры FPN и PSPNet используют пирамидальную структуру признаков для интеграции информации на разных масштабах, но FPN является более универсальной, так как подходит для семантической сегментации и обнаружения объектов, а PSPNet специализируется на задачах анализа контекста сцен.

LinkNet — архитектура сверточной нейронной сети, разработанная для задач семантической сегментации, в которых требуется высокая точность и эффективность. Она также построена на основе энкодер-декодерной структуры с использованием расширенных сверток для увеличения поля зрения сети и учета объектов разных размеров. Основным фокусом данной архитектуры является эффективная сегментация изображений при как можно меньшем числе параметров сети, что позволяет применять архитектуру в условиях ограниченных вычислительных мощностей.

Кроме рассмотренных выше архитектур для исследования предлагается модификация архитектуры U-Net с Inception-блоками, схема которой представлена на рис. 2. Далее будем называть ее U-Net-M. На этом рисунке использованы следующие условные обозначения:

- Input — входной слой;
- Conv2D — операция свертки;
- MaxPooling — операция субдискретизации выбором максимального значения в окрестности 2×2 пикселей;
- AvgPooling — операция субдискретизации выбором среднего значения в окрестности 2×2 пикселей;
- Softmax — выходной слой с функцией активации Softmax;
- Up-sampling — операция увеличения масштаба методом ближайшего соседа;
- Concatenation — конкатенация тензоров;
- BN + ReLu — использование пакетной нормализации с функцией активации ReLu;
- x2 (x4) — двукратное (четырехкратное) использование операций с тензорами либо использование блока из наборов операций (при наличии фигурной скобки).

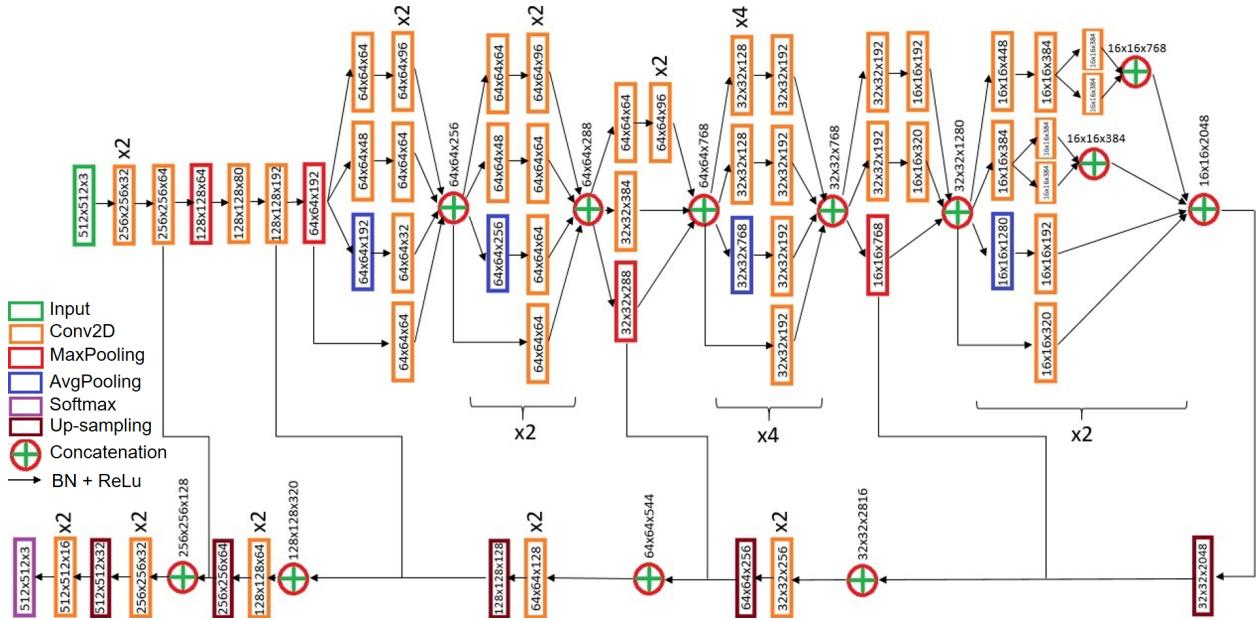


Рис. 2. Схема архитектуры сверточной нейронной сети U-Net-M

Fig. 2. Schematic diagram of the U-Net-M convolutional neural network architecture

Архитектура состоит из двух соединенных между собой сетей: энкодера для извлечения из изображения семантической информации в виде вектора признаков и декодера для превращения вектора признаков в матрицу нового изображения — маски классов. Энкодер представляет собой обычную сверточную сеть, состоящую из повторяющихся Inception-блоков со слоями активации ReLu. После каждого слоя активации идет слой пакетной нормализации. Декодер представляет собой последовательность шагов из операций повышающей дискретизации карт признаков, за которой следует объединение с соответствующим сверточным слоем из энкодера. Последний слой декодера работает с функцией активации Softmax.

Для оценки эффективности предложенных сверточных нейронных сетей использовались две метрики: Intersection over Union (IoU) и F-score.

Метрика IoU является общепринятой метрикой эффективности при решении задачи семантической сегментации цифровых изображений [26], принимает значения из интервала $[0, 1]$, найденные по формуле

$$\text{IoU}(P, T) = \frac{|P \cap T|}{|P \cup T|},$$

где P – множество пикселей, сегментированных сетью; T – множество пикселей маски, выделенной экспертом.

Метрика F-score — взвешенное гармоническое среднее полноты (Recall) и точности (Precision). Precision, Recall и F-score вычисляются по следующим формулам:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{F-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}.$$

Здесь TP (true positive) — истинно-положительное решение, TN (true negative) — истинно-отрицательное решение, FP (false positive) — ложно-положительное решение, FN (false negative) — ложно-отрицательное решение.

В качестве функции потерь выбрана функция Дайса – Серенсена (DSL):

$$DSL = 1 - \frac{2|X \cap Y|}{|X| + |Y|},$$

где X — множество пикселей маски, выделенной экспертом; Y — множество пикселей, сегментированных сетью.

В качестве оптимизатора выбран алгоритм Adam. Обучение сети происходило с использованием высокоуровневого фреймворка PyTorch, который представляет собой гибкий и мощный способ для создания и обучения нейронных сетей. PyTorch содержит широкий набор инструментов для построения сетей, включающий всевозможные виды слоев нейронных сетей, функции активации, оптимизаторы и пр.

3. Процесс обучения

Для формирования выборок для обучения и тестирования все изображения случайным образом были разделены на три части: обучающую (50 изображений), валидационную (9 изображений) и тестовую (9 изображений).

Исходные изображения размечались вручную с использованием веб-сервиса Supervisely [27] (рис. 3), а затем нарезались на фрагменты 512×512 пикселей. Таким образом, было получено 1170 обучающих, 280 валидационных и 110 тестовых фрагментов. Далее расширение выборки происходило в полуавтоматическом режиме: сеть обучалась на имеющейся обучающей выборке, затем на ее вход подавались новые фрагменты. Если результат предсказания был удовлетворительным, то фрагмент добавлялся в обучающую выборку на следующей итерации. На рис. 4 представлена схема полу-



Рис. 3. Пример размеченного изображения

Fig. 3. An example of a labelled image

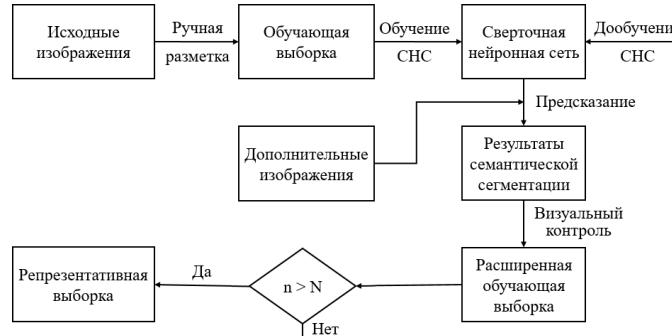


Рис. 4. Схема полуавтоматического расширения обучающей выборки

Fig. 4. Scheme of semi-automatic expansion of the training sample

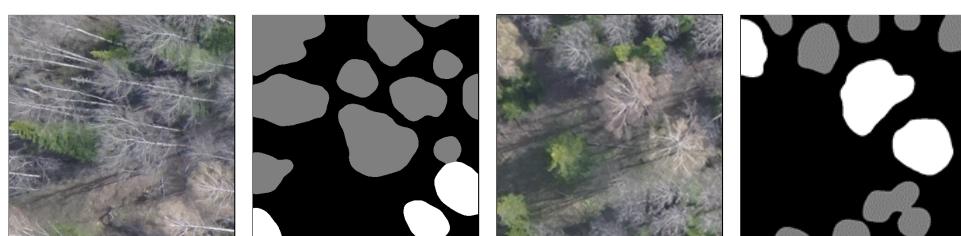


Рис. 5. Примеры пар фрагментов и масок из обучающей выборки (белым цветом представлены березы, серым — осины)

Fig. 5. Examples of pairs of fragments and masks from the training sample (birch trees are represented in white, aspen trees in gray)

автоматического расширения обучающей выборки. В результате обучающая выборка была расширена до 1560 уникальных пар фрагментов.

Для увеличения объема обучающей выборки выполнялась аугментация данных, которая включала следующие операции:

- 1) поворот фрагментов на случайно выбранный угол $\alpha \in (0, 180^\circ)$;
- 2) масштабирование на случайный процент от 0 до 10;
- 3) случайные отображения относительно горизонтальной и вертикальной осей.

В результате всех проведенных операций объем обучающей выборки был увеличен до 17 000 пар. Примеры пар фрагментов и масок из обучающей выборки представлены на рис. 5.

4. Экспериментальные исследования

Все вычисления проводились на узле кластера ФИЦ ИВТ, который обладает следующими характеристиками: HPE Apollo 6500 G10+, 2×AMD EPYC 7452 (32 ядра по 2.35 ГГц), 1 ТБ ОЗУ, 8×Nvidia A100 80 ГБ SXM4.

Для определения оптимальных параметров сверточных нейронных сетей выполнен ряд экспериментов, в ходе которых модели СНС обучались с разными настройками мини-пакета (8, 16 и 32) в течение 150 эпох с постепенно поникающейся скоростью обучения от 10^{-3} до 10^{-6} . Результаты экспериментальных исследований представлены в таблице. Значения метрик IoU и F-score приведены для трех наборов выборки: обучающей, валидационной и тестовой. Представленные в таблице вычисления были произведены на одной видеокарте.

Результаты экспериментальных исследований СНС
Experimental results of CNNs

| Архитектура сети | IoU | F-score | Число эпох | Время, мин | Память, МБ |
|------------------|-------|---------|------------|------------|------------|
| U-Net | 0.919 | 0.958 | 57 | 9.3 | 6802 |
| | 0.845 | 0.914 | | | |
| | 0.785 | 0.871 | | | |
| FPN | 0.684 | 0.811 | 62 | 17.1 | 3604 |
| | 0.662 | 0.789 | | | |
| | 0.651 | 0.762 | | | |
| PSPNet | 0.816 | 0.899 | 85 | 10.4 | 2520 |
| | 0.794 | 0.883 | | | |
| | 0.733 | 0.829 | | | |
| Linknet | 0.921 | 0.959 | 73 | 13.8 | 4544 |
| | 0.816 | 0.897 | | | |
| | 0.742 | 0.821 | | | |
| DeepLabV3 | 0.932 | 0.965 | 88 | 16.7 | 4854 |
| | 0.851 | 0.917 | | | |
| | 0.797 | 0.889 | | | |
| DeepLabV3+ | 0.949 | 0.972 | 110 | 20 | 3904 |
| | 0.878 | 0.932 | | | |
| | 0.832 | 0.907 | | | |
| U-Net-M | 0.935 | 0.966 | 123 | 30.7 | 10038 |
| | 0.873 | 0.929 | | | |
| | 0.828 | 0.905 | | | |

Анализ полученных результатов показывает, что архитектуры U-Net-M и DeepLabV3+ продемонстрировали лучшее качество семантической сегментации. На рис. 6 приведены графики зависимости метрики IoU от номера эпохи для архитектур U-Net-M и DeepLabV3+, пунктирной линией выделены эпохи с наибольшими показателями метрик на валидации.

На рис. 7 представлены примеры результатов семантической сегментации с помощью архитектур U-Net-M и DeepLabV3+. Сравнительный анализ полученных результатов семантической сегментации показывает, что осиновые и березовые насаждения

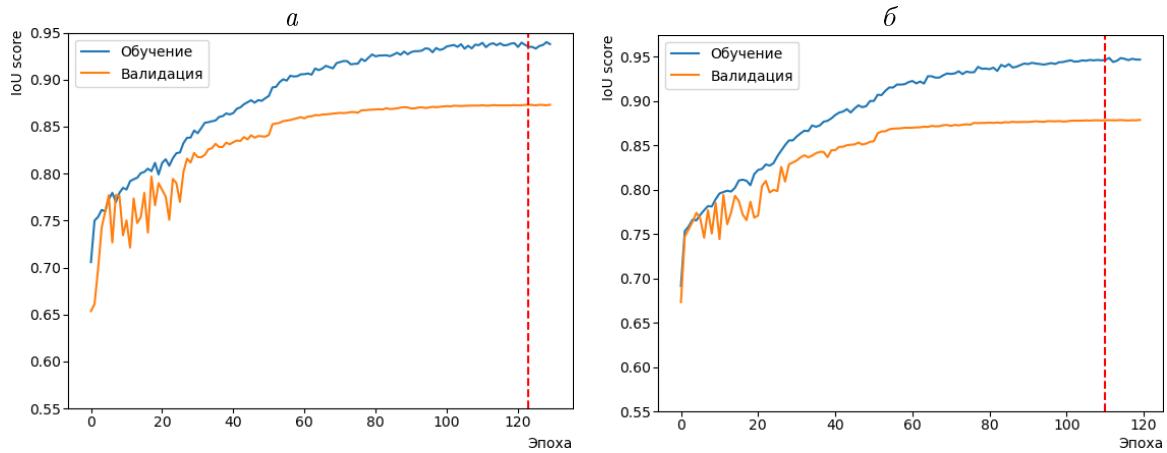


Рис. 6. Графики зависимости метрики IoU от номера эпохи для архитектур СНС U-Net-M (a) и DeepLabV3+ (b)

Fig. 6. Plots of the dependence of IoU metric on epoch number for U-Net-M (a) and DeepLabV3+ (b) CNN architectures

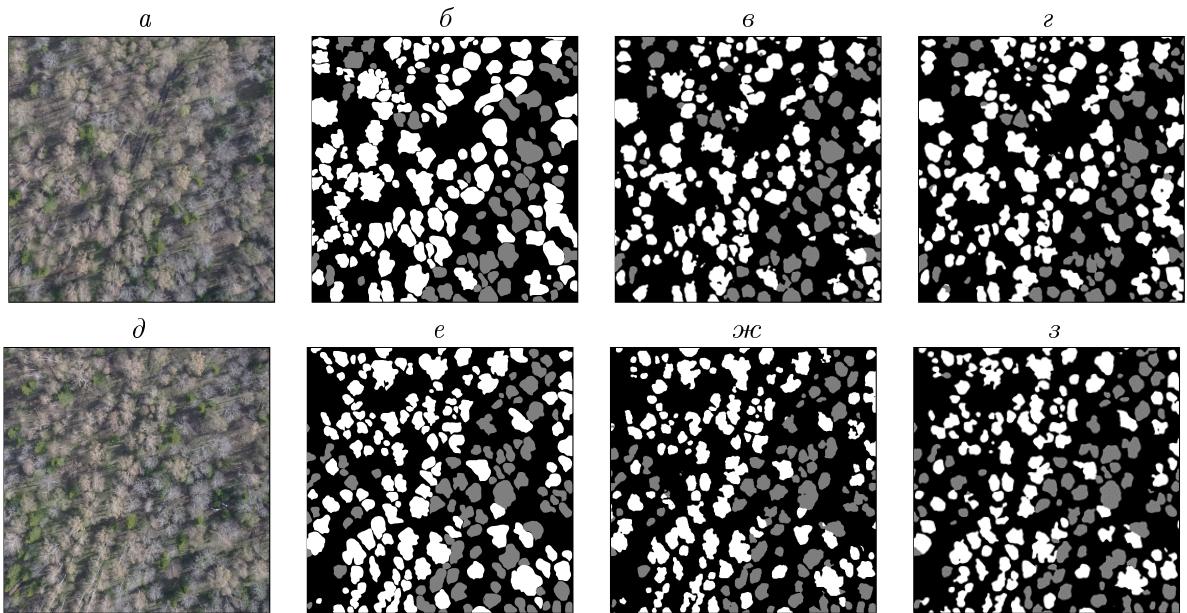


Рис. 7. Результаты семантической сегментации исходных изображений (a, д) с помощью архитектур U-Net-M (в, жс) и DeepLabV3+ (г, з) и эталонная карта сегментации (б, е). На рисунках б–г, е–з белым цветом представлены березы, а серым – осины

Fig. 7. Results of semantic segmentation of the original images (a, d) using U-Net-M (в, жс) and DeepLabV3+ (г, з) architectures and ground-truth segmentation mask (б, е). In pictures б–г, е–з birch trees are represented in white, aspen trees in gray

распознаются практически безошибочно. Наблюдаются небольшие различия в контурах выделяемых объектов, что может быть обусловлено нечеткостью и сложностью их границ.

Заключение

Рассмотрена проблема автоматизированного дешифрирования осиновых и березовых древостоев на RGB-изображениях сверхвысокого (около 5 см) пространственного разрешения, получаемых с помощью беспилотных летательных аппаратов. Для решения этой задачи использованы сверточные нейронные сети различных архитектур: U-Net, DeepLabV3, DeepLabV3+, FPN, LinkNet, PSPNet. Также была предложена архитектура U-Net-M, являющаяся модификацией U-Net с Inception-блоками. Проведено экспериментальное исследование перечисленных СНС. Для оценки качества результатов сегментации использовались метрики IoU и F-score. Результаты экспериментов показали, что наилучшее качество семантической сегментации осиновых и березовых древостоев обеспечивают архитектуры DeepLabV3+ и предложенная U-Net-M.

Результаты работы могут найти практическое применение, например, при определении размера арендной платы за лесные участки, используемые для лесозаготовки древесины.

В дальнейшем на основе этой работы планируется решать задачу подсчета числа отдельных деревьев на изображениях, полученных с БПЛА.

Благодарности. Исследование выполнено за счет гранта Российского научного фонда (проект № 22-17-20012, <https://rscf.ru/project/22-17-20012/>) при паритетной финансовой поддержке Правительства Республики Хакасия.

Программные расчеты выполнены с использованием оборудования ЦКП “Центр научных ИТ-сервисов ФИЦ ИВТ”.

Список литературы

- [1] Fassnacht F.E., Latifi H., Sterénczak K., Modzelewska A., Lefsky M., Waser L.T., Straub C., Ghosh A. Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*. 2016; (186):64–87.
- [2] Onishi M., Watanabe S., Nakashima T., Ise T. Practicality and robustness of tree species identification using UAV RGB image and deep learning in temperate forest in Japan. *Remote Sensing*. 2022; (14):1710. DOI:10.3390/rs14071710.
- [3] Weinstein B.G., Marconi S., Bohlman S., Zare A., White E. Individual tree-crown detection in RGB imagery using semi-supervised deep learning neural networks. *Remote Sensing*. 2019; (11):1309. DOI:10.3390/rs1111309.
- [4] Safanova A., Tabik S., Alcaraz-Segura D., Rubtsov A., Maglinets Y., Herrera F. Detection of fir trees (*abies sibirica*) damaged by the bark beetle in unmanned aerial vehicle images with deep learning. *Remote Sensing*. 2019; 11(6):643. DOI:10.3390/rs11060643.
- [5] Марков Н.Г., Маслов К.А., Керчев И.А., Токарева О.С. Модели U-Net для семантической сегментации поврежденных деревьев сосны сибирской кедровой на снимках с БПЛА. Современные проблемы дистанционного зондирования Земли из космоса. 2022; 19(1):65–75.
- [6] Dalponte M., Frizzera L., Gianelle D. Individual tree crown delineation and tree species classification with hyperspectral and LiDAR data. *Peer Journal*. 2019; (6):e6227.

- [7] **Maschler J., Atzberger C., Immitzer M.** Individual tree crown segmentation and classification of 13 tree species using Airborne hyperspectral data. *Remote Sensing*. 2018; (10):1218. DOI:10.3390/rs10081218.
- [8] **Дмитриев Е.В., Козуб В.А., Мельник П.Г., Соколов А.А., Сафонова А.Н.** Классификация и оценка состояния смешанных древостоев по аэроизображениям сверхвысокого пространственного разрешения. *Лесной журнал*. 2019; (5):9–24. DOI:10.17238/issn0536-1036.2019.5.9.
- [9] **Kislov D.E., Korznikov K.A., Altman J., Vozmischcheva A.S., Krestov P.V.** Extending deep learning approaches for forest disturbance segmentation on very high-resolution satellite images. *Remote Sensing in Ecology and Conservation*. 2021. DOI:10.1002/rse2.194. Available at: <https://zslpublications.onlinelibrary.wiley.com/doi/10.1002/rse2.194>.
- [10] **Harmon I., Marconi S., Weinstein B., Bai Y., Zhe Wang D., White E., Bohlman S.** Improving rare tree species classification using domain knowledge. *IEEE Geoscience and Remote Sensing Letters*. 2023; (20):8500305. DOI:10.1109/LGRS.2023.3278170.
- [11] **Dalponte M., Bruzzone L., Gianelle D.** Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data. *Remote Sensing of Environment*. 2012; (123):258–270.
- [12] **Shen X., Cao L.** Tree-species classification in subtropical forests using airborne hyperspectral and LiDAR data. *Remote Sensing*. 2017; (9):1180.
- [13] **Tuominen S., Näsi R., Honkavaara E., Honkavaara E., Balazs A., Hakala T., Viljanen N., Pölönen I., Saari H., Ojanen H.** Assessment of classifiers and remote sensing features of hyperspectral imagery and stereo-photogrammetric point clouds for recognition of tree species in a forest area of high species diversity. *Remote Sensing*. 2018; (10):714. DOI:10.3390/rs10050714.
- [14] **Sothe C., Dalponte M., de Almeida C.M., Schimalski M.B., Lima C.L., Liesenberg V., Miyoshi G.T., Tommaselli A.M.G.** Tree species classification in a highly diverse subtropical forest integrating UAV-based photogrammetric point cloud and hyperspectral data. *Remote Sensing*. 2019; (11):1338. DOI:10.3390/rs11111338.
- [15] **Nevalainen O., Honkavaara E., Tuominen S., Viljanen N., Hakala T., Yu X., Hyppä J., Saari H., Pölönen I., Imai N.N., Tommaselli A.M.G.** Individual tree detection and classification with UAV-based photogrammetric point clouds and hyperspectral imaging. *Remote Sensing*. 2017; (9):185. DOI:10.3390/rs9030185.
- [16] **Aubry-Kientz M., Laybros A., Weinstein B., Ball J.G.C., Jackson T., Coomes D., Vincent G.** Multisensor data fusion for improved segmentation of individual tree crowns in dense tropical forests. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*. 2021; (14):3927.
- [17] **Schiefer F., Kattenborn T., Frick A., Frey J., Schall P., Koch B., Schmidlein S.** Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks. *SPRS Journal of Photogrammetry and Remote Sensing*. 2020; (170):205–215.
- [18] **Dietenberger S., Mueller M.M., Bachmann F., Adam M., Metz F., Nestler M., Hese S., Thiel C.** Digital forest inventory based on UAV imagery. *Copernicus Meetings*. EGU23-1418. DOI:10.5194/egusphere-egu23-1418.
- [19] **Pajares G.** Overview and current status of remote sensing applications based on unmanned aerial vehicles. *Photogrammetric Engineering & Remote Sensing*. 2015; 81(4):281–329.
- [20] **Зотов С.А., Дмитриев Е.В., Мельник П.Г., Кондранин Т.В.** Повышение информативности мультиспектральных спутниковых изображений с использованием

- ем данных текстурного анализа. Известия вузов. Лесной журнал. 2022; (2):84–104. DOI:10.37482/0536-1036-2022-2-84-104.
- [21] **Борзов С.М., Гурьянов М.А., Потатуркин О.И.** Исследование эффективности классификации трудноразличимых типов растительности по гиперспектральным изображениям. Компьютерная оптика. 2019; 43(3):464–473. DOI:10.18287/2412-6179-2019-43-3-464-473.
- [22] **Синявский Ю.Н., Пестунов И.А., Дубровская О.А., Рылов С.А., Мельников П.В., Ермаков Н.Б., Полякова М.А.** Методы и технология сегментации мультиспектральных изображений высокого разрешения для исследования природных и антропогенных объектов. Вычислительные технологии. 2016; 21(1):127–140.
- [23] **Пестунов И.А., Рылов С.А.** Алгоритмы спектрально-текстурной сегментации спутниковых изображений высокого пространственного разрешения. Вестник Кемеровского государственного университета. 2012; (4–2(52)):104–109.
- [24] **Ma L., Liu Y., Zhang X., Ye Y., Yin G., Johnson B.A.** Deep learning in remote sensing applications: a meta-analysis and review. ISPRS Journal of Photogrammetry and Remote Sensing. 2019; (152):166–177. DOI:10.1016/j.isprsjprs.2019.04.015.
- [25] **Ronneberger O., Fischer P., Brox T.** U-Net: convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention — MICCAI 2015. Lecture Notes in Computer Science. Springer: Cham, Switzerland; 2015; (9351):234–241.
- [26] **Rahman M.A., Wang Y.** Optimizing intersection-over-union in deep neural networks for image segmentation. Advances in Visual Computing: International Symposium on Visual Computing. 2016; (10072):234–244.
- [27] Supervisely — user guide. Available at: <https://docs.supervisely.com>.

Semantic segmentation of aspen and birch tree stands on UAV RGB images using convolutional neural networks

I. A. PESTUNOV^{1,2,*}, R. A. KALASHNIKOV¹, S. A. RYLOV^{1,3}

¹Federal Research Center for Information and Computational Technologies, 630090, Novosibirsk, Russia

²Nikita Botanical Garden — National Scientific Center RAS, 298648, Yalta, Russia

³Katanov Khakass State University, 655017, Abakan, Russia

*Corresponding author: Igor A. Pestunov, e-mail: pestunov@ict.nsc.ru

Received December 14, 2023, revised February 05, 2024, accepted February 12, 2024.

Abstract

The problem of determining the species composition for forest stands using remote sensing data has been attracting significant attention for decades. The relevance of this problem is due to the constant improvement and development of remote sensing tools and technologies including the intensive spread of unmanned imagery technologies in recent years.

Remote sensing based on small unmanned aerial vehicles (UAVs) is a rapidly developing technology. Compared to manned aircraft, UAVs are an easy-to-use and low-cost tool for remote sensing of forests. Survey cameras mounted on UAVs allow data collection even in cloudy conditions. UAVs can produce flexible temporal resolution and extremely high (up to several centimeters) spatial resolution

of images, where tree features can be seen at the level of branches and even leaves. Therefore, not only spectral features but also spatial (textural and geometric) features play a significant role in tree species recognition in ultra-high spatial resolution images.

Currently, the most effective approach to analyse high spatial resolution aerospace images is to apply deep learning methods based on convolutional neural networks (CNNs). This is due to the fact that CNNs are specifically designed to analyse spatial patterns and they don't require "manual" extraction of spatial features. Unlike traditional image segmentation and recognition algorithms, CNNs provide the ability to analyse spectral and spatial features of objects in an image jointly.

The paper investigates the possibility of automatic identification and classification of aspen and birch stands in RGB images of ultra-high spatial resolution obtained from unmanned aerial vehicles. To solve the problem of semantic segmentation we compared the performance of convolutional neural networks based on different architectures: U-Net, FPN, PSPNet, Linknet, DeepLabV3, DeepLabV3+. The results showed that DeepLabV3+ architecture and modifications of U-Net with Inception-blocks allow to achieve the best results of semantic segmentation, reaching the highest values of IoU (~ 0.83) and F-score (~ 0.91) quality metrics.

Keywords: image processing, drone photography, semantic segmentation, classification, UAV, identification of tree stands, convolutional neural network.

Citation: Pestunov I.A., Kalashnikov R.A., Rylov S.A. Semantic segmentation of aspen and birch tree stands on UAV RGB images using convolutional neural networks. 2024; 29(5):100–112. DOI:10.25743/ICT.2024.29.5.008. (In Russ.)

Acknowledgements. The research was funded by the Russian Science Foundation (project No. 22-17-20012, <https://rscf.ru/project/22-17-20012/>) with parity financial support from the Government of the Republic of Khakassia.

Calculations were carried out using the equipment of the Center for the collective use of scientific equipment "Scientific IT-services center of FRC ICT".

References

1. Fassnacht F.E., Latifi H., Sterénczak K., Modzelewska A., Lefsky M., Waser L.T., Straub C., Ghosh A. Review of studies on tree species classification from remotely sensed data. *Remote Sensing of Environment*. 2016; (186):64–87.
2. Onishi M., Watanabe S., Nakashima T., Ise T. Practicality and robustness of tree species identification using UAV RGB image and deep learning in temperate forest in Japan. *Remote Sensing*. 2022; (14):1710. DOI:10.3390/rs14071710.
3. Weinstein B.G., Marconi S., Bohlman S., Zare A., White E. Individual tree-crown detection in RGB imagery using semi-supervised deep learning neural networks. *Remote Sensing*. 2019; (11):1309. DOI:10.3390/rs11111309.
4. Safonova A., Tabik S., Alcaraz-Segura D., Rubtsov A., Maglinets Y., Herrera F. Detection of fir trees (*abies sibirica*) damaged by the bark beetle in unmanned aerial vehicle images with deep learning. *Remote Sensing*. 2019; 11(6):643. DOI:10.3390/rs11060643.
5. Markov N.G., Maslov K.A., Kerchev I.A., Tokareva O.S. U-Net models for semantic segmentation of damaged *Pinus Sibirica* trees in UAV imagery. *Sovremennye Problemy Distantionnogo Zondirovaniya Zemli iz Kosmosa*. 2022; 19(1):65–75. (In Russ.)
6. Dalponte M., Frizzera L., Gianelle D. Individual tree crown delineation and tree species classification with hyperspectral and LiDAR data. *Peer Journal*. 2019; (6):e6227.
7. Maschler J., Atzberger C., Immitzter M. Individual tree crown segmentation and classification of 13 tree species using Airborne hyperspectral data. *Remote Sensing*. 2018; (10):1218. DOI:10.3390/rs10081218.
8. Dmitriev E.V., Kozub V.A., Melnik P.G., Sokolov A.A., Safonova A.N. Classification and assessment of the state of mixed forests from very high spatial resolution airborne images. *Lesnoy Zhurnal*. 2019; (5):9–24. DOI:10.17238/issn0536-1036.2019.5.9. (In Russ.)
9. Kislov D.E., Korznikov K.A., Altman J., Vozmischcheva A.S., Krestov P.V. Extending deep learning approaches for forest disturbance segmentation on very high-resolution satellite images.

- Remote Sensing in Ecology and Conservation. 2021. DOI:10.1002/rse2.194. Available at: <https://zslpublications.onlinelibrary.wiley.com/doi/10.1002/rse2.194>.
10. **Harmon I., Marconi S., Weinstein B., Bai Y., Zhe Wang D., White E., Bohlman S.** Improving rare tree species classification using domain knowledge. IEEE Geoscience and Remote Sensing Letters. 2023; (20):8500305. DOI:10.1109/LGRS.2023.3278170.
 11. **Dalponte M., Bruzzone L., Gianelle D.** Tree species classification in the Southern Alps based on the fusion of very high geometrical resolution multispectral/hyperspectral images and LiDAR data. Remote Sensing of Environment. 2012; (123):258–270.
 12. **Shen X., Cao L.** Tree-species classification in subtropical forests using airborne hyperspectral and LiDAR data. Remote Sensing. 2017; (9):1180.
 13. **Tuominen S., Näsi R., Honkavaara E., Honkavaara E., Balazs A., Hakala T., Viljanen N., Pölönen I., Saari H., Ojanen H.** Assessment of classifiers and remote sensing features of hyperspectral imagery and stereo-photogrammetric point clouds for recognition of tree species in a forest area of high species diversity. Remote Sensing. 2018; (10):714. DOI:10.3390/rs10050714.
 14. **Sothe C., Dalponte M., de Almeida C.M., Schimalski M.B., Lima C.L., Liesenberg V., Miyoshi G.T., Tommaselli A.M.G.** Tree species classification in a highly diverse subtropical forest integrating UAV-based photogrammetric point cloud and hyperspectral data. Remote Sensing. 2019; (11):1338. DOI:10.3390/rs11111338.
 15. **Nevalainen O., Honkavaara E., Tuominen S., Viljanen N., Hakala T., Yu X., Hyppä J., Saari H., Pölönen I., Imai N.N., Tommaselli A.M.G.** Individual tree detection and classification with UAV-based photogrammetric point clouds and hyperspectral imaging. Remote Sensing. 2017; (9):185. DOI:10.3390/rs9030185.
 16. **Aubry-Kientz M., Laybros A., Weinstein B., Ball J.G.C., Jackson T., Coomes D., Vincent G.** Multisensor data fusion for improved segmentation of individual tree crowns in dense tropical forests. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. 2021; (14):3927.
 17. **Schiefer F., Kattenborn T., Frick A., Frey J., Schall P., Koch B., Schmidlein S.** Mapping forest tree species in high resolution UAV-based RGB-imagery by means of convolutional neural networks. SPRS Journal of Photogrammetry and Remote Sensing. 2020; (170):205–215.
 18. **Dietenberger S., Mueller M.M., Bachmann F., Adam M., Metz F., Nestler M., Hese S., Thiel C.** Digital forest inventory based on UAV imagery. Copernicus Meetings. EGU23-1418. DOI:10.5194/egusphere-egu23-1418.
 19. **Pajares G.** Overview and current status of remote sensing applications based on unmanned aerial vehicles. Photogrammetric Engineering & Remote Sensing. 2015; 81(4):281–329.
 20. **Zotov S.A., Dmitriev E.V., Melnik P.G., Kondranin T.V.** Increasing the informativity of multispectral satellite images using texture analysis data. Lesnoy Zhurnal. 2022; (2):84–0104. DOI:10.37482/0536-1036-2022-2-84-104. (In Russ.)
 21. **Borзов S.M., Гурянов М.А., Потатурик О.И.** Study of the classification efficiency for different-to-distinguish vegetation types using hyperspectral data. Computer Optics. 2019; 43(3):464–473. DOI:10.18287/2412-6179-2019-43-3-464-473. (In Russ.)
 22. **Sinyavskiy Yu.N., Pestunov I.A., Dubrovskaya O.A., Rylov S.A., Melnikov P.V., Ermakov N.B., Polyakova M.A.** Methods and technology for segmentation of images with high spatial resolution for studies of nature and man-made objects. Computational Technologies. 2016; 21(1):127–140. (In Russ.)
 23. **Pestunov I.A., Rylov S.A.** Spectral-texture segmentation algorithms for high spatial resolution satellite images. Bulletin of Kemerovo State University. 2012; (4-2(52)):104–109. (In Russ.)
 24. **Ma L., Liu Y., Zhang X., Ye Y., Yin G., Johnson B.A.** Deep learning in remote sensing applications: a meta-analysis and review. ISPRS Journal of Photogrammetry and Remote Sensing. 2019; (152):166–177. DOI:10.1016/j.isprsjprs.2019.04.015.
 25. **Ronneberger O., Fischer P., Brox T.** U-Net: convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention — MICCAI 2015. Lecture Notes in Computer Science. Springer: Cham, Switzerland; 2015; (9351):234–241.
 26. **Rahman M.A., Wang Y.** Optimizing intersection-over-union in deep neural networks for image segmentation. Advances in Visual Computing: International Symposium on Visual Computing. 2016; (10072):234–244.
 27. Supervisely — user guide. Available at: <https://docs.supervisely.com>.