

*Д. А. Шатилов<sup>1</sup>, С. А. Рылов<sup>2\*</sup>*

## **Определение видового состава лесных массивов на спутниковых снимках высокого разрешения с помощью сверточных нейронных сетей**

<sup>1</sup> Новосибирский государственный университет, г. Новосибирск, Российская Федерация

<sup>2</sup> Федеральный исследовательский центр информационных и вычислительных технологий, г. Новосибирск, Российская Федерация

\*e-mail: RylovS@mail.ru

**Аннотация.** Получаемые данные со средств дистанционного зондирования Земли являются очень полезными для решения задач, связанных с мониторингом лесной растительности и лесотаксацией. Основным этапом тематической обработки таких данных является сегментация, проводимая зачастую вручную на основе экспертной оценки. Методы автоматической сегментации на основе глубокого обучения, учитывающие пространственные и спектральные характеристики изображения, позволяют избежать данного трудоемкого процесса и обеспечивают высокую точность построения карт сегментации. В настоящей работе исследуется возможность применения СНС-моделей (СНС – сверточная нейронная сеть) на мультиспектральных спутниковых снимках высокого пространственного разрешения с целью определения видового состава древостоев исследуемой местности. Для составления площадных карт сегментации использовались модели архитектуры U-Net, которые обучены на датасете, состоящем из снимков с российского спутника «Ресурс-П» и экспертной разметки. С использованием кросс-валидации и механизма аугментации произведена семантическая сегментация тестируемого участка и оценена точность сегментации с помощью IoU метрики. Показано, что оптимальной моделью для сформированного датасета является архитектура Attention U-Net. Используя трансферное обучение на основе снимков со спутника «WorldView-2» и метод коррекции границ, было дополнительно улучшено качество результатов сегментации.

**Ключевые слова:** семантическая сегментация, мониторинг леса, аэрокосмические изображения, глубокое обучение, U-Net

*D. A. Shatilov<sup>1</sup>, S. A. Rylov<sup>2\*</sup>*

## **Tree Species Forest Mapping in High-Resolution Satellite Images Using Convolutional Neural Networks**

<sup>1</sup> Novosibirsk State University, Novosibirsk, Russian Federation

<sup>2</sup> Federal Research Center for Information and Computational Technologies, Novosibirsk, Russian Federation

\* e-mail: RylovS@mail.ru

**Abstract.** The data obtained from Earth remote sensing tools is very useful in solving forest vegetation monitoring and forest taxation problems. The main stage of thematic processing of such data is segmentation, which is often done manually based on an expert assessment. Automatic segmentation methods based on deep learning allow avoiding this labor-intensive process and providing highly accurate segmentation maps, taking into account spatial and spectral image characteristics. This paper examines CNN (convolutional neural network) models application on

multispectral satellite images of high spatial resolution in order to determine the tree species composition in the study area. In order to build segmentation maps, U-Net architecture models were trained on a compiled dataset consisting of images from the Russian «Resurs-P» satellite and expert mark-up. Using cross-validation and augmentation mechanism, semantic segmentation of the tested area was carried out. The segmentation quality was assessed using the IoU metric. It is shown that the optimal model for the compiled dataset is "Attention U-Net". Using transfer learning based on «WorldView-2» satellite images and the boundary correction method, the segmentation quality was additionally improved.

**Keywords:** semantic segmentation, forest monitoring, aerospace imagery, deep learning, U-Net

### *Введение*

В настоящее время оптические спутниковые данные высокого пространственного разрешения незаменимы для мониторинга обширных и труднодоступных лесных территорий. Для проведения анализа получаемых аэрокосмических снимков и выделения лесообразующих пород в задаче лесотаксации зачастую используют экспертную оценку, но ее ручное составление затруднено ввиду мультиспектральности и крупномасштабности изображений. В связи с чем для их обработки уместно применять технологии компьютерного зрения, проводящие распознавание и классификацию в полуавтоматическом или автоматическом режиме. Методы сегментации позволяют построить для исследуемой территории площадную маску с выделенными кластерами, соответствующими различным классам объектов. При этом процесс распознавания происходит на основе сходства спектральных и пространственных признаков [1] различных областей.

Для задачи семантической сегментации в данный момент активно [2] применяются методы на основе глубокого обучения, позволяющие получить высокое качество сегментации изображений. Оценку точности построения карт сегментации зачастую проводят путем сравнения с масками, составленными по экспертной оценке. Наибольшее сходство с исходной разметкой достигается применением методов обучения с учителем, когда часть экспертной разметки интерпретируется как обучающая выборка. Наиболее эффективные среди моделей данного типа [2] с точки зрения реализации и вычислительных затрат являются сверточные нейронные сети, среди которых выделяется симметричная пирамидальная U-Net модель [3]. Авторы статьи [3], применяя модель U-Net в задачах сегментации биомедицинских изображений, получили значения IoU-метрики от 77.6% до 92% на разных датасетах, более чем на 10% опередив другие сверточные нейронные сети. При проведении семантической сегментации аэрокосмических снимков U-Net также демонстрирует высокие результаты [4-7]. Активно развиваются и различные модификации U-Net, наиболее популярные из которых это U-Net++ [8] и Attention U-Net [9]. Применение данных моделей на ряде датасетов позволяет улучшить качество предсказания оригинальной нейросети [10, 11]. Таким образом, в литературе делается акцент на высокой способности U-Net подобных архитектур к семантической сегментации различных изображе-

ний благодаря способности к выделению разномасштабных текстурных признаков. И, как показано в [2], выбор оптимальной архитектуры U-Net сильно зависит от тематической области и тестируемого датасета.

Цель данного исследования заключается в разработке метода автоматической сегментации и построении оптимальной модели на основе U-Net для распознавания видового состава лесов с использованием космических снимков высокого разрешения на примере российского спутника «Ресурс-П». В связи с чем были поставлены задачи составления датасета на основе специфики выбранных моделей, определения гиперпараметров для максимизации точности, оценки вклада источников ошибок и их минимизации посредством корректировки построенных площадных карт древостоев.

### *Методы и материалы*

Рассматриваемый тестовый участок является частью Бронницкого лесничества в Московской области, значительную площадь которого занимают искусственные насаждения различных лесных пород. Такие насаждения характеризуются чистым видовым составом и имеют четкие границы произрастания, отделяющие их от основных лесных массивов. При создании датасета обследуемой территории использовался 4-канальный крупномасштабный снимок высокого пространственного разрешения (рис. 1,а), сделанный спутником «Ресурс-П». В качестве исходной карты сегментации используются данные, полученные в работе [12], всего выделено 14 следующих классов (рис. 1,б): «не лес» – особый класс для учета неразмеченных областей; «нет информации» – специально введенная метка для прочих видов (согласно оценке [12] большая часть таких пикселей «соответствует границам древостоев и просекам»); «сосна обыкновенная»; «ель европейская»; «пихта сибирская»; «лиственница европейская»; «лиственница сибирская»; «лиственница Сукачева»; «дуб черешчатый»; «береза бородавчатая»; «осина»; «ольха черная»; «липа мелколистная» и «ива белая».

При подготовке датасета метка «ива белая» была заменена на класс «нет информации», так как ей соответствует малый по размерам лесной массив, изменения остальных классов не производились. Важно отметить, что учет метки «нет информации» при дальнейшей сегментации опущен согласно приведенной выше оценке и дополнением исходного класса граничными пикселями лесных насаждений. Для составления обучающих фрагментов (тайлов) из полученной мультиклассовой разметки и спутниковых данных, разметка была поделена на области размером  $256 \times 256$  по пиксельной сетке, которым поставлен в соответствие участок снимка «Ресурс-П» с учетом географического преобразования. Таким образом, был подготовлен датасет, состоящий из 81 пары 4-канальных тайлов спутникового изображения и разметки, переведенной в "one-hot" формат.

Для построения площадных масок видового состава рассматриваемого датасета были протестированы две модели: U-Net++ и Attention U-Net с энкодером на базе ResNet-50 [13].

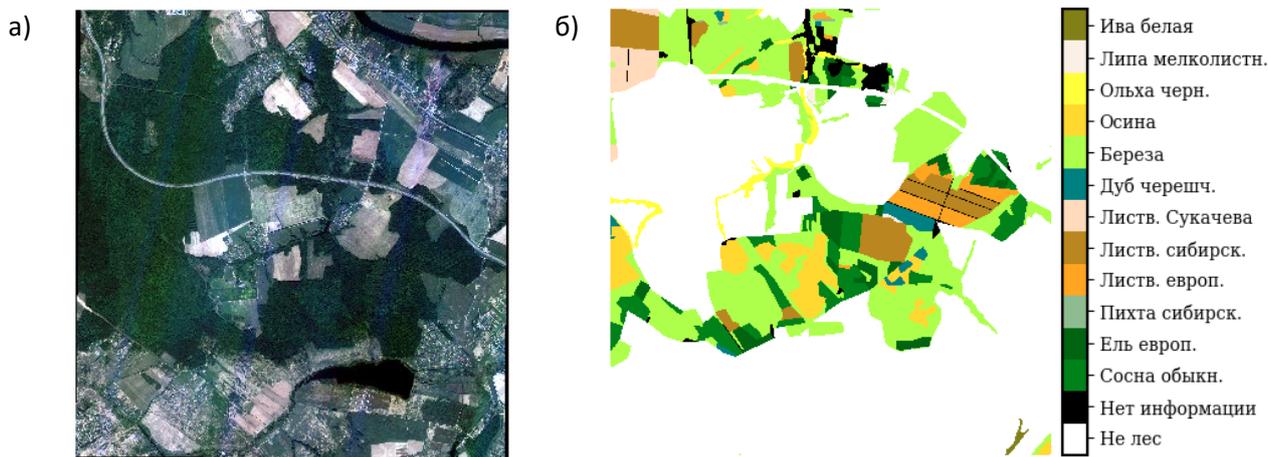


Рис. 1. Исходные данные: а) – RGB-композит снимка спутника «Ресурс-П»; б) – экспертная разметка территории

В качестве метрики нейросетей был выбран коэффициент Жаккара или IoU, усредняемый по  $N$  классам, а функцией потерь служила линейная комбинация функции потерь Дайса ( $DL$ ) и модифицированной перекрёстной энтропии (CSE) [14]:

$$CL = (1 - \alpha) DL + \alpha \sum_{i=0}^{N-1} w_i \sum_{j,k} (1 - pred_i[j, k])^\gamma \cdot CSE_i[j, k], \quad (1)$$

где  $\alpha$  – коэффициент пропорциональности,  $\gamma$  – параметр фокусировки,  $pred_i$  –  $i$ -ый слой маски предсказания. Представленная модификация кросс-энтропии также учитывает веса классов  $w_i$ , определяемых обратно пропорционально количеству пикселей  $i$ -го класса.

Ввиду присутствия в датасете малого количества изображений, содержащих редко встречаемые классы, для оценки точности модели использовалась процедура кросс-валидации, где количество разбиений было установлено равным 4. На каждом шаге перекрестной проверки размер тренировочного датасета увеличивался с помощью аугментаций, включавших случайное смещение центра, поворот и отражение по двум направлениям. Части изображения или маски, обрезанные в результате аугментации, заполнялись с помощью зеркального отражения.

### Результаты

С целью оптимального обновления весов и предотвращения переобучения для каждой нейросети составлялись индивидуальные планировщики скорости обучения. После обучения моделей лучшие результаты распознавания видового состава по кросс-валидации для подготовленного датасета составили: для U-Net++ сети – 0.252, для Attention U-Net – 0.292 (рис. 2).

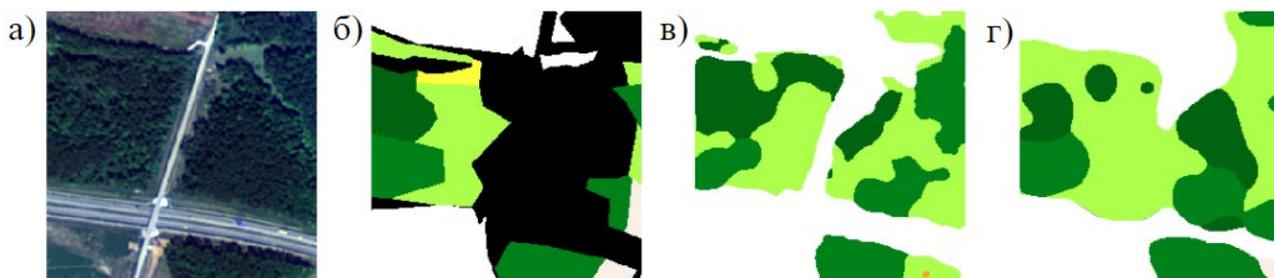


Рис. 2. Пример исходного изображения (а), разметки (б) и построенных карт сегментации для Attention U-Net (в) и U-Net++ (г) моделей

Для повышения качества предсказания также применялась процедура трансферного обучения: модели заранее инициализировали весами, полученными при проведении сегментации на другом датасете, после чего сети дополнительно дообучали. В качестве данных для трансферного обучения использовался датасет, составленный по той же экспертной разметке и снимку со спутника «WorldView-2», для которого были взяты каналы [Blue (2), Green (3), Red (5), NIR\_1 (7)], как наиболее соответствующие по спектральному диапазону каналам спутника «Ресурс-П». Использование предложенного метода позволило увеличить IoU метрику для U-Net++ и Attention U-Net до 0.267 и 0.302 соответственно (рис. 3). Достигнутая точность на тренировочном датасете для сетей составила 0.512 и 0.498.

Также было проведено сведение предсказания модели к доминантному виду на выделенных лесотаксационных наделах произведено путем попиксельной коррекции границ отдельных кластеров карты сегментации с помощью нахождения наиболее распространенной метки предсказания для заданного выдела, пример преобразования продемонстрирован на рис. 4.

Применяя описанный алгоритм, удалось повысить точность сегментации до значений 0.271 и 0.317, а на тренировочном наборе – до 0.524 и 0.543 для U-Net++ и Attention U-Net соответственно.

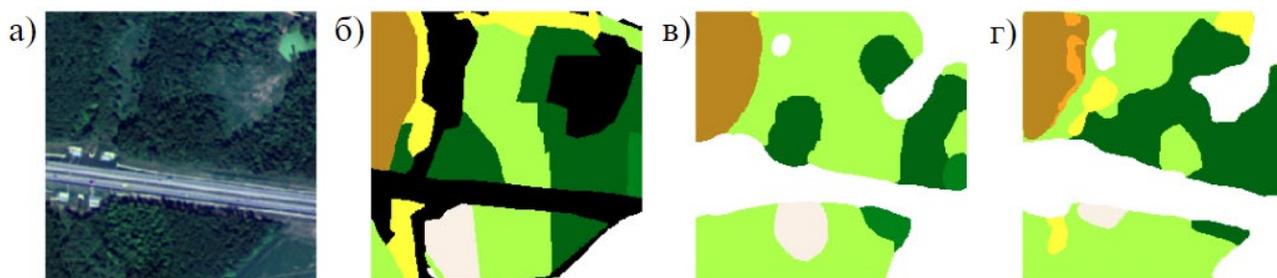


Рис. 3. Пример исходного изображения (а), разметки (б) и построенных карт сегментации для Attention U-Net (в) и U-Net++ (г) моделей с использованием трансферного обучения

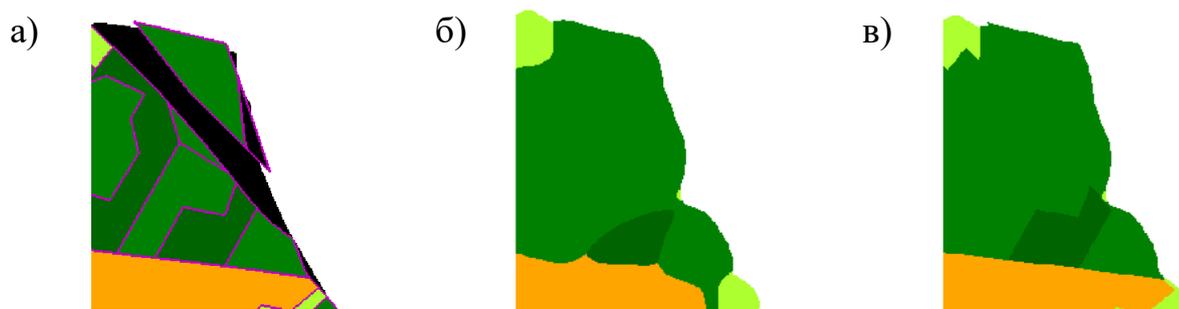


Рис. 4. Преобразование построенной маски сегментации: а) – экспертная разметка с контурами лесотаксационных выделов (фиолетовые линии); б) – карта предсказания модели; в) – преобразованная карта

### *Обсуждение*

Анализ полученных результатов сегментации показал, что низкое качество сегментации снимка «Ресурс-П» связано с влиянием характерных шумов при съемке синего канала (рис. 1, а), а также из-за ошибочного выделения лесных пород, ввиду появления антропогенных объектов и проведения вырубок на исследуемой территории (рис. 3). Последнее связано с различием в датах съемки местности и составления разметки, в частности, ошибки при разметке граничных пикселей около антропогенных объектов особенно сильно влияют на качество предсказания U-Net++ (рис. 2), что приводит к ошибочному выделению лесных пород на очищенных участках.

Более высокие результаты модели Attention U-Net в сравнении с U-Net++, обусловлены способностью сети подавлять нерелевантные пространственные области и отдельные каналы во время обучения [9]. Для указанной модели основные ошибки при составлении площадных карт разметки связаны с разметкой, составленной на уровне лесотаксационных выделов с указанием доминантной породы, в то время как использование СНС-моделей предполагает проведение попиксельной сегментации. Применение трансферного обучения и метода коррекции границ позволяет увеличить качество семантической сегментации, которое происходит за счет подавления наименее распространенных классов, приводя к тому, что метка, к примеру, «липы мелколистной» полностью заменяется другими классами при построении карты видового состава. Приведенная нейросеть особенно явно различает классы «не лес» (IoU: 0.98), «сосну обыкновенную» (0.83), «березу» (0.52) и «ель европейскую» (0.46), но при выделении различных фенотипов лиственницы происходит смешение видового состава. Для более точного выделения данных классов, по всей видимости, необходимо проводить сегментацию с учетом разновременных снимков рассматриваемого участка, когда породы деревьев различны – в период вегетации и перед опадом листвы [12, 15].

### *Заключение*

В данной работе были апробированы СНС-модели на основе архитектуры U-Net для проведения семантической сегментации 4-канального космического

снимка со спутника «Ресурс-П». Для экспериментальных исследований был подготовлен малоразмерный датасет, состоящий из пар изображение-маска, на которых выделено 12 сильно несбалансированных классов древостоев. Снижение влияния несбалансированности разметки достигалось с помощью взвешенной функции потерь (1) и метода кросс-валидации, а также расширением тренировочного датасета при помощи аугментации на каждом шаге. На подготовленных данных наиболее высокую эффективность показала модель Attention U-Net, достигая значения по кросс-валидации в 0.292 по средней IoU метрике. Результат распознавания видового состава данной модели был дополнительно улучшен с применением трансферного обучения и метода выделения доминантной породы на лесотаксационных выделах. Полученная точность составила 0.317 и 0.543 для валидационного и тренировочного наборов.

В дальнейших исследованиях планируется разработка архитектуры, совмещающей подавление нерелевантных областей со способностью анализа разновременных изображений исследуемой территории, что должно существенно увеличить качество семантической сегментации древостоев.

### *Благодарности*

Авторы выражают признательность к.ф.-м.н. Егору Владимировичу Дмитриеву за предоставленные данные разметки.

### БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Пестунов И.А., Рылов С.А. Сегментация спутниковых изображений высокого разрешения по спектральным и текстурным признакам // Интерэкспо Гео-Сибирь. – 2012. – Т. 1. – № 4. – С. 86-91.
2. Huang L., Jiang B., Lv S., Li Y. Deep Learning-based Semantic Segmentation of Remote Sensing Images: A Survey // IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing. – 2023. – Vol. 10. – P. 1-28.
3. Ronneberger, O., Fischer, P., Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation // Lecture Notes in Computer Science. – Springer, 2015. – Vol. 9351. – P. 234-241.
4. Långkvist M, Kiselev A, Alirezaie M, Loutfi A. Classification and Segmentation of Satellite Orthoimagery Using Convolutional Neural Networks // Remote Sensing. – 2016. – Vol. 8(4). – P. 329.
5. Iglovikov V., Shvets A. TerausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation [Электронный ресурс]. – 2018. – Режим доступа: <https://arxiv.org/abs/1801.05746> (дата обращения: 01.05.2024).
6. Iglovikov V., Mushinskiy S., Osin V. Satellite Imagery Feature Detection using Deep Convolutional Neural Network: A Kaggle Competition [Электронный ресурс]. – 2017. – Режим доступа: <https://arxiv.org/abs/1706.06169> (дата обращения: 01.05.2024).
7. Safarov F., Temurbek K., Djumanov J. et al. Improved Agricultural Field Segmentation in Satellite Imagery Using TL-ResUNet Architecture. // Sensors. – 2022. – Vol. 22(24). – P. 9784.
8. Zhou Z., Siddiquee R., Tajbakhsh N., Liang J. UNet++: A Nested U-Net Architecture for Medical Image Segmentation [Электронный ресурс]. – 2018. – Режим доступа: <https://arxiv.org/abs/1807.10165> (дата обращения: 01.05.2024).
9. Oktay O., Schlemper J., Folgoc L. et al. Attention U-Net: Learning Where to Look for the Pancreas [Электронный ресурс]. – 2019. – Режим доступа: <https://arxiv.org/abs/1804.03999> (дата обращения: 01.05.2024).

10. Lokula, B., Tirumuri R. Satellite image segmentation using Unet++ and MobileNetV2 deep learning model [Электронный ресурс]. – 2024. – Режим доступа: <https://www.researchsquare.com/article/rs-4144393/v1> (дата обращения: 01.05.2024).
11. David J., Ce Z. An attention-based U-Net for detecting deforestation within satellite sensor imagery // International Journal of Applied Earth Observation and Geoinformation. – 2022. – Vol. 107. – P. 102685.
12. Дмитриев Е. В., Кондранин Т. В., Мельник П. Г., Донской С. А. Определение видового состава смешанного леса на основе совместной обработки публичных спутниковых карт и многовременных изображений Sentinel-2 // Современные проблемы дистанционного зондирования Земли из космоса. – 2023. – Т. 21. – № 1. – С. 31-50.
13. Agrawal S., Rewaskar V., Agrawal R. et al. Advancements in NSFW Content Detection: A Comprehensive Review of ResNet-50 Based Approaches // Journal of intelligent systems and applications in engineering. – 2023. – Vol. 11 (4). – P. 41-45.
14. Azad R., Heidary M., Yilmaz K. et al. Loss Functions in the Era of Semantic Segmentation: A Survey and Outlook [Электронный ресурс]. – 2021. – Режим доступа: <https://arxiv.org/abs/2312.05391> (дата обращения: 01.05.2024).
15. Калашников Р.А., Пестунов И.А., Ермаков Н.Б., Радчиков А.Н. Распознавание осиновых и березовых насаждений на RGB-изображениях с БПЛА с использованием глубокого обучения // Сборник трудов XVIII Российской конференции с международным участием «Распределенные информационно-вычислительные ресурсы (DICR-2022)». – Новосибирск, 2022. – С. 104-108.

© Д. А. Шатилов, С. А. Рылов, 2024