

Разработка алгоритма определения изменений лесного покрова на изображениях со спутника Sentinel-2

Р.В. Ларионов¹, А.В. Сенников¹, В.В. Хрящев¹

¹ Ярославский государственный университет им. П.Г. Демидова, Советская улица, 14, г. Ярославль 150003, Россия

Аннотация

В работе представлен алгоритм детектирования произошедших за определенный период времени изменений лесного покрова с использованием спутниковых изображений. В качестве изменений рассматриваются лесные вырубки, как наиболее часто встречающееся изменение лесного покрова в средней полосе России. Предлагаемое решение основано на использовании парных изображений, состоящих из двух спутниковых снимков одной и той же территории в разные моменты времени, которые обрабатываются нейросетевыми моделями с целью получения бинарной маски сегментации, соответствующей произошедшим вырубкам. Для проведения эксперимента был собран набор из 109 парных изображений. В качестве нейросетевых моделей в работе рассматриваются три модели с архитектурами ResNet-34+U-Net, SegFormer_b5 и SegNeXt_1. Для оценки работы моделей использовались такие метрики, как коэффициент Серенсена, F-мера, точность и полнота. Сеть SegNeXt_1 показала лучшие результаты по коэффициенту Серенсена и полноте со значениями 0,84 и 0,80, а сеть ResNet-34+U-Net – лучшие результаты по точности и F-мере – 0,71 и 0,74.

Ключевые слова

Детектирование изменений, сегментация изображений, спутниковые изображения, лесные вырубки, свёрточные нейронные сети, компьютерное зрение.

Forest Change Detection Methods in Sentinel-2 Images

R.V. Larionov¹, A.V. Sennikov¹, V.V. Khryashchev¹

¹ P. G. Demidov Yaroslavl State University, Sovetskaya Str. 14, Yaroslavl, 150003, Russian Federation

Abstract

This paper presents an algorithm for the detection of deforestation using satellite imagery. The proposed solution is based on the use of paired images, consisting of two images of the same area taken at different times, which are processed by neural network models to obtain a binary segmentation mask corresponding to the deforestation that has occurred. A set of 109 paired images was collected for the experiment. Three models with the architectures ResNet-34+U-Net, SegFormer_b5 and SegNeXt_1 are considered as neural network models in this paper. Metrics such as Dice, F-score, precision and recall were used to evaluate the performance of the models. The SegNeXt_1 network performed best in Dice coefficient and recall with values of 0.84 and 0.80, while the ResNet-34+U-Net network performed best in precision and F-measure with values of 0.71 and 0.74.

Keywords

Change detection, image segmentation, satellite images, forest cuts, convolutional neural networks, computer vision.

ГрафиКон 2023: 33-я Международная конференция по компьютерной графике и машинному зрению, 19-21 сентября 2023 г., Институт проблем управления им. В.А. Трапезникова Российской академии наук, г. Москва, Россия

EMAIL: r.larionov3@uniyar.ac.ru (Р.В. Ларионов); alexeysennikov76@yandex.ru (А.В. Сенников); v.khryashchev@uniyar.ac.ru (В.В. Хрящев)

ORCID: r.larionov3@uniyar.ac.ru (Р.В. Ларионов); 0009-0001-8771-6233 (А.В. Сенников); 0000-0002-7132-135X (Р.В. Ларионов); 0000-0003-2978-2459 (В.В. Хрящев)



© 2023 Copyright for this paper by its authors.
Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

1. Введение

В настоящее время задача мониторинга лесного покрова становится все более актуальной для оперативного картографирования. Оперативное получение информации о состоянии лесных массивов крайне важно лесному хозяйству для оптимального решения проблем, связанных с различными повреждениями лесов таких как вырубки, гари, ветровалы [1].

Задачу автоматического контроля изменений лесного покрова можно решить с помощью методов глубокого обучения с использованием данных дистанционного зондирования земли (ДЗЗ) [2]. На современном этапе развития спутниковых систем данные ДЗЗ находятся в открытом доступе и используются в широком спектре задач, связанных с поиском объектов определенного класса, детектированием изменений. В частности, обученные на таких данных нейросетевые модели позволяют с высокой точностью решать задачи детектирования незаконных свалок [3], маскирования облаков [4] и сегментации сельскохозяйственных полей [5].

Задача, рассматриваемая в данной работе, относится к категории задач, называемой детектированием изменений. Данное направление является важным инструментом для долгосрочных наблюдений за Землей [6]. Ее основным отличием от других задач, связанных с распознаванием и поиском, является временной контекст. Это означает, что интересующими нас объектами являются различные изменения в изображениях одной сущности, полученных в разные моменты времени. Также данная задача подразумевает разделение произошедших изменений на классы.

Целью данной работы является разработка алгоритма детектирования изменений лесного покрова на основе анализа данных ДЗЗ. В качестве изменений в работе рассматривается лесная вырубка, как наиболее часто встречающееся нарушение лесного покрова в средней полосе России.

2. Набор изображений

Для проведения эксперимента был собран набор из парных изображений, полученных с использованием ресурса Copernicus [7], предоставляющего изображения со спутника Sentinel-2. Данные снимки не подвергались никакой предобработке, кроме орторектификации и атмосферной коррекции. Пространственный размер изображений составлял 100x100 километров, а пространственное разрешение – 10 метров на пиксель. Парное изображение состоит из двух 16-битных снимков одной территории, снятых в разные моменты времени. Каждый из двух снимков представляет из себя многоканальный растр, состоящий из 4 каналов – красного, зеленого, синего и ближнего инфракрасного. Таким образом, парное изображение является восьмиканальным. Изображения из набора были сняты в период с 2019 по 2021 год. Средний промежуток времени в снимках одной территории составлял полгода. Каждому парному изображению была экспертно проставлена маска вырубки, соответствующая изменениям между двумя снимками. Фрагменты из набора с соответствующей маской вырубок представлены на рисунке 1.

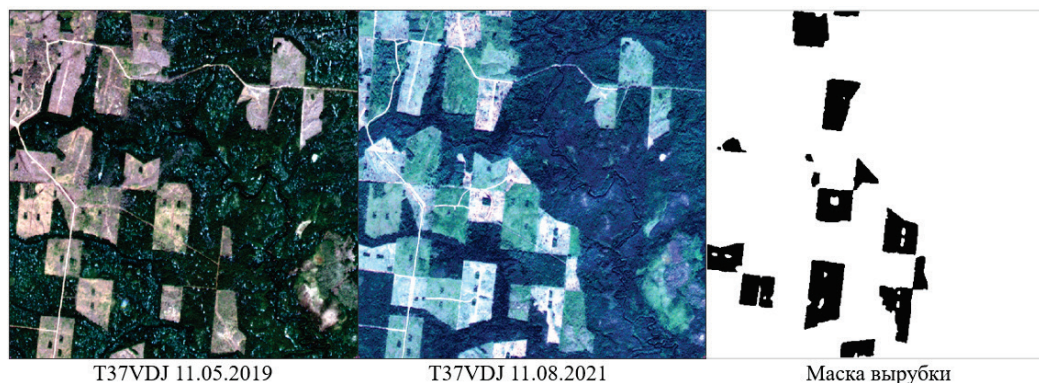


Рисунок 1 – Пример изображений выборки

Для обучения из исходных больших изображений с шагом 256 пикселей были сформированы патчи размером 512x512 пикселей. Затем все изображения были перенормированы с 16 бит на 8 в формате чисел с плавающей запятой так, что потери информации не было. Это сделано для того, чтобы на входе моделей были каналы со значениями яркостей пикселей в диапазоне от 0 до 255 для осуществления трансферного обучения. Полученный набор данных был поделен на обучающую и тестовую выборку в соотношении 101 к 8. Полная статистика по данному набору представлена в таблице 1. Из нее видно, что количество патчей с наличием вырубок заметно меньше количества патчей без вырубков. В связи с этим для эффективного обучения моделей производилась балансировка батча, использованная для того, чтобы в одном батче половина патчей имела вырубку, а другая половина нет [8].

Таблица 1 – Статистика сформированных наборов

	Обучающая выборка	Тестовая выборка
Количество парных изображений	101	8
Количество патчей	178164	14112
Количество патчей с вырубкой	48113	3236
Количество патчей без вырубки	130051	10876

3. Описание нейросетевых моделей

Для решения задачи сегментации лесных вырубок было использовано три нейросетевых моделей типа кодировщик-декодировщик. В качестве первой модели была использована широко используемая в различных задачах сегментации данных ДЗЗ полностью сверточная нейронная сеть U-Net [9], на входе которой стоит базовая сеть с архитектурой ResNet-34 [10]. Архитектура данной сети описана в предыдущей работе [1].

Двумя другими алгоритмами выбраны нейронные сети на базе трансформеров, в основе которых лежит наличие механизма внимания [11]. Принцип работы подобных алгоритмов построен на позиционном кодировании различных участков входного изображения. В рамках текущего исследования выбраны две архитектуры на основе трансформеров: SegFormer [12] и SegNeXt [13].

SegFormer относится к типу нейронных сетей, полностью работающих с использованием трансформеров. Трансформеры набирают все большую популярность в задачах машинного обучения. Их главной особенностью является наличия механизма внимания, позволяющего ускорить обучение и связать все вводимые данные между собой. Эта особенность дает возможность выполнять параллельную обработку данных и учитывать контекст. Используемая архитектура SegFormer с сетью извлечения признаков `mit_b5` представлена на рисунке 2.



Рисунок 2 – Архитектура SegFormer_b5

SegFormer состоит из кодировщика и декодировщика. Кодировщик выполняет роль извлечения грубых и точных признаков из входного изображения и состоит из блока разбиения на патчи и 4 блоков-трансформеров [12]. Блок разбиения на патчи дробит входное изображение на патчи размером 4x4 пикселей, которые поступают на вход блока-трансформера. Данный блок включает в себя повторенные N раз слои внимания и слои Mix-FFN, которые смешивают свертку размером 3x3 и сеть прямого распространения. На выходе блока-трансформера стоит блок слияния, который объединяет признаки, полученные на предыдущем этапе, и понижает

размерность, получая более высокоуровневое представление. Декодер в SegFormer объединяет многоуровневые признаки, выработанные кодировщиком, для прогнозирования маски сегментации. Он состоит только из легковесных слоев, включающих в себя многослойные перцептроны и слои повышения размерности.

В качестве третьей модели в данной работе использовалась модель с архитектурой SegNeXt [13]. Она также относится к типу трансформеров и является улучшением архитектуры SegFormer. В кодировщике SegNeXt вместо блока внимания из SegFormer используется многомасштабный модуль сверточного внимания – MSCA. Данный модуль содержит свертку 5×5 для объединения локальной информации, 3 ветки сверток по глубине для захвата контекста и свертку 1×1 для моделирования взаимосвязи между различными каналами. Совместно с 2 слоями нормализации батча и сетью прямого распространения, модуль MSCA образует блок MSCAN, являющийся основным структурным элементом кодера в SegNeXt, аналогичный трансформерному блоку в SegFormer. MSCAN также повторен 4 раза. Декодировщик в SegNeXt отличается тем, что объединяет признаки, полученные только с 3 последних этапов, а не со всех 4, как в SegFormer, что повышает производительность и отбрасывает низкоуровневую информацию с первого этапа, а также наличием операции разложения матриц извлеченных признаков. Архитектура SegNeXt_1 представлена на рисунке 3.



Рисунок 3 – Кодер и декодер архитектуры SegNeXt_1

4. Описание эксперимента

Рассмотрим результаты обучения и тестирования представленных моделей. Все представленные модели перед обучением имели веса, полученные при их обучении на наборах изображений ImageNet [8] и DeepGlobe Land Cover [9]. Обучение на наборе ImageNet дало способность решать задачу поиска низкоуровневых признаков, а на наборе DeepGlobe Land Cover – способность решать задачу сегментации на спутниковых снимках. Обучение происходило с размером батча 16. На изображения с наличием вырубок и с отсутствием вырубок приходилось по 8 патчей в батче. В качестве алгоритма оптимизации для сети ResNet-34+U-Net был использован алгоритм Adam с начальной скоростью сходимости 0,001. Для сетей-трансформеров в качестве такого алгоритма выступал AdamW с начальной скоростью сходимости в 0,000001 и линейно уменьшающейся до нуля.

Все обучаемые модели принимают на вход 8-канальный патч размером 512×512 пикселей. Первые 4 канала – красный, зеленый, синий и ближний инфракрасный относятся к изображению, снятому со спутника в более ранний момент времени, а вторые аналогичные каналы – к изображению, снятому в более поздний момент времени. На выходе нейронные сети выдают одноканальную бинарную маску с тем же размером, что и патч на входе. Результат работы алгоритма представлен на рисунке 4 на примере модели ResNet-34+U-Net. На рисунке наглядно видно сравнение эталонной и предсказанных масок, из которого можно сделать вывод о том, что модель хорошо локализует изменения, но при этом ошибается в деталях.

Для объективной оценки качества работы нейросетевых моделей использовались такие метрики, как точность (precision, P), полнота (recall, R), F-мера ($F1$), коэффициент Серенсена (Dice) и IoU [5].

В таблице 2 приведены результаты тестирования трех моделей. Для каждой из них был использован порог уверенности, при котором результаты являются наилучшими. Для модели ResNet-34+U-Net порог уверенности составил 0,34, для моделей SegFormer_b5 и SegNeXt_1 – 0,50. Значение метрики IoU бралось как среднее IoU для верно определенных объектов.

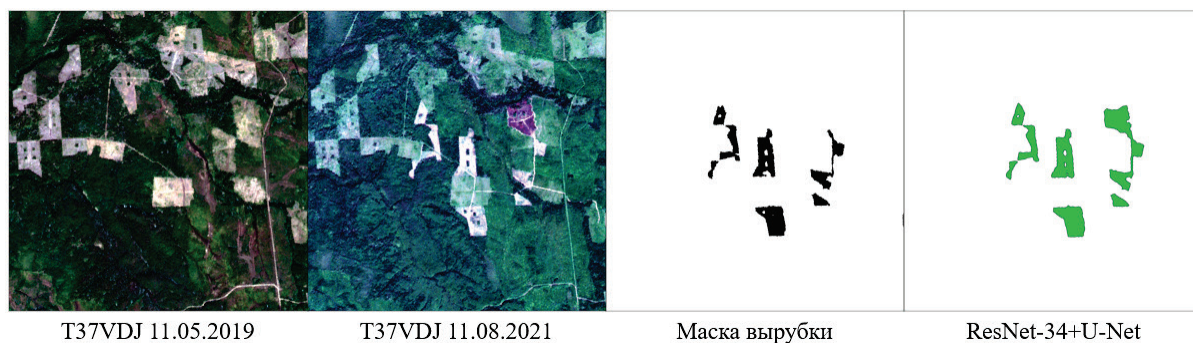


Рисунок 4 – Результат работы алгоритма на примере U-Net+ResNet

Из результатов видно, что все модели достаточно хорошо определяют границы найденных объектов, на что указывает высокое значение коэффициента Серенсена и метрики IoU. Трансформеры в данном случае отличаются тем, что у них очень много ложных срабатываний по отношению к ResNet-34+U-Net. Однако при этом у модели SegNeXt_1 меньше ложных пропусков, что также сказывается на большем значении метрики полноты. Наиболее слабой из 3 моделей оказался трансформер с архитектурой SegFormer_b5, обладающий меньшим количеством верных срабатываний и ложных пропусков.

Таблица 2 – Результаты тестирования моделей

Метрика	ResNet-34+U-Net	SegFormer b5	SegNeXt 1
Dice	0,82	0,82	0,84
TP_IoU	0,90	0,90	0,90
F1	0,74	0,65	0,67
Precision	0,71	0,57	0,57
Recall	0,78	0,77	0,80
TP	13608	13359	13894
FP	5435	10015	10270
FN	3719	3968	3433

5. Выводы

В ходе данного исследования был разработан алгоритм детектирования лесных вырубок на основе анализа спутниковых изображений с использованием методов глубокого обучения. В рамках работы рассматривались три нейросетевых модели типа кодировщик-декодировщик, популярных в задачах сегментации. Первая модель обладала архитектурой ResNet-34+U-Net, вторая – SegFormer_b5, а третья – SegNeXt_1. Для их обучения был подготовлен набор из 109 спутниковых изображений, которые были поделены на патчи размером 512x512 пикселей. Оценка качества работы обученных моделей происходила с помощью таких метрик как точность, полнота, f-мера и коэффициент Серенсена. Исследование показало, что удается с высокой точностью обучить алгоритм, способный находить изменения лесных массивов на парных изображениях, состоящих из снимков одной территории в разные моменты времени, что подтверждает актуальность исследования для задач оперативного картографирования. В данной работе сегментировался только один класс изменений – лесные вырубки. Наиболее эффективной с точки зрения количества детектированных вырубок оказался трансформер с архитектурой SegNeXt_1, который обладает наибольшим значением полноты и коэффициента Серенсена, равным 0,80 и 0,84 соответственно. При этом более сбалансированной является модель ResNet-34+U-Net, точность предсказаний которой выше остальных и составляет 0,71 в то время, как у моделей-трансформеров точность равно 0,57. Дальнейшими направлениями развития работы является расширение выборки классов для сегментации и исследование влияния времени года на качество сегментации.

6. СПИСОК ИСТОЧНИКОВ

- [1] Khryashchev V., Larionov R. Wildfire Segmentation on Satellite Images using Deep Learning, 2020 Moscow Workshop on Electronic and Networking Technologies (MWENT), 2020, pp. 1–5, doi: 10.1109/MWENT47943.2020.9067475.
- [2] Eremchenko E. What is and What is not the Digital Earth? // GraphiCon 2020: Proceedings of the 30th International Conference on Computer Graphics and Machine Vision (Saint Petersburg), 2020. Pp. 1-11.
- [3] Ларионов Р. В. Использование изображений с БПЛА для обнаружения незаконных свалок твердых бытовых отходов / Р. В. Ларионов, В. В. Хрящев, В. А. Павлов // Цифровая обработка сигналов и её применение (ЦОС-2021) : Доклады 23-ей Международной конференции, Москва, 24–26 марта 2021 года. – Москва: Российское научно-техническое общество радиотехники, электроники и связи им. А.С. Попова, 2021. – С. 209-213. – EDN MWJNLS.
- [4] Тарасов, А. В. Оценка точности методов маскирования облачности по снимкам Sentinel-2 и PlanetScope // Региональные проблемы дистанционного зондирования Земли материалы VI междунар. научн. конференции (Красноярск), 2019. С. 139–143.
- [5] Khryashchev V., Larionov R., Kotov N., Nazarovskiy A. Segmentation of Agricultural Fields on Microwave C-Band SAR Images // 2022 International Siberian Conference on Control and Communications (SIBCON), 2022. Pp. 1-5. Doi: 10.1109/SIBCON56144.2022.10002999.
- [6] Dergacheva E., Demidenko E. Visualizing Global Socio-Technogenic Human Transformation: Digital Challenges of Living Earth// GraphiCon 2020: Proceedings of the 30th International Conference on Computer Graphics and Machine Vision (Saint Petersburg), 2020. Pp. 1-14.
- [7] Copernicus [Электронный ресурс]. URL: <https://www.copernicus.eu/en> (дата обращения 12.07.2023).
- [8] Tavera A., Arnaudo E., Masone C., Caputo B. Augmentation Invariance and Adaptive Sampling in Semantic Segmentation of Agricultural Aerial Images // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2022. Pp. 1656-1665.
- [9] 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. 2015. Vol 9351. Pp. 234–241.
- [10] He. K., Zhang X., Ren S., Sun J. Deep Residual Learning for Image Recognition // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. Pp. 770-778.
- [11] Attention is all you need / Vaswani A. et al. // NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017. Pp. 6000-6010.
- [12] SegFormer: Simple and Efficient Design for Semantic Segmentation with Transformers / Enze X. et al. // Neural Information Processing Systems (NeurIPS), 2021. Pp. 1-14.
- [13] SegNeXt: Rethinking convolutional attention design for semantic segmentation / Guo M. et al. // Neural Information Processing Systems (NeurIPS), 2022. Pp. 1-15.
- [14] Imagenet: A large-scale hierarchical image database / Deng J. et al. // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2009. Pp. 248–255.
- [15] DeepGlobe 2018: A Challenge to Parse the Earth through Satellite Images / Demir I. et al. // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2018. Pp. 172-17209.