**Сегментация лесных массивов на спутниковых снимках   
с помощью глубокого обучения**

**Аннотация** – целью исследования являлось создание алгоритма глубокого обучения для автоматической сегментации лесных массивов на спутниковых снимках высокого разрешения. Для оценки качества разработанной модели использовался коэффициент Дайса, который сравнивает результаты алгоритмов с экспертной разметкой. Данная статья демонстрирует, как сверточная нейронная сеть, реализованная на современных графических процессорах, может применяться для обнаружения лесов на спутниковых изображениях.

**Ключевые слова** – компьютерное зрение, сегментация изображений, спутниковые снимки, обнаружение лесных массивов.

Введение

Дистанционное зондирование не может полностью заменить наземный сбор данных, но может помочь в мониторинге больших площадей труднодоступных регионов. И хотя алгоритмы компьютерного зрения для обнаружения объектов на изображениях быстро развиваются, методы автоматической сегментации спутниковых снимков все еще уступают человеку по качеству работы: задача сегментации снимков дистанционного зондирования земной поверхности не была доведена до автоматизма с такой же точностью, как при ручной разметке [1]. Человек способен решить проблему сегментации лучше, чем компьютер, однако это занимает слишком много времени. Кроме того, в этом случае невозможно получить результаты в режиме реального времени, поэтому на сегодняшний день задача сегментации спутниковых снимков с использованием алгоритмов компьютерного зрения особенно актуальна.

Существует множество алгоритмов детектирования объектов на изображениях. Проблема сегментации снимков успешно решается в биометрии, медицине и робототехнике [2]. Большинство алгоритмов сегментации могут быть применены и к задачам дистанционного зондирования. Сегментация спутниковых снимков является сложной задачей. Основным подходом к его решению на основе методов машинного обучения является классификация пикселей в соответствии с принадлежащим ему классом объекта. На сегодняшний день наилучшие результаты в этой области были продемонстрированы сверточными нейронными сетями. Уникальность данного подхода основана на автоматическом формировании признаков в процессе обучения алгоритма. Таким образом, достигается повышенная точность сегментации [3].

Более того, есть несколько причин, по которым сверточная нейронная сеть не является тривиальным решением сегментации изображения. Для решения проблемы пространственной протяженности обнаруженных объектов необходим уникальный подход, учитывающий инвариантность к повороту изображения или масштабированию [4]. Такие алгоритмы должны [5, 6]:

* Иметь достаточное количество обучающих примеров.

Как правило, в открытых базах данных недостаточно спутниковых изображений. Обучающие наборы данных могут быть увеличены путем самостоятельной разметки или объединения изображений из нескольких баз данных.

* Быть инвариантными к вращению.

Объекты на спутниковых снимках могут быть повернуты на любой угол. Алгоритм сегментации должен выделять границы объектов независимо от их положения на изображении.

* Учитывать малый размер некоторых объектов.

Большинство алгоритмов глубокого обучения способны выделять объекты больших размеров, как, например, в базе данных ImageNet [7]. Однако спутниковые изображения охватывают большую область, на которой, как правило, требуется выделить небольшие по сравнению со всей сценой объекты.

В проведенном исследовании были использованы спутниковые снимки, которые имеют следующие особенности:

* высокая повторяемость съемки, благодаря которой можно многократно получать данные интересующей территории, что увеличивает вероятность получения безоблачных или малооблачных изображений;
* большая площадь заснятой поверхности Земли с высоким пространственным разрешением;
* многоканальные фотографии в видимом и невидимом диапазонах, включая ультрафиолетовые и инфракрасные каналы.

Информация, полученная в процессе съемки по RGB-каналам, а также по ближнему инфракрасному каналу (NIR), имеет ряд особенностей с точки зрения их использования при анализе лесных площадей.

Синяя зона спектрального излучения активно поглощается хлорофиллом (преимущественно хлорофиллом В). Эта область очень чувствительна к атмосферным условиям, таким как туман или дымка. По сравнению с красными или NIR-каналами синий канал менее чувствителен к изменениям содержания хлорофилла. В результате, для решения задач сегментации он используется только для специальных целей, например, для мониторинга воды. Синий канал также облегчает распознавание лесных пожаров на безоблачных изображениях территории [8]. Поэтому чтобы решить задачу выделения лесных массивов, лучше всего использовать зеленые и красные каналы.

Растения в основном поглощают красный и синий спектры изображения, отражая большую часть зеленого. Зеленый канал в таком случае не только формирует изображение леса в целом, но и позволяет выделять на нем растительность. Это также необходимо для оценки общего состояния леса.

Красный канал очень важен для анализа растительности (в основном леса). Длина волны красного канала больше, чем синего. По этой причине атмосфера Земли влияет на него гораздо меньше, чем на остальные цветовые спектры. Красный канал играет решающую роль в анализе изменений лесного покрова, например, при оценке ущерба от стихийных бедствий, классификации видов растительности, мониторинга лесного покрова и т. д. [9].

Способность листвы отражать свет различается для разных видов деревьев. Отражающая способность хвои значительно ниже, чем у листвы. Значения (NDVIRE) ближнего инфракрасного канала молодого хвойного леса значительно выше, чем у старого. В связи с этим, ближний инфракрасный канал очень важен для классификации лесов, определения видового состава и для мониторинга заражения лесов вредителями. Ближний инфракрасный канал также играет ключевую роль в обнаружении последствий ураганных ветров и в настоящее время становится важным компонентом при расчете некоторых показателей, определяющих биофизические параметры растительности.

В данной статье представлена сверточная нейронная сеть, которую можно использовать для сегментации леса. Здесь также описан процесс обучения, тестирования и специальные метрики для оценки качества работы алгоритма.

Архитектура нейронной сети

В данном разделе описывается архитектура нейронной сети, которая использовалась для выделения лесных массивов на аэрофотоснимках, сделанных с двух разных спутников.

Разработанный алгоритм глубокого обучения основан на очень популярной и широко распространенной архитектуре под названием U-Net, которая представляет собой сверточную нейронную сеть, используемую для изображений. Первоначально он был создан для сегментации медицинских изображений [10].

U-Net представляет собой U-образную сверточную нейронную сеть, которая состоит из двух частей: кодировщика и декодировщика. Кодировщик и декодировщик представляют собой нейронные сети, состоящие из шести блоков. Каждый блок кодировщика включает в себя два сверточных слоя, за которыми следует функция активации ReLU и операция понижения дискретизации. Декодировщик также состоит из шести блоков, каждый из которых включает в себя операцию повышения дискретизации для увеличения пространственного размера карты признаков, слияние с соотвествующей картой признаков кодировщика, два сверточных слоя, за которыми следует функция активации ReLU. Последний слой сети выполняет операцию свертки с фильтром размера 1x1 для соотносения каждого пикселя определенному классу.

Поскольку наши изображения – четырехканальные (RGB и ближний инфракрасный (NIR) каналы), нами была модифицирована классическая архитектура U-Net, путем добавления ещё одного кодировщика, который отдельно принимает данные с ближнего инфракрасного канала. Результаты с обоих кодировщиков объединяются в центре сети, а также в каждом блоке декодировщика, как и в исходной сети U-Net (см. рис. 1).



Рис. 1. Архитектура сети U-Net с двумя кодировщиками

Модифицированная версия сети U-Net была реализована на языке Python с использованием фреймворка Tensorflow. Tensorflow - это высокопроизводительная библиотека, использующая графические процессоры для решения задач глубокого машинного обучения.

Численные результаты

Перед непосредственным обучением сети была выполнена предобработка имеющихся спутниковых снимков. База данных состояла из 17 изображений. Каждый канал спутникового снимка содержал 16-битные значения с импульсными шумами. Наличие таких выбросов отрицательно влияет на результаты и общее время обучения модели. Поэтому для начала было выполнено поканальное выравнивание цветовой гистограммы относительно значения канала с минимальными значениями, выбранными с помощью кумулятивной функции порогового распределения интенсивностей каналов. После этого этапа предобработки снимков каждый канал изображения содержал значения в диапазоне [0, 1] с выровненной цветовой гистограммой.

Второй этап состоял в разделении базы данных на обучающий и тестовый наборы. Такое разделение данных осуществлялось путем разрезания каждого исходного изображения по горизонтали, с учетом того, что 15% пикселей, принадлежащих необходимым объектам, гарантированно имелись в тестовой выборке.

Третий этап состоял в извлечении фрагментов из снимков из обучающей и тестовой выборки, поскольку размеры изображений в них варьировались от 700 до 12000 пикселей, что мешало бы обучить сеть с использованием таких данных из-за ограниченного объема памяти графического процессора. В итоге, каждый нарезанный фрагмент представлял собой изображение размером 512x512 пикселей, которое было извлечено из исходного спутникового снимка с помощью метода скользящего окна с шагом 256 пикселей. Примеры нарезанных фрагментов и соответствующих им масок показаны на рис. 1.



Рис. 1. Примеры фрагментов, нарезанных из спутниковых снимков, и соответствующих масок

Из обучающего набора данных также была выделена валидационная выборка для вычисления метрик и контроля процесса обучения. Размеры обучающей и тестовой выборки приведены в таблице 1.

Таблица 1. Размеры обучающей и тестовой выборки

|  |  |
| --- | --- |
| **Обучающая выборка** | |
| Общее количество нарезанных фрагментов | 14139 |
| Количество нарезанных фрагментов с объектами класса «лес» | 8655 |
| Количество нарезанных фрагментов без объектов класса «лес» | 5484 |
| **Тестовая выборка** | |
| Общее количество нарезанных фрагментов | 3282 |
| Количество нарезанных фрагментов с объектами класса «лес» | 1785 |
| Количество нарезанных фрагментов без объектов класса «лес» | 1497 |

Из таблицы 1 видно, что набор данных для обучения является несбалансированным, что негативно влияет на процесс обучения сети. Для уменьшения дисбаланса, каждый батч для обучения строился путем случайного выбора нарезанных фрагментов без объектов класса «лес» и такого же количества нарезанных фрагментов с объектами класса «лес». Такой способ формирования обучающих батчей показал лучшие результаты сегментации и обнаружения по сравнению со стандартной последовательной подачей изображений в сеть.

Кроме того, для дальнейшего увеличения размера обучающей выборки производилась аугментация данных: случайные зеркальные отражения (RF), повороты, пространственные сдвиги, сдвиги в масштабе (SSR) и добавление случайного шума в цветовой схеме HSV (RN), что значительно повышало качество итоговой сегментации.

Обучающий процесс контролировался с помощью функции потерь и коэффициента Дайса, вычисляемого на валидационной выборке данных. Коэффициент Дайса рассчитывался по формуле

,

где и - экспертная разметка и результаты работы алгоритма соответственно. Для обучения разработанной модели использовались две различные функции потерь:

1. Бинарная кросс-энтропия с использованием коэффициента Дайса (BCE + DL):



,

где

,

- метка класса для пикселя , а - значения вероятности принадлежности пикселя к классу «лес».

2. Фокальная функция потерь с использованием коэффициента Дайса (FL + DL):

где

фокальная функция потерь, а параметр .

Фокальная функция потерь была специально разработана для задач с сильно несбалансированными данными. Эта метрика превосходит других своих конкурентов на наборе данных COCO.

Были обучены четыре модели (BB – с учетом сбалансированного батча):

1. BCE + DL / BB / RF + SSR (BCE #1)

2. BCE + DL / BB / RF + SSR + RN (BCE #2)

3. FL + DL / RF + SSR (FL #1)

4. FL + DL / RF + SSR + HSV (FL #2)

В качестве алгоритма численной оптимизации был выбран оптимизатор адаптивной оценки моментов (Adam). На каждой итерации обучения модель обновляла свои веса после прогона через сеть сформированного батча из 16 нарезанных фрагментов. Классификатор заканчивал свое обучение после выполнения 100 эпох.

Модели обучались на суперкомпьютере NVIDIA DGX-1 Центра искусственного интеллекта и цифровой экономики ЯрГУ им. П.Г. Демидова. Графики функции потерь и коэффициента Дайса на валидационной выборке показаны на рис. 2.



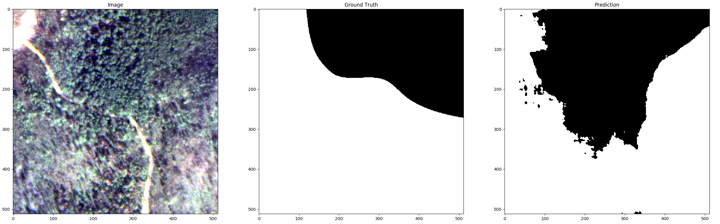
Рис. 2. Графики функции потерь и коэффициента Дайса на валидационной выборке

На выходе с каждой обученной модели выдавалось черно-белое изображение со значениями вероятности принадлежности пикселя классу «лес» в диапазоне от 0 до 1. Необходимо было выбрать соответствующее значение порога , для формирования бинарной маски. Для решения этой задачи использовался простой поиск нужного значения по сетке с шагом 0,01 с учетом максимизации значения . Выбранные значения порога для обученных моделей приведены в таблице 2.

Таблица 2. Результаты работы алгоритмов на тестовой выборке с учетом подобранных значений порога

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model** | **T** | **D** | **F1** | **P** | **R** |
| BCE #1 | 0.39 | 0.8439 | 0.1641 | 0.09475 | 0.6136 |
| BCE #1 | 0.5 | 0.8446 | 0.17 | 0.09873 | 0.6105 |
| **BCE #2** | **0.62** | **0.7652** | **0.3488** | **0.248** | **0.5876** |
| BCE #2 | 0.5 | 0.7554 | 0.3319 | 0.2309 | 0.59 |
| FL #1 | 0.52 | 0.8222 | 0.1967 | 0.1168 | 0.6211 |
| FL #1 | 0.5 | 0.8181 | 0.1775 | 0.1038 | 0.6144 |
| FL #2 | 0.56 | 0.7874 | 0.1549 | 0.09027 | 0.5443 |
| FL #2 | 0.5 | 0.7632 | 0.1288 | 0.07269 | 0.5664 |

Результаты сегментации лесных массивов для нарезанных фрагментов показаны на рис. 3. В первой колонке – нарезанные фрагменты спутниковых снимков, во второй – экспертная разметка, в третьей – результаты работы алгоритма глубокого обучения (BCE #2).



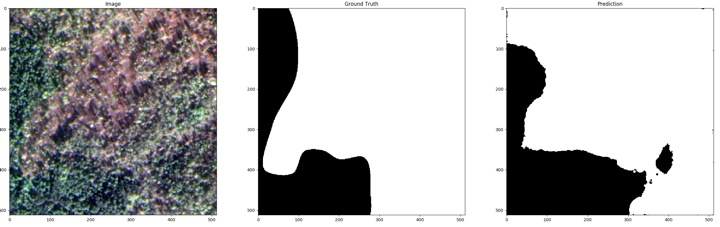


Рис. 3. Результаты сегментации лесных массивов для нарезанных фрагментов

Заключение

В статье были показаны результаты обучения и тестирования разработанной сверточной нейронной сети, предназначенной для сегментации спутниковых изображений. Для работы с четырехканальными снимками была предложена модифицированная архитектура U-Net с двумя кодировщиками. Разработанный алгоритм был предварительно обучен на наборе данных Spacenet. Значения коэффициента Дайса и индекса Жаккара на данных из приватной базы были равными 0,78 и 0,67 соответственно.

В этой статье было продемонстрировано, что разработанная модифицированная сеть U-Net с обучением на сбаланированной выборке способна проводить сегментацию лесных массивов на спутниковых снимках высокого разрешения. Невысокие результаты алгоритма глубокого обучения (максимальное значение составляет всего 0,34), могут быть связаны с неправильной или неполной сегментацией спутниковых снимков экспертами. Необходимы дальнейшие исследования для повышения качества сегментации снимков. Будущие направления исследования могут включать в себя увеличение числа спутниковых снимков в наборе данных с помощью дополнительных методов аугментации, а также обучение модифицированной версии U-Net с другими функциями потерями, которые смогут разрешить проблему несбалансированных данных.

Благодарности

Статья была подготовлена при финансовой поддержке Министерства образования Российской Федерации в рамках научного проекта № 14.575.21.0167 (идентификатор RFMEFI57517X0167).

Литература

1. Sayfeddin D., Bulgkov A., Kruglova T. Neirosetevaya sistema otslejivaniya mestopolozheniya dinamicheskogo agenta na baze kvadrokoptera // Inzhenernyi vestnik Dona, 2014, No. 1, URL: http://ivdon.ru/magazine/archive/n1y2014/2293/
2. Plutogarenko N., Varnavskiy A. Primenenie neironnykh setei dlya postroeniya modeli prognozirovaniya sostoyaniya gorodskoi vozdushnoi sredy // Inzhenernyi vestnik Dona, 2012, No. 4-2, URL: http://ivdon.ru/magazine/archive/n4p2y2012/1351/
3. L. Deng, D. Yu Deep Learning : Methods and Application Foundations and Trends in Signal Processing. 2014, vol. 7, no. 3-4, pp. 197-387.
4. Solov’ev R. A., Telpukhov D. V., Kustov A. G., Avtomaticheskaya segmentatsiya sputnikovykh snimkov na baze modifitsirovannoi svertochnoi neironnoi seti UNET (Automatic segmentation of satellite images based on the modified UNET convolutional neural network), Inzhenernyi vestnik Dona, 2017, Vol. 47, No. 4(47), URL: http://ivdon.ru/uploads/article/pdf/IVD\_56\_soloviev\_N.pdf\_116222c2f5.pdf.
5. V.Khryashchev, L.Ivanovsky, V.Pavlov, A.Ostrovskaya, A.Rubtsov Comparison of Different Convolutional Neural Network Architectures for Satellite Image Segmentation // Proceedings of the FRUCT’23, Bologna, Italy, 13–16 November 2018. pp.172–179.
6. Khryashchev V., Pavlov V., Priorov A., Ostrovskaya A. Deep learning for region detection in high-resolution aerial images // IEEE East-West Design & Test Symposium (EWDTS). 2018. pp. 1–5.
7. ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC2012), Web: http://image-net.org/challenges/LSVRC/2012/
8. W G.. Rees Physical principles of Remote Sensing. Cambridge University Press, 2006, 336 p.
9. Campell J.B. Introduction to remote sensing / J.B. Campell.–N.Y.–London: The Guilford press, 1996–P. 120–549.
10. O. Ronneberger, P. Fischer, T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, LNCS, vol. 9351, 2015, pp. 234–241.