Глубокое обучение для детектирования зданий на спутниковых снимках высокого разрешения

Владимир Хрящев 1, Леонид Ивановский 1, Анна Островская 2

1 Ярославский государственный университет им П.Г. Демидова, 150003 Ярославль, Россия,

ул. Советская д. 14,

e-mail: v.khryashchev@uniyar.ac.ru, leon.ivanovsky@yahoo.com

2 Российский университет дружбы народов (РУДН), 117198 Москва, Россия, ул. Миклухо-Маклая д. 6, e-mail: ostrovskaya-aa@rudn.university

Аннотация. В данной статье представлены результаты исследования сверточной нейронной сети для детектирования зданий на спутниковых снимках высокого разрешения базы данных Planet. Индекс Жакара был использован для анализа качества алгоритма машинного обучения. Этот индекс сравнивает результаты алгоритмов с реальными масками. Маски были разрезаны на более мелкие части вместе с изображениями перед обучением разработанной модели. Разработанная сверточная нейронная сеть была запущена на суперкомпьютере NVIDIA DGX-1, предоставленном Центром искусственного интеллекта и цифровой экономики ЯрГУ им. П.Г. Демидова. Задача детектирования зданий на спутниковых изображениях может быть применена на практике для городского планирования, контроля строительства муниципальных объектов, поиска лучших мест для торговых точек и т. д.

Ключевые слова: машинное обучение, сегментация спутниковых снимков, детектирование зданий.

1. Введение

В настоящее время задача обнаружения объектов на спутниковых изображениях высокого разрешения находится в центре внимания исследователей. Автоматическая сегментация изображения позволяет выделять интересующие объекты, такие как здания, дороги или транспортные средства. Алгоритмы глубокого обучения способны решить эту проблему. Тем не менее, использование этих алгоритмов вместо традиционных методов машинного обучения нетривиально [1]. Такие модели должны

* Учитывать малый размер некоторых объектов.
* Быть инвариантным к поворотам объектов.
* Иметь достаточно обучающих примеров.
* Иметь достаточно обучающих примеров.
* Иметь достаточно обучающих примеров.

В машинном обучении задача сегментации изображений обычно переформулируется как задача классификации на уровне пикселей. Наиболее простой и медленный подход к решению этой проблемы – ручная сегментация изображений. Тем не менее, это трудоемкий и долгий процесс, который подвержен ошибкам, неизбежным при использовании монотонного ручного труда. Поэтому большой интерес исследователей в области машинного обучения направлен на разработку систем автоматической сегментации снимков. Подобного рода сегментация позволяет обеспечить обработку изображений «на лету» сразу его после получения. Сегментация спутниковых снимков находит свое применение в области городского планирования, контроля строительства муниципальных объектов и вырубки лесов, грамотного использования сельскохозяйственных угодий метеорологии и т.д.

В данной статье представлены результаты исследования работы сверточных нейронных сетей. Основным их преимуществом является то, что они могут обнаруживать и классифицировать объекты в режиме реального времени, и в то же время они превосходят по качеству традиционные методы машинного обучения. Дескрипторы изображений формируются алгоритмом самостоятельно, в процессе обучения. Структура этих моделей параллельна и как следствие она отлично подходит для обучения на графических процессорах, которые состоят из тысячи ядер, которые вместе способны решать множество задач одновременно [2].

Один из самых успешных алгоритмов сегментации основан на полносвязных нейронных сетей. Основная идея таких сетей заключается в использовании в конце полносвязного слоя вместе с операцией свертки, в то время как предыдущие слои сети извлекают необходимые признаки из входных данных [3]. Полносвязные нейронные сети показывают отличные результаты обнаружения объектов на наборе данных PASCAL VOC 2011 [4].

Вскоре полносвязные нейронные сети были улучшены до сети U-Net, который получила свое развитие при решении задачи сегментации биомедицинских изображений. [5]. Позже эта модель показала приемлемые результаты обнаружения объектов на спутниковых снимках. Архитектура U-Net использует пропускаемые соединения для объединения признаков, полученных на разных этапах работы сети. Используя архитектуру U-Net, авторы статьи получили значение меры Серенсена равное 0,75 для детектирования зданий на спутниковых снимках из набора данных Inria [7]. В работе [8] U-Net показывает хорошие результаты по обнаружению деревьев, сельскохозяйственных угодий и водных объектов: значение коэффициента Серенсена тоже оказалось равным 0,75.

Эта статья состоит из шести частей. Первая часть вводит читателей в суть поставленной проблемы. В ней также приведен обзор работ, связанных с задачей детектирования объектов на изображениях высокого разрешения. Второй раздел посвящен доступным базам данных спутниковых снимков. В третьей части описываются разработанные архитектуры сверточных нейронных сетей для выделения зданий на спутниковых снимках и некоторые особенности их обучения. Более того, в этом разделе упоминаются фреймворк Keras для создания алгоритмов глубокого обучения. В четвертой части представлены результаты численных экспериментов для разработанных алгоритмов. В заключении подводятся итоги исследования. И наконец, в конце содержится список использованных литературных источников.

1. Базы спутниковых снимков

База данных SpaceNet [9] покрывает собой территорию 6 крупных городских агломераций: Рио-де-Жанейро, Лас-Вегаса, Парижа, Шанхая, Хартума и Аталанты. Восьмиканальные снимки, сделанные со спутников WorldView-2 и WorldView-3, имеют различное пространственное разрешение. База данных также разбита на подмножества данных, в зависимости от сегментированных на них объектов. Так, например, в SpaceNet содержится два набора данных, суммарно покрывающих территории в 3011 км² и 5555 км², для решения задачи выделения зданий на спутниковых снимках. Примеры изображений из базы SpaceNet показаны на рис. 1.

|  |
| --- |
| commenter |

Рис. 1: Примеры изображений из базы данных SpaceNet.

База данных Pleiades-1B [10] содержит цветные изображения со спутника Pleiades-1B. Этот спутник был успешно запущен 2 декабря 2012 года для аэрофотосьемки крупных территорий, включая инженерные объекты, горно-промышленные комплексы, а также стихийные бедствия и спасательные работы. Этот набор данных был запечатлен под разными углами съемки. Каждое изображение базы данных Pleiades-1B имеет пространственное разрешение 0,5 м/пиксель. Примеры изображений из базы данных Pleiades-1B показаны на рис. 2.

|  |
| --- |
| commenter |

Рис. 2: Примеры изображений из базы данных Pleiades-1B.

Для исследования работы алгоритмов выделения зданий были использованы спутниковые снимки из базы данных PlanetScope. 14 изображений из PlanetScope размером порядка 8192x8192 пикселей имеют пространственное разрешение 0.5 м/пиксель и покрывают территорию трех российских городов: Москвы, Ярославля и Рыбинска. Примеры изображений из базы PlanetScope показаны на рис. 3.

|  |
| --- |
| commenter |

Рис. 3: Примеры изображений из базы данных PlanetScope.

1. Алгоритм глубокого обучения
   1. Сверточная нейронная сеть

В этой статье представлена разработанная сверточная нейронная сеть - U-Net. Сверточные нейронные сети направлены на быстрое и качественное обнаружение различных объектов [11]. Проделанное исследование работы U-Net продолжает исследования, которые были представлены в работах [8, 12].

Библиотека Keras с использованием фреймворка Tensorflow в качестве бэкэнда была использована для разработки сверточных нейронных сетей. Keras - это библиотека с открытым исходным кодом, написанная на языке Python. Keras предоставляет множество реализаций широко используемых структурных блоков нейронной сети, а также готовых инструментов для предварительной обработки изображений и текстовых данных. Другими словами, Keras предлагает более интуитивный набор абстракций более высокого уровня для разработки алгоритмов глубокого обучения [13]. Кроме того, эта библиотека позволяет обучать разработанные модели на графическом процессоре.

Как показано на рис.4, U-Net состоит из двух частей: кодировщика (слева) и декодировщика (справа). Кодировщик представляет собой нейронную сеть с типичной архитектурой сверточной нейронной сети. Декодировщик имеет схожую структуру, как и кодировщик, не считая операция слияния и операцию понижения дискретизации, которая была заменена слоем повышения дискретизации. Последний слой сети выполняет операцию свертки с фильтром размера 1x1 для соотносения каждого пикселя определенному классу. В итоге, сеть имеет 19 сверточных слоев, 18 функций активации ReLU, 4 операций понижения дискретизации, 4 операций upsampling и 4 операций слияния признаков. Суммарно, сеть U-Net имеет около 7.9 млн параметров обучения.

|  |
| --- |
| commenter |

Рис. 4: U-Net.

* 1. Особенности обучения

Подход, основанный на использовании сверточных нейронных сетей, требует немалых вычислительных ресурсов. В связи с этим, этапы обучения и тестирования выполнялись на большом количестве независимых потоков графического процессора с использованием технологии параллельных вычислений NVIDIA CUDA. Эта кросс-платформенная технология, поддерживаемая всеми современными видеокартами NVIDIA [14]. В качестве алгоритма численной оптимизации был выбран оптимизатор адаптивной оценки моментов. Этот оптимизирующая функция объединяет лучшие подходы от градиентного спуска и импульсных оптимизаторов, показывая быструю сходимость для большинства задач машинного обучения [15]. В качестве функции потерь была выбрана бинарная кросс-энтропия. Классификатор заканчивал свое обучение после выполнения 100 эпох, обновляя свои веса после прогона через сеть сформированного батча из 36 образцов

|  |
| --- |
| commenter |

Рис. 5: Примеры изображения и соответствующей ему сгенерированной маски.

Обучение и тестирование разработанных моделей было осуществлено на спутниковых снимках из базы данных PlanetScope. Информация о местоположении зданий была получена из json-файла и сгенерирована в виде черно-белых масок, в которых каждый пиксель, относящийся к зданиям, окрашивался в белый цвет. Примеры изображения из базы данных PlanetScope и соответствующей ему маски показаны рис. 5.

Традиционный подход при сегментации изображений основан на использовании частей снимков, которые соответствуют входу сверточной нейронной сети. Разработанная сеть U-Net требует на вход данные размером 512 × 512 пикселей, поэтому перед обучением моделей, каждый спутниковый снимок и соответствующая ему маска нарезались на части соответствующего размера с помощью метода скользящего окна. Примеры нарезанных спутниковых снимков и соответствующих им масок показаны на рис. 6.

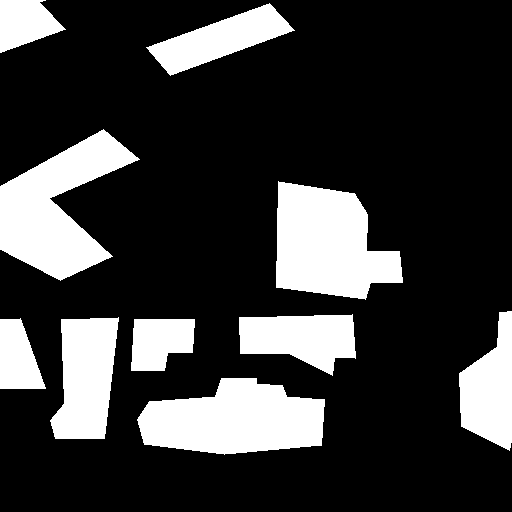


Рис. 6: Примеры нарезанных изображений и соответствующих им масок.

В результате подготовленная выборка данных содержала 3264 картинки размером 512 × 512 пикселей. Для моделирования численных экспериментов множество данных было разбито на обучающую и валидационную выборки в отношении 80/20. Таким образом, обучающая выборка содержала 2611 картинок, а валидационная – 653 картинки. Обучающая и валидационная выборки содержали одинаковых данных. Для поставленной задачи брались только 2 класса: «здания» и «не здания».

1. Результаты численных экспериментов

Разработанные сверточные нейронные сети были обучены и протестированы на суперкомьютере NVIDIA DGX-1, предоставленным Центром искусственного интеллекта и цифровой экономики ЯрГУ им. П.Г. Демидова.

Как правило, качество работы алгоритмов сегментации обычно оценивается специальными метриками, осуществляющими сравнение предсказанных и экспертных масок изображений между собой. Для оценки разработанных моделей использовался индекс Жаккара (). Этот показатель принимает значения из отрезка [0, 1] и фактически является бинарной мерой сходства двух множеств. Индекс Жаккара рассчитывается по следующей формуле:

где – множество масок, размеченных экспертами, и - результаты предсказаний [16]. Зависимость значения индекса Жаккара от числа выполненных обучающих эпох () для сети U-Net показана на рис. 7.

|  |
| --- |
| commenter |

Рис. 7: Зависимость индекса Жаккара от числа выполненных обучающих эпох для разработанной сети

Согласно результатам, приведенным на рис. 8, U-Net показала приемлемые результаты: значение индекса Жаккара оказалось равным 0.65. Несмотря на сложность архитектуры сети, обучение модели длилось порядка 2 ч.

|  |
| --- |
| commenter |

Fig. 8: Test results.

Результаты работы сети U-Net по детектированию зданий показаны на рис. 8. Разработанная модель детектирует хуже вблизи границ нарезанных изображений. Это хорошо известная проблема сети U-Net, которая может возникнуть при сегментации спутниковых снимков. Эта проблема решается нарезанием различных частей из большого спутникового снимка и сочетанием полученных результатов.

Разработанная модель может быть использована для поиска на спутниковых снимках лучших мест для торговых точек. U-Net способна выделять жилые районы, а следовательно с ее помощью можно рассчитать примерное количество жителей, проживающих в квартале, и как следствие, количество возможных покупателей при размещении магазина в определенном месте. Это позволяет выбрать лучшее место для расположения торговых точек.

|  |
| --- |
| commenter |

Рис. 9: Применение разработанной сверточной нейронной сети для поиска лучшего места для торговых точек.

1. Заключение

В этой статье представлены результаты исследований U-Net для обнаружения зданий на спутниковых снимках высокого разрешения базы данных PlanetScope. Перед обучением разработанной модели изображения и соответствующие им маски разрезали на более мелкие части. Обучение сверточной нейронной сети длилось около 2 ч на графическом процессоре суперкомпьютера NVIDIA DGX-1, предоставленного Центром искусственного интеллекта и цифровой экономики ЯрГУ им. П.Г. Демидова. Индекс Жакара был использован для анализа качества алгоритма машинного обучения. Этот индекс является бинарной мерой сходства результатов алгоритмов с масками, размеченными экспертами. Для U-Net значение индекса Жакара оказалось равным 0,65. Разработанная модель может быть использована для обнаружения зданий на спутниковых снимках. Задача детектирования зданий на спутниковых изображениях может быть применена на практике для городского планирования, контроля строительства муниципальных объектов или поиска лучших мест для торговых точек.

1. Благодарности

Численные эксперименты для сравнительного анализа разработанных алгоритмов проводились на спутниковых снимках приватной базы данных. Для выполнения численных экспериментов каждый снимок и соответствующая ему маска, сгенерированная из json-файла, нарезались на части меньшего размера. Использование специальной метрики сходства предсказанных и экспертных масок изображений показало, что алгоритм U-Net позволяет получить лучшие результаты в сравнении с LinkNet. Значение коэффициента Серенсена для сети U-Net оказалось равным 0.77. Обе разработанные модели оказались просты в реализации.

Статья была подготовлена при финансовой поддержке Министерства образования Российской Федерации в рамках научного проекта № 14.575.21.0167 (идентификатор RFMEFI57517X0167). Авторы также благодарны Центру Искусственного Интеллекта Ярославского государственного университета им. П.Г. Демидова за обеспечение доступа к суперкомпьютеру NVIDIA DGX-1.

1. Список литературы
2. Я. Гудфеллоу, И. Бенджио, А. Курвилль, Глубокое обучение. М.: ДМК Пресс, 2017, 652 с.
3. Н. Шакла, Машинное обучение & Tensorflow. СПб: Питер, 2019, 336 с.
4. E. Shelhamer, J. Long, T. Darrell, “Fully Convolutional Networks for Semantic Segmentation”, Web: https://arxiv.org/pdf/1605.06211.pdf.
5. PASCAL VOC 2011 challenge, Web: http://host.robots.ox.ac.uk/pascal/VOC/voc2011/index.html.
6. 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. 234–341.
7. G. Chhor, C. Bartolome Aramburu, I. Bougdal-Lambert, “Satellite Image Segmentation for Building Detection using U-net”, Web: http://cs229.stanford.edu/proj2017/final-reports/5243715.pdf.
8. Inria Aerial Image Labelling Database, Web: https://project.inria.fr/aerialimagelabeling/.
9. V. Khryashchev, L. Ivanovsky, V. Pavlov, A. Ostrovskaya and A. Rubtsov, "Comparison of Different Convolutional Neural Network Architectures for Satellite Image Segmentation. Proceedings of the 23rd Conference of Open Innovations Association FRUCT’23, Bologna, Italy. 2018, pp. 172-179.
10. SpaceNet Database, Web: http://explore.digitalglobe.com/spacenet.
11. Pleiades-1B Satellite Sensor, Web: https://www.satimagingcorp.com/satellite-sensors/pleiades-1b/.
12. J. Patravali, S. Jain, S. Chilamkurthy, “2D-3D Fully Convolutional Neural Networks for Cardiac MR Segmentation”, Qure.AI, 2017.
13. V. Khryashchev, A. Priorov, V. Pavlov, A. Ostrovskaya, “Deep learning for region detection in high-resolution aerial images”, Proceedings of 16-th IEEE East-West Design & Test Symposium (EWDTS’2016), Kazan, Russia, September 14 - 17, 2018. P. 792-796.
14. А. Джулли, С. Пал, Библиотека Keras – инструмент глубокого обучения. М.: ДМК Пресс, 2018, 294 с.
15. Дж. Сандерс, Э. Кэндрот, Технология CUDA в примерах: введение в программирование графических процессоров. М.: ДМК Пресс, 2013, 232 с.
16. D. P. Kingma, J. Ba. Adam: A Method for Stochastic Optimization. Web: https://arxiv.org/abs/1412.6980.
17. J. Bell, H.M. Dee, "The subset-matched Jaccard index for evaluation of Segmentation for Plant Images", Web: https://arxiv.org/pdf/1611.06880.pdf.
18. Дж. Вандер Плас, Python для сложных задач: наука о данных и машинное обучение. СПб.: Питер, 2018. 576 с.
19. А. Мюллер, С. Гвидо, Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными. СПб: ООО "Алфа-книга", 2018, 480 с.

Сведения об авторах

|  |  |  |  |
| --- | --- | --- | --- |
| Имя, фамилия | Title | Research Field | Personal website |
| Владимир Хрящев | Канд. техн. наук | Компьютерное зрение, глубокое обучение, большие данные |  |
| Леонид Ивановский | Аспирант | Машинное обучение, большие данные, программирование |  |
| Анна Островская |  |  |  |