Анализ урбанистических объектов с помощью  
сегментации спутниковых снимков и глубокого обучения

Леонид Иванеовский1, Анна Островская2, Владимир Хрящев1

1 ЯрГУ им. П.Г Демидова, ул. Советская 14,   
150003 Ярославль, РФ

2 Российский Университет Дружбы Народов (РУДН), ул. Миклухо-Маклая д.6, 117198 Москва, РФ  
[leon.ivanovsky@yahoo.com](mailto:leon.ivanovsky@yahoo.com), ostrovskaya\_aa@rudn.university, v.khryashchev@uniyar.ac.ru

**Аннотация.** Целью нашего исследования была разработка методов на основе сверточных нейронных сетей для автоматического выделения зданий на спутниковых снимках. Для анализа качества разработанных алгоритмов использовался коэффициент подобия Серенсена, сопоставляющий друг с другом полученные алгоритмом результаты и области, ранее размеченные экспертами. Маски были автоматически сгенерированы из json-файлов и порезаны на более мелкие части вкупе со спутниковыми снимками перед обучением сверточных нейронных сетей. Такой подход позволил решить проблему сегментации изображений большого разрешения. В статье было показано, как сверточные нейронные сети, реализованные на современных графических процессорах, могут использоваться для эффективного обнаружения объектов. Проблема детектирования зданий на спутниковых снимках может быть применена на практике для городского планирования, контроля строительства муниципальных объектов, поиска наилучшего расположения магазинов и т. д.

**Keywords:** алгоритмы глубокого обучения, сегментация спутниковых снимков, детектирование зданий.

1. Введение

На сегодняшний день проблема детектирования зданий на спутниковых снимках высокого разрешения находится в фокусе научных исследований. Автоматическая сегментация изображений позволяет выделять интересующие области на снимке. Большинство подходов по решению данной проблемы предполагают разработку алгоритмов глубокого машинного обучения.

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. Автоматическая 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, 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. 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 [1]. Сверточные нейронные сети стали повсеместно использоваться в компьютерном зрении с тех пор, как алгоритм AlexNet [2] победил в конкурсе ImageNet Challenge: ILSVRC 2012 [3].

Задача сегментации спутниковых снимков довольно трудная. Большинство подходов по решению данной проблемы предполагают использование сверточных нейронных сетей. Особенность таких сетей заключается в том, что дескрипторы изображений формируются алгоритмом самостоятельно, в процессе обучения. В последние годы были предложены различные алгоритмы глубокого обучения, нацеленные на сегментацию изображений.

Один из самых успешных алгоритмов сегментации основан на использовании пирамидальных нейронных сетях. Пирамидальные сети показали приемлемые результаты детектирования объектов на спутниковых снимках [4]. Использование пирамидальных сетей позволило получить значение индекса Жаккара равное 0.49 для сегментации изображений со спутника DeepGlobe [5].

В статье [6] представлена архитектура U-Net - особый тип пирамидальных сетей, который получила свое развитие при решении задачи сегментации биомедицинских изображений. Позже эта модель была применена для классификации спутниковых изображений на уровне пикселей [7]. Архитектура U-Net использует пропускаемые соединения для объединения признаков, полученных на разных этапах работы сети.

В работе [8] представлена нейронная сеть LinkNet. Это специальная архитектура, которая состоит из кодировщика и декодировщика, также как и U-Net. Она эффективно получает информацию о признаках из кодировщика после выполнения соответствующего блока декодировщика. Такая техника извлечения признаков позволила получить высокие результаты по детектированию объектов на базе CamVid [9].

Эта статья состоит из шести частей. Первая часть посвящена использованию сверточных нейронных сетей как методу машинного обучения и особенностям задачи сегментации спутниковых снимков. 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 и библиотека Tensorflow 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

Базы изображений являются важнейшей составляющей успешного обучения и оценки качества работы алгоритмов. На сегодняшний день существует несколько доступных баз спутниковых снимков.

База данных SuperView-1 [10] содержит цветные многоканальные изображения с одноименного спутника, который был запущен на орбиту 9 января 2018 г. с космодрома Тайюань в Китае. Спутниковые изображения покрывают территорию площадью 900'000 км². Примеры снимков из этой базы показаны на рис. 1.



**Рис. .** Примеры изображений базы данных SuperView-1.

База данных GeoEye-1 [11] содержит четырехканальные изображения (красный, зеленый, синий и инфракрасный каналы). Спутник GeoEye-1 был запущен 6 сентября 2008 г. с военно-воздушной базы Ванденберг в США. Спутниковые снимки GeoEye-1 применяются для мониторинга окружающей среды, а также в области горного дела, машиностроения, археологии и сельского хозяйства. Примеры изображений из базы данных GeoEye-1 показаны на рис. 2.



**Рис. 2.** Примеры изображений базы данных GeoEye-1.

Для нашего исследования были использованы 14 цветных спутниковых снимка из приватной базы данных российских городов. Каждый снимок размером порядка 8192x8192 пикселей имеет пространственное разрешение 0.5 м/пиксель. Изображения этой базы данных покрывают территорию трех российских городов: Москвы, Ярославля и Рыбинска. Примеры снимков из этой базы показаны на рис. 3.



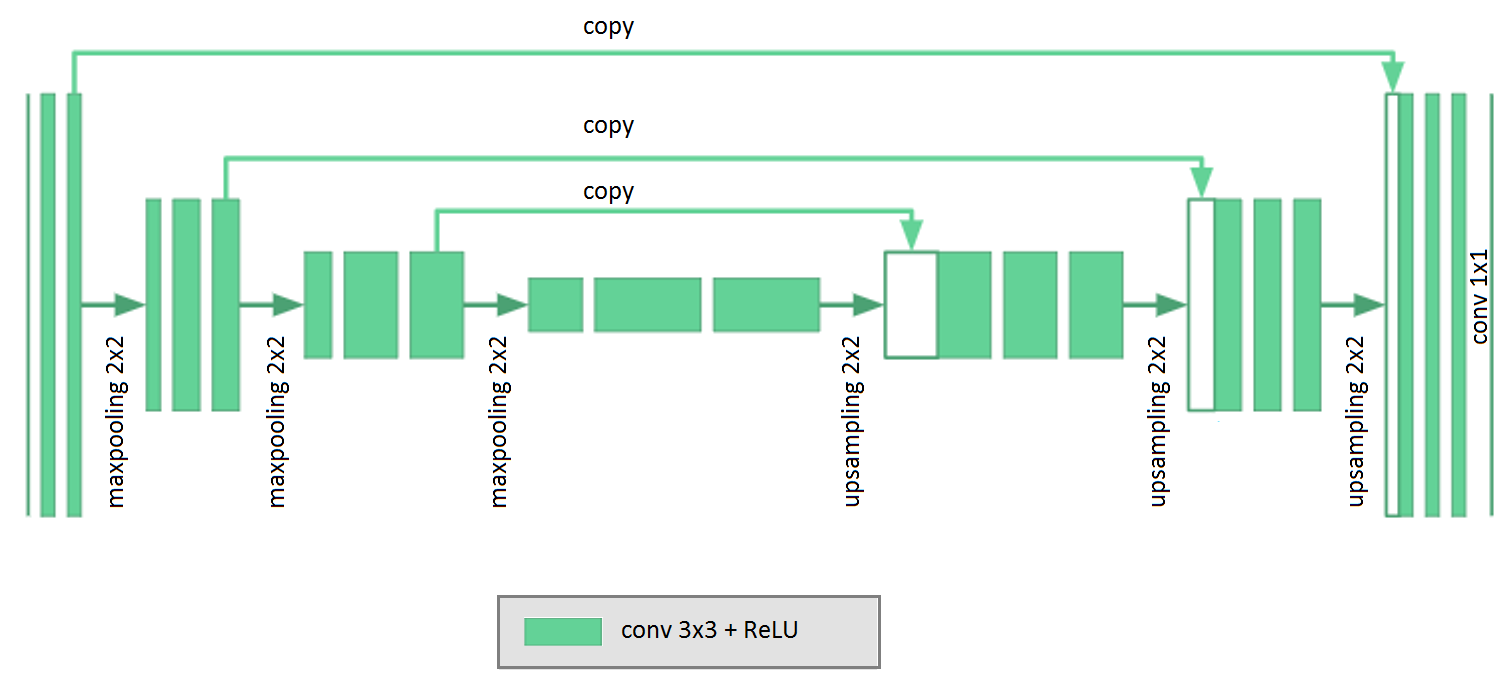
**Рис. 3.** Примеры изображений из приватной базы данных.

1. Алгоритмы глубокого обучения

В данной статье представлены разработанные архитектуры сверточных нейронных сетей. Это специальная архитектура нейронной сети, направленная на быстрое и качественное обнаружение различных объектов [12]. В данном исследовании был проделан сравнительный анализ работы сетей U-Net [7] и LinkNet [8]. Результаты, приведенные в данной работе, продолжают исследования, представленные в статьях [13, 14].

Все разработанные сверточные нейронные сети были реализованы с помощью библиотеки Keras с использованием фреймворка Tensorflow в качестве бэкэнда. Keras - это библиотека с открытым исходным кодом, написанная на языке программирования Python. Она предоставляет множество реализаций широко используемых структурных блоков нейронной сети, а также готовых инструментов для предварительной обработки изображений и текстовых данных [15]. Кроме того, эта библиотека позволяет обучать модели на графических процессорах видеокарты.

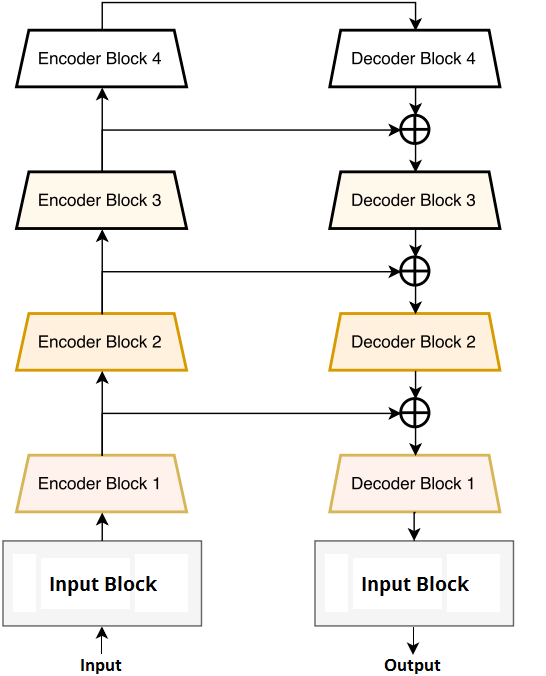
TensorFlow - это программная библиотека с открытым исходным кодом для высокопроизводительных численных расчетов. Этот фреймворк используется также в области машинного обучения, например для создания нейронных сетей. Tensorflow нацелен на быстрое обнаружение и классификацию изображений, достигая качества человеческого восприятия [16].



**Рис. 4.** Архитектура сети U-Net.

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, состоящей из четырех блоков. Каждый такой блок состоит из двух сверточных слоев с фильтром 3 × 3, с примененной к каждому из них функции активации ReLU, а также операции понижения дискретизации с помощью оператора maxpooling с размером фильтра 2 × 2 и шагом 2. Декодировщик содержит такое же количество блоков, как и кодировщик. Каждый блок декодировщика состоит из операций повышения дискретизации с помощью оператора upsampling с размером фильтра 2 × 2 и объединения с соответствующим набором признаков от кодировщика, двух сверточных слоев с фильтром 3 × 3 и примененной к каждому из них функции активации ReLU. 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.

Как и описанная выше архитектура сети U-Net, LinkNet имеет две части: кодировщик и декодировщик. Согласно архитектуре сети, указанной на рис. 5, обе подсети состоят из 4 блоков. Каждый блок кодировщика содержит 4 сверточных слоя, 2 слоя слияния признаков и 1 операции понижения дискретизации с помощью оператора maxpooling. В соответствии со структурой блока кодировщика блок декодировщика имеет схожую архитектуру, за исключением слоев слияния признаков и применения оператора maxpooling, который был заменен на операцию повышения частоты дискретизации с помощью оператора upsampling. Более того, перед отправлением карты признаков в кодировщик, к картинке последовательно применялись операции пакетной нормализации, активации ReLU, свертки с размером фильтра 2 × 2 и понижения дискретизации с помощью оператора maxpooling. После выполнения последнего блока декодировщика, над картой признаков дважды подряд выполнялась так же последовательность из операций повышения дискретизации с помощью оператора upsampling, пакетной нормализации, активации ReLU и сверточного слоя с размером фильтра 2 × 2.



**Рис. 5.** Архитектура сети LinkNet.

В качестве алгоритма численной оптимизации был выбран оптимизатор адаптивной оценки моментов (Adam) со скоростью обучения 1e-3. Этот оптимизирующая функция объединяет лучшие подходы от градиентного спуска и импульсных оптимизаторов, показывая быструю сходимость для большинства задач машинного обучения [17]. В качестве функции потерь была выбрана бинарная кросс-энтропия На каждой итерации обучения модель обновляла свои веса после прогона через сеть сформированного батча из 18 образцов. Классификатор заканчивал свое обучение после выполнения 96 эпох.

1. Результаты экспериментов

Численные эксперименты для разработанных моделей глубокого обучения были проведены на спутниковых снимков приватной базы данных. Информация о местоположении зданий была извлечена из json-файлов, в результате чего были сгенерированы черно-белые маски, где белый пиксель обозначал принадлежность к строениям.

Традиционный подход при сегментации изображений основан на использовании частей снимков, которые соответствуют входу сверточной нейронной сети. Разработанные алгоритмы требуют 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.

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. В нашем исследовании учитывались только 2 класса: «здания» и «не здания».

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 коэффициент подобия Серенсена (DSC). Этот показатель принимает значения из отрезка [0, 1] и фактически является бинарной мерой сходства двух множеств. Коэффициент Серенсена рассчитывается по следующей формуле:

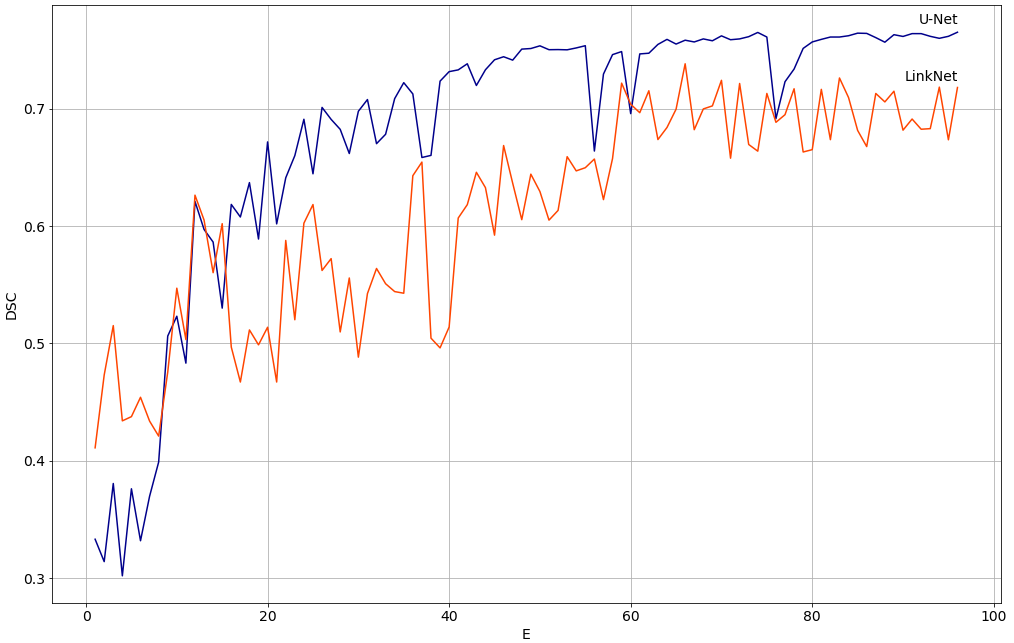
()

где - мощность пересечения и - сумма мощностей экспертной разметки и предсказаний . Для поставленной задачи числитель и знаменатель можно вычислить по следующим формулам

, (2)

, (3)

где - значения вероятностей принадлежности пикселей к тому или иному классу экспертной разметки и предсказаний соответственно. Зависимость значения коэффициента Серенсена от числа выполненных обучающих эпох (E) для разработанных нейронных сетей показана на рис. 6.



**Рис. 6.** Зависимость значения коэффициента Серенсена от числа выполненных обучающих эпох для разработанных алгоритмов глубокого обучения

Согласно результатам, приведенным на рис. 6, более худшие результаты сегментации спутниковых снимков были показаны алгоритмом LinkNet (0,72), тогда как лучшие результаты были получены с помощью U-Net (0,77). Некоторые тестовые результаты работы сети U-Net показаны на рис. 7.



**Рис. 7.** Тестовые результаты.

1. Заключение

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.

Список литературы

1. Я. Гудфеллоу, И. Бенджио, А. Курвилль, Глубокое обучение. М.: ДМК Пресс, 2017, 652 с.
2. Krizhevsky, A., Sutskever, I., Hinton G.: Imagenet classification with deep convolutional neural networks. In: Advances in neural information processing systems, pp. 1097–1105. Curran Associates Inc., USA (2012).
3. ImageNet Large Scale Visual Recognition Challenge 2012 (ILSVRC2012), http://image-net.org/challenges/LSVRC/2012/.
4. Seferbekov, S., Iglovikov, V., Buslaev, A., Shvets, A.: Feature Pyramid Network for Multi-Class Land Segmentation, https://arxiv.org/pdf/1806.03510.pdf.
5. DeepGlobe CVPR 2018 – Satellite Challenge, <http://deepglobe.org>.
6. Ronneberger O., Fischer P., Brox T.: U-Net: Convolutional Networks for Biomedical Image Segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) Medical Image Computing and Computer-Assisted Intervention – MICCAI-2015, LNCS, vol. 9351, pp. 234–341, Springer, Munich (2015).
7. Chhor, G., Bartolome Aramburu, C., Bougdal-Lambert, I.: Satellite Image Segmentation for Building Detection using U-net, http://cs229.stanford.edu/proj2017/final-reports/5243715.pdf.
8. Chaurasia, A., Culurciello, E.: LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation, https://arxiv.org/pdf/1707.03718.pdf.
9. The Cambridge-driving Labeled Video Database (CamVid), http://mi.eng.cam.ac.uk/research/projects/VideoRec/CamVid/.
10. SuperView-1 Satellite Sensor (0.5m), https://www.satimagingcorp.com/satellite-sensors/superview-1/.
11. GeoEye-1 Satellite Images, https://www.satimagingcorp.com/gallery/geoeye-1/.
12. Дж. Вандер Плас, Python для сложных задач: наука о данных и машинное обучение. СПб.: Питер, 2018. 576 с.
13. Khryashchev, V., Priorov A., Pavlov V., Ostrovskaya A.: Deep learning for region detection in high-resolution aerial images. In: Proceedings of 16-th IEEE East-West Design & Test Symposium (EWDTS’2016), pp. 792–796. Kazan, Russia (2016).
14. Khryashchev, V., Ivanovsky, L., Pavlov, V., Ostrovskaya, A., Rubtsov, A.: Comparison of Different Convolutional Neural Network Architectures for Satellite Image Segmentation. Proceedings of the 23rd Conference of Open Innovations Association FRUCT’23, pp. 172-179. Bologna, Italy (2018).
15. А. Джулли, С. Пал, Библиотека Keras – инструмент глубокого обучения. М.: ДМК Пресс, 2018, 294 с.
16. Н. Шакла, Машинное обучение & Tensorflow. СПб: Питер, 2019, 336 с.
17. Kingma, D., Ba, J.: Adam: A Method for Stochastic Optimization, https://arxiv.org/abs/1412.6980.