Детектирование зданий на спутниковых снимках высокого разрешения с помощью   
U-Net подобных нейронных сетей

Леонид Ивановский, Владимир Хрящев, Владимир Павлов,

ЯрГУ им. П.Г. Демидова

Ярославль, Россия

leon.ivanovsky@yahoo.com, vhr@yandex.ru, i@yajon.ru

Анна Островская

Российский университет дружбы народов (РУДН)

Москва, Россия,

ostrovskaya\_aa@rudn.university

***Аннотация — в данной статье представлены результаты исследования работы двух сверточных нейронных сетей для выделения зданий на спутниковых снимках базы данных PlanetScope. Для анализа качества работы алгоритмов, был использован коэффициент Серенсена, сопоставляющий друг с другом полученные алгоритмом результаты и области, ранее размеченными экспертами. Перед обучением моделей размеченные маски, как и соответствующие им снимки высокого разрешения, были порезаны на более мелкие части. Задача детектирования зданий на спутниковых снимках находит свое применение в области городского планирования и контроля строительства муниципальных объектов.***

# I. 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. Однако для проблемы сегментации спутниковых снимков, использование таких алгоритмов вместо традиционных алгоритмов машинного обучения нетривиально по ряду причин [1]. Такие методы должны:

* Учитывать малый размер некоторых объектов.

На больших спутниковых снимках, малые объекты обычно сгруппированы. Расстояние до поверхности Земли определяет то, что изображено в одном пикселе такого изображения. Это означает, что такие объекты, как здания, могут занимать всего несколько пикселей на спутниковом снимке.

* Быть инвариантным к поворотам объектов.

Объекты на спутниковых снимках могут быть расположены различным образом. Так, здания или транспортные средства могут быть повернутыми на любой угол, в то время как деревья чаще всего расположены вертикально к поверхности Земли.

* Иметь достаточно обучающих примеров.

Для большинства имеющихся баз изображений, таких как Inria, имеется недостаток размеченных снимков. В то же время, методы аугментации данных, таких как зеркальное отражение, поворот или аффинные преобразования, позволяют решить данную проблему.

* Иметь достаточно обучающих примеров.

Спутниковые снимки очень большие. Порой их размер превышает 150 Мб, а разрешение составляет более 16000 × 16000 пикселей. Например, каждое изображение со спутника DigitalGlobe [3] покрывает территорию более 64 км² и содержит как минимум 250 млн пикселей.

* Иметь достаточно обучающих примеров.

Для задачи детектирования объектов на снимках высокого разрешения собраны специальные базы данных. Они содержат спутниковые снимки, которые были сделаны в условиях ясной погоды. Однако, на некоторых снимках все равно присутствует шум (небольшая облачность неба или блики с крыш зданий). Очевидно, что методы глубокого машинного обучения должны быть устойчивыми к шуму.

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. 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. И хотя о сверточных нейронных сетях было известно несколько десятилетий тому назад, только благодаря недавним разработкам в области создания высокопроизводительных компьютеров с графическими ускорителями позволил исследователям работать со сверточными нейронными сетями, которые имеют миллионы параметров. На сегодняшний день, при решении современных задач компьютерного зрения подходы, основанные на использовании сверточных нейронных сетей, все чаще превосходят не только классические алгоритмы машинного обучения, но и в некоторых случаях даже экспертов в соответствующих областях [1].

Эта статья состоит из шести частей. Первая часть вводит читателей в суть поставленной проблемы, а также описывает преимущества использования сверточных нейронных сетей, обученных на графических процессорах в сравнении с традиционными методами машинного обучения. Вторая часть осуществляет обзор работ, связанных с задачей сегментации спутниковых изображений. Третий раздел посвящен доступным базам данных спутниковых снимков. В четвертом разделе описываются разработанные архитектуры сверточных нейронных сетей для выделения зданий на спутниковых снимках. Также в этой части были указаны особенности обучения моделей. В пятом разделе представлены результаты численных экспериментов для разработанных алгоритмов на базе данных PlanetScope. И наконец, в заключении подводятся итоги исследования.

# II. ОБЗОР РАБОТ

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 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. Такая особенность полносвязных нейронных сетей позволяет успешно их использовать в задаче сегментации изображений [4].

Вскоре полносвязные нейронные сети были преобразованы в пирамидальные сети. Подобного рода сети используют пирамидальную архитектуру для выделения более сложных признаков. Пирамидальные сети показали значительно лучшие результаты в сравнении с полносвязными сетями. Пирамидальные сети позволили получить значение индекса Жаккара равное 0.49 для сегментации изображений со спутника DeepGlobe [6].

Полносвязные нейронные сети были также усложнены до архитектуры U-Net. Впервые U-Net была упомянута в работе [7] по сегментации медицинских изображений. Позже эта модель оказалась также очень эффективной и для пиксельной классификации спутниковых снимков [8]. Архитектура U-Net использует пропускаемые соединения для объединения признаков, полученных на разных этапах работы сети. Такие соединения обеспечивают более точную локализацию объектов на спутниковых снимках. Используя архитектуру U-Net, авторы статьи [8] получили значение меры Серенсена равное 0,75 для детектирования зданий на спутниковых снимках.

В работе [9] 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. В некоторых случаях, такой подход оказывается лучше, чем использование полносвязных нейронных сетей. Нейронная сеть LinkNet позволила получить отличные результаты по детектированию объектов на базе CamVid [10].

# III. 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 Inria [3] содержит цветные спутниковые снимков, покрывающие территорию суммарной площадью 810 км². Обучающая и тестовая выборка составляет 180 изображений размером 1000x1000 пикселей при пространственном разрешении 0,3 м/пиксель. Все снимки разделены на 2 класса: «здания» и «не здания». Изображения из базы данных Inria охватывают разные городские поселения, начиная от густонаселенных районов и заканчивая альпийскими городами: Сан-Франциско (США), Чикаго (США), Вена (Австрия), Инсбук (Австрия), Беллингэм (США) и Тироль (Австрия). Примеры изображений из базы данных Inria показаны на рис. 1.

База данных SpaceNet [4] покрывает собой территорию 6 крупных городских агломераций: Рио-де-Жанейро (Бразилия), Лас-Вегаса (США), Парижа (Франция), Шанхая (Китай), Хартума (Судан) и Аталанты (США). 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. 2.

Набор данных DSTL содержит 50 спутниковых снимков в формате GEOTIFF размером 3300x3000 пикселей при пространственном разрешении 1 км x 1 км. Впервые эта база данных была представлена в конкурсе Kaggle «DSTL Satellite Imagery Feature Detection» [15]. Изображения этой базы данных разбиты на 10 различных классов: «здания», «искусственные сооружения», «дороги», «трассы», «деревья», «сельскохозяйственные угодья», «водные пути», «стоячая вода», «большие транспортные средства» (например грузовые автомобили или автобусы) и «небольшие транспортные средства» (автомобили, микроавтобусы или велосипеды). Примеры изображений из базы данных DSTL показаны на рис. 4.

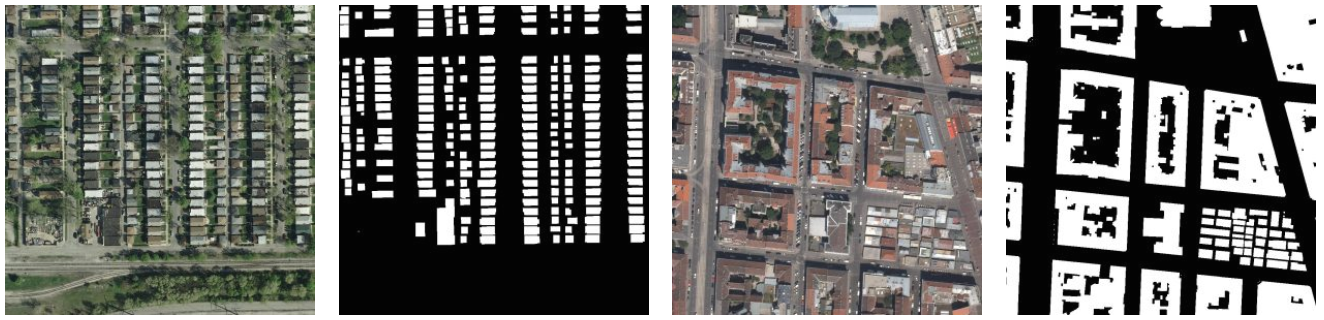


Рис. 1. Примеры изображений из базы данных Inria



Рис. 2. Примеры изображений из базы данных SpaceNet

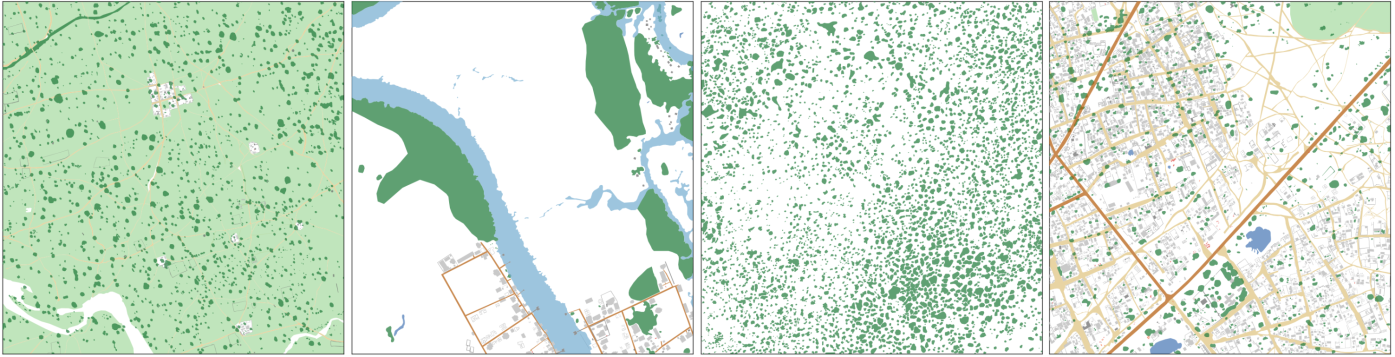


Рис. 3. Примеры изображений из базы данных DSTL



Рис. 4. Примеры изображений из базы данных PlanetScope

Для исследования работы алгоритмов выделения зданий были использованы спутниковые снимки из базы данных PlanetScope, предоставленной Агентством космических систем России. 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. 4.

# IV. АЛГОРИТМЫ ГЛУБОКОГО ОБУЧЕНИЯ

В данной статье представлены модели, основанные на использовании сверточной нейронной сети. Это специальная архитектура нейронной сети, направленная на быстрое и качественное обнаружение различных объектов [1]. Сверточные нейронные сети относятся к алгоритмам глубокого обучения, популярным сейчас для решения большинства современных задач компьютерного зрения. В ходе исследования были использованы две модели нейронных сетей: U-Net [6] и, LinkNet [13]. Данная статья является продолжением работы [13], посвященной обнаружению крупных объектов на спутниковых снимках.

Для построения архитектур сетей использовалась библиотека 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, таких как слои, функции оптимизации и активации, а также готовых инструментов для предварительной обработки изображений и текстовых данных [14]. Кроме того, эта библиотека позволяет обучать модели на GPU.

Как показано на рис.5, 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.

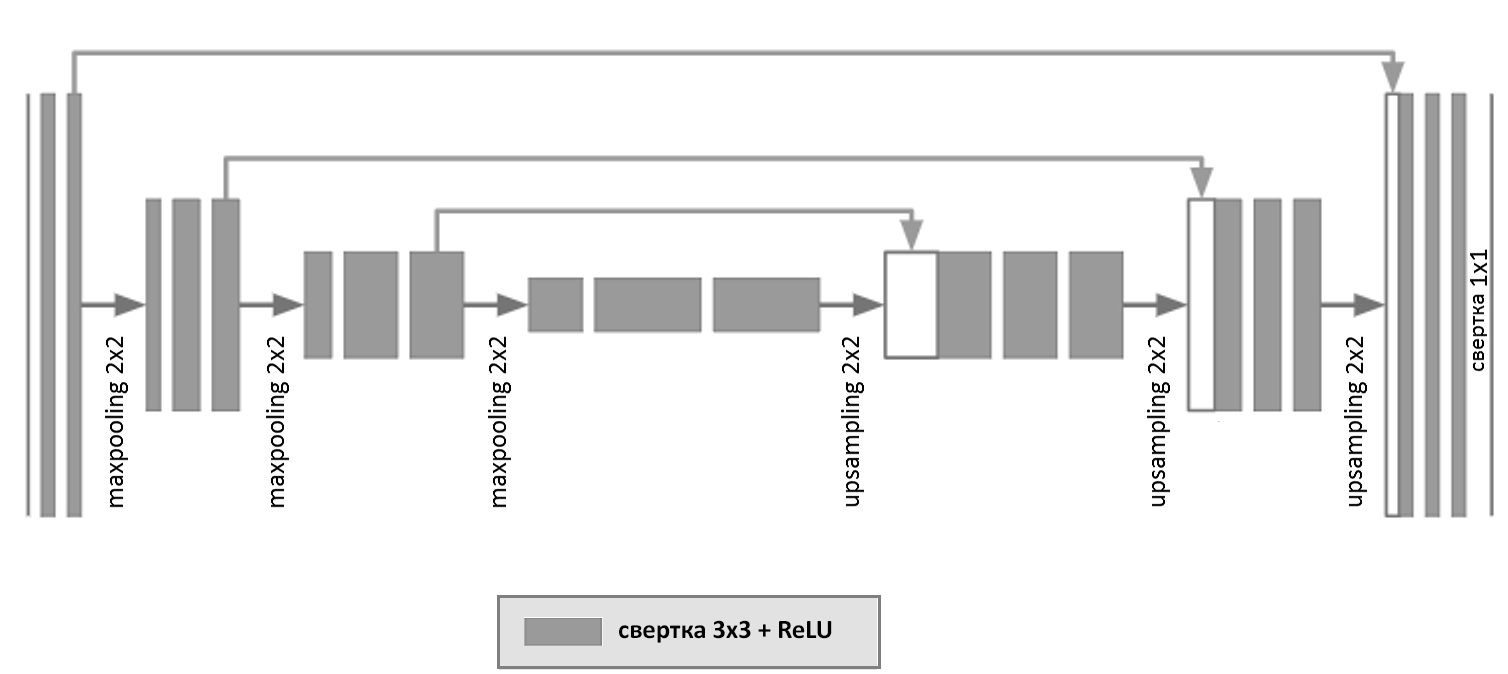
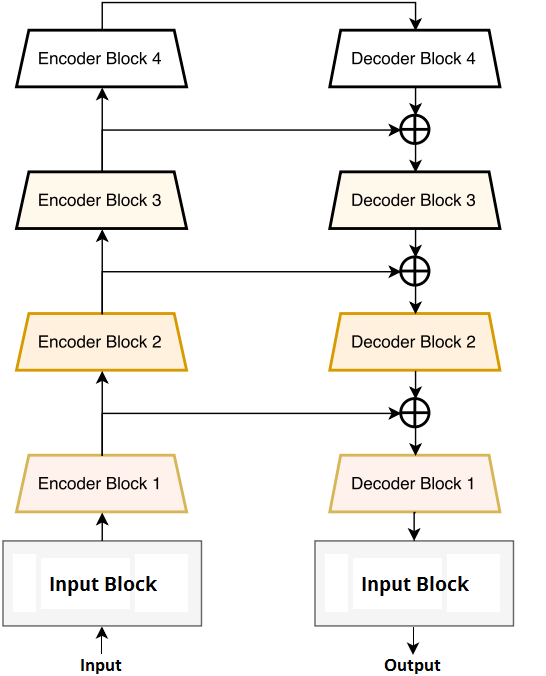
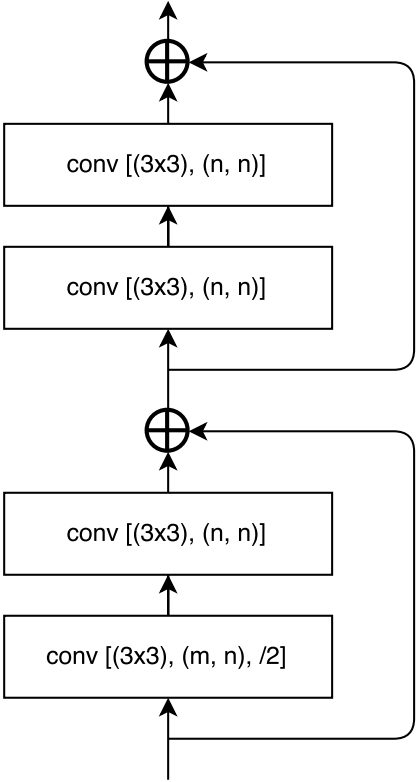


Рис. 5. Архитектура сети U-Net

Как и описанная выше архитектура сети U-Net, LinkNet имеет две части: кодировщик и декодировщик. Согласно архитектуре сети, указанной на рис. 6, 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. Архитектура LinkNet и схема кодировщика этой сети показаны на рис. 6.



a)



б)

Fig. 6. LinkNet: а) архитектура сети, б) схема кодировщика

Согласно табл. 1., 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 [15].

Таблица 1. Количество параметров обучения в разработанных моделях

|  |  |
| --- | --- |
| **Алгоритм** | **Приблизительное количество параметров обучения (в млн)** |
| U-Net | 7.8 |
| LinkNet | 17,2 |

В качестве алгоритма численной оптимизации был выбран оптимизатор адаптивной оценки моментов (Adam) [16] со скоростью обучения 1e-3. В качестве функции потерь была выбрана бинарная кросс-энтропия [17]. 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.



a)



b)

Рис. 7. Данные из базы PlanetSScope: a) снимок из базы, b) сгенерированная маска

# V. РЕЗУЛЬТАТЫ ЧИСЛЕННЫХ ЭКСПЕРИМЕНТОВ

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. 7.

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. Примеры нарезанных спутниковых снимков и соответствующих им масок показаны на рис. 8.

В результате подобных преобразований, подготовленные выборки данных содержали 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 включала в себя 2611 фотографий, а валидационная – 653. Тренировочное и тестовое множества данных не имели общих картинок. Для поставленной задачи брались 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.

Доля правильных ответов (A) моделей рассчитывалась по формуле

где - количество правильно распознанных пикселей, а – общее число пикселей на картинке. Результаты численных экспериментов для каждой разработанной модели на валидационной выборке показаны в табл. 2.

Таблица 2. Результаты численных экспериментов

|  |  |
| --- | --- |
| **Алгоритм** | **Доля правильных ответов (A, %)** |
| U-Net | 96,31% |
| LinkNet | 95,85% |

Качество работы алгоритмов сегментации обычно оценивается специальными метриками. Для оценки разработанных моделей использовался коэффициент подобия Серенсена (DSC), осуществляющий сравнение предсказанных и экспертных масок изображений между собой. 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, а - сумма мощностей экспертной разметки и предсказаний [19]. Для поставленной задачи числитель и знаменатель можно вычислить по следующим формулам.

, ,

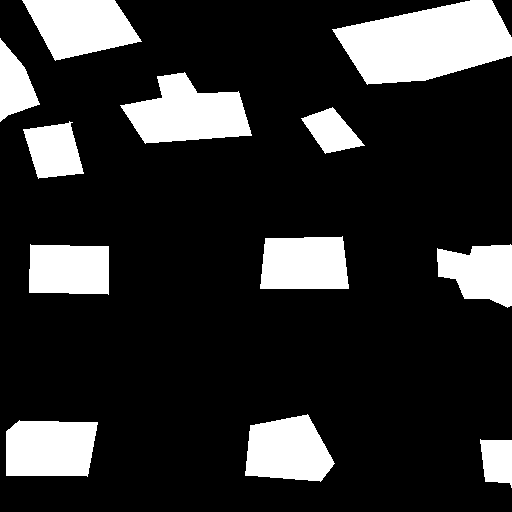
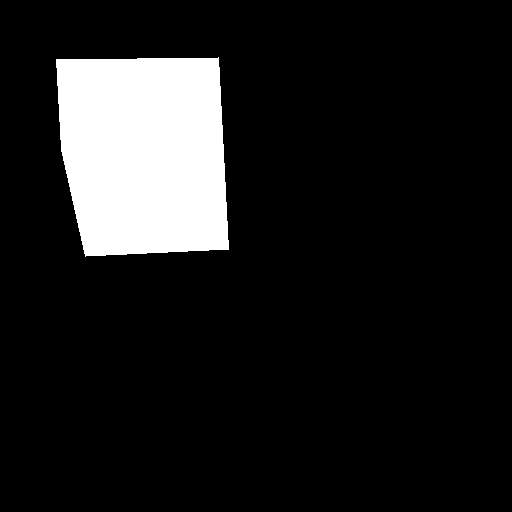
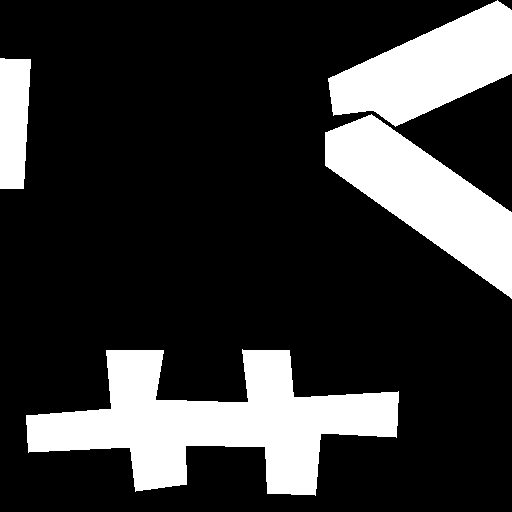
   

Рис. 8. Нарезанные образцы: в первой строке – подготовленные картинки, вторая строка – соответствующие им маски.

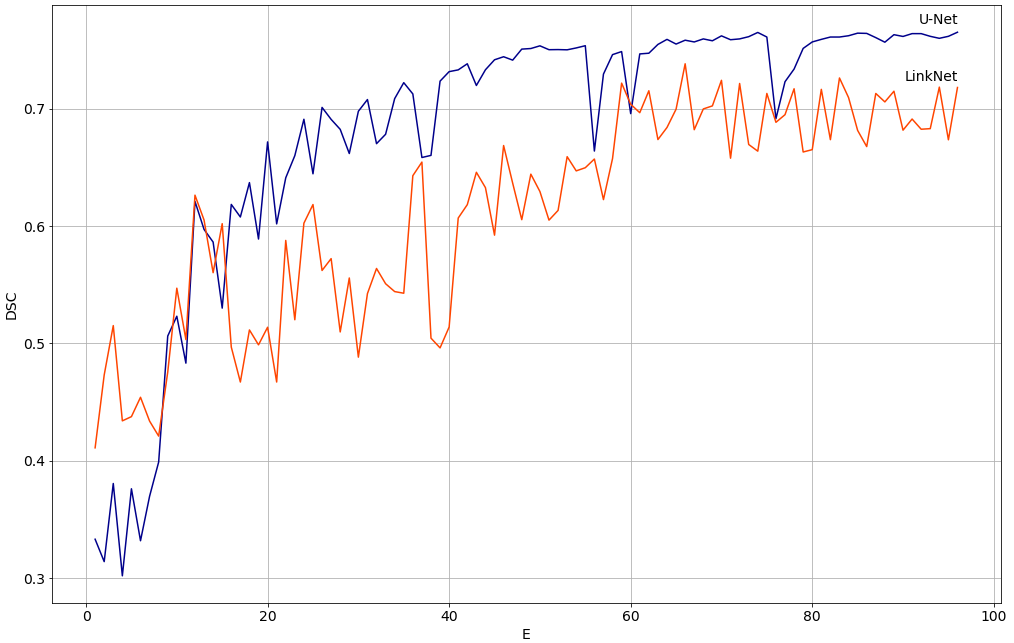


Рис.9. Зависимость значения коэффициента Серенсена от числа выполненных обучающих эпох для разработанных нейронных сетей

где - значения вероятностей принадлежности пикселей к тому или иному классу экспертной разметки и предсказаний соответственно. Зависимость значения коэффициента Серенсена от числа выполненных обучающих эпох для разработанных нейронных сетей показана на рис. 9.

Таблица 3. Результаты численных экспериментов

|  |  |
| --- | --- |
| **Алгоритм** | **Коэффициент Серенсена (DSC)** |
| U-Net | 0,77 |
| LinkNet | 0,72 |

Согласно результатам, приведенным на рис. 8 и в табл. 3, более худшие результаты сегментации спутниковых снимков были показаны алгоритмом LinkNet, тогда как лучшие результаты были получены с помощью U-Net. Этот факт объясняется сложностью архитектур разработанных моделей. Однако нужно отметить, то обучение сети LinkNet занимало всего 1.5 ч, что на 1 ч меньше, чем обучение U-Net, несмотря на то, что LinkNet имеет больше параметров для обучения. Обе архитектуры пробрасывают информацию из кодировщика в декодировщик, что позволяет сетям использовать больше полезной информации из входных данных.

# VI. ЗАКЛЮЧЕНИЕ

В статье было показано, как глубокие нейронные сети, реализованные на современных графических процессорах, могут быть эффективно использованы для решения задачи выделения зданий на спутниковых снимках высокого разрешения. Численные эксперименты по оценке качества работы разработанных алгоритмов были проведены на изображениях из базы данных PlanetScope. Для выполнения численных экспериментов каждый спутниковый 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. Для обучения и тестирования сверточных нейронных сетей был использован суперкомьютер NVIDIA DGX-1, предоставленный Центром искусственного интеллекта и цифровой экономики ЯрГУ им. П.Г. Демидова.

# БЛАГОДАРНОСТИ

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.

# ЛИТЕРАТУРА

[1] Я. Гудфеллоу, И. Бенджио, А. Курвилль, *Глубокое обучение*. М.: ДМК Пресс, 2017, 652 с.

[2] Inria Aerial Image Labelling Database, Web: https://project.inria.fr/aerialimagelabeling/.

[3] DigitalGlobe satellite, Web: https://www.digitalglobe.com.

[4] E. Shelhamer, J. Long, T. Darrell, “Fully Convolutional Networks for Semantic Segmentation”, Web: https://arxiv.org/pdf/1605.06211.pdf.

[5] S. Seferbekov, V. Iglovikov, A. Buslaev, A. Shvets, Web: https://arxiv.org/pdf/1806.03510.pdf.

[6] DeepGlobe. CVPR 2018 – Satellite Challenge, Web: http://deepglobe.org.

[7] 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.

[8] 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.

[9] A. Chaurasia, E. Culurciello, “LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation”, Web: https://arxiv.org/pdf/1707.03718.pdf.

[10] The Cambridge-driving Labeled Video Database (CamVid), Web: http://mi.eng.cam.ac.uk/research/projects/VideoRec/CamVid/.

[11] SpaceNet Database, Web: http://explore.digitalglobe.com/spacenet.

[12] DSTL Satellite Imagery Feature Detection, Web: https://www.kaggle.com/c/dstl-satellite-imagery-feature-detection.

[13] Sovzond, Web: https://sovzond.ru/services/remote-sensing/.

[14] V. Khryashchev, L. Ivanovsky, V. Pavlov, A. Ostrovskaya and A. Rubtsov, "Comparison of Different Convolutional Neural Network Architectures for Satellite Image Segmentation," *2018 23rd Conference of Open Innovations Association (FRUCT), Bologna, 2018*, pp. 172-179.

[15] А. Джулли, С. Пал, *Библиотека Keras – инструмент глубокого обучения*. М.: ДМК Пресс, 2018, 294 с.

[16] Дж. Сандерс, Э. Кэндрот, Технология CUDA в примерах: введение в программирование графических процессоров. М.: ДМК Пресс, 2013, 232 с.

[17] 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.

[18] T.S.K.M. Rabie, “Implementation of some similarity coefficients in conjunction with multiple upgma and neighbor-joining algorithms for enhancing phylogenetic trees”, *Egypt. Poult. Sci*. Vol. 30 (II), pp.607-621.