Сравнение различных архитектур сверточных нейронных

сетей для сегментации спутниковых снимков

Владимир Хрящев, Леонид Ивановский, Владимир Павлов

ЯрГУ им. П.Г. Демидова, Ярославль, Россия

v.khryashchev@uniyar.ac.ru, leon.ivanovsky@yahoo.com, vladimir@1pavlov.com

Анна Островская

РУДН, Москва, Россия

ostrovskaya\_aa@rudn.university

Антон Рубцов

АО «РКС», Москва, Россия

rubtsov493@gmail.com

*Аннотация*— в данной статье представлены результаты исследования работы трех различных архитектур сверточных нейронных сетей для обнаружения объектов на спутниковых снимках из баз данных DSTL, Landsat-8 и PlanetScope. Для анализа точности работы алгоритмов, полученные результаты были сопоставлены с областями, ранее размеченными экспертами. Сегментация спутниковых снимков находит свое применение в области градостроительства, для моделирования климата и т.д.

I. ВВЕДЕНИЕ

Прогресс в области создания высокопроизводительных компьютеров с графическими ускорителями (GPU) позволил исследователям работать со сверточными нейронными сетями (CNNs), которые имеют миллионы параметров. При решении современных задач компьютерного зрения подходы, основанные на использовании сверточных нейронных сетей, превосходят не только классические методы, но и в некоторых случаях даже экспертов в соответствующих областях. Сверточные нейронные сети продемонстрировали свое превосходство при решении задач классификации изображений, обнаружении объектов и распознавания сцены. В настоящее время почти все новые задачи компьютерного зрения пытаются решить именно с помощью методов глубокого обучения [1]. Сегментация изображений - одна из таких задач.

Сегментация спутниковых снимков является сложной задачей. В приложениях машинного обучения поставленная проблема обычно переформулируется как задача маркировки пикселей. В настоящее время задача обнаружения объектов на спутниковых снимках с высоким разрешением находится в центре внимания исследовательского общества. Между тем, большинство подходов к решению этой проблемы заключается в использовании сверточной нейронной сети. Особенность таких сетей заключается в том, что дескрипторы изображений формируются алгоритмом самостоятельно, в процессе обучения.

Сегментация спутниковых снимков по пикселям требует применение алгоритмов глубокого обучения. При этом использование таких методов вместо традиционных решений является нетривиальным по целому ряду причин. Необходимы уникальные подходы по решению задачи пространственной протяженности детектируемых объектов с учетом инвариантности к вращению картинки или изменению масштаба снимка [2]. Такие алгоритмы должны:

* Учитывать малую пространственную протяженность объектов. Обнаружение небольших объектов на больших снимках является одной из основных проблем. В отличие от объектов из базы данных ImageNet,, на картинках которой запечатлены большие объекты, на спутниковых снимках интересующие объекты часто очень маленькие по размерам, но при этом они плотно сгруппированны. Это объясняется тем, что для спутниковых снимков разрешение определяется, исходя из расстояния до земли. Оно определяет то, что запечатлено в одном пикселе изображения. Размер территории, запечатленной на снимках, обычно варьируется от 30 см² до 16 м². Это означает, что, размер такого объекта, как автомобиль, будет занимать всего 15 пикселей.
* Быть инвариантными к вращению. Объекты на спутниковых снимках могут быть ориентированы в пространстве совершенно по-разному. Так, например, корабли могут быть повернуты на любой угол, от 0 до 360 градусов, тогда как деревья расположены вертикально.
* Иметь достаточное количество обучающих примеров. Для большинства наборов имеющихся данных, например, таких как LandSat [3] и Inria [4], наблюдается нехватка размеченных снимков. Однако усилия по созданию большого количества обучающих примеров, как, например, в базе данных SpaceNet [5], позволяют решить эту проблему.
* Иметь возможность обрабатывать картинки высокого разрешения. Спутниковые снимки, поступающие на вход алгоритмам машинного обучения, очень велики, их размер часто превышает разрешение 2000x2000 пикселей. Так, например, снимки со спутника DigitalGlobe охватывают более 64 км², что составляет более 250 миллионов пикселей.

Сегментация изображения по существу включает разделение изображений на значимые области, которые можно рассматривать как задачу классификации уровня пикселей. Наиболее простой (и медленный) подход к решению этой проблемы - ручная сегментация изображений. Тем не менее, это трудоемкий и долгий процесс, который подвержен ошибкам и несоответствиям, неизбежным при использовании ручного труда. В настоящее время большой интерес исследователей в области машинного обучения связан с разработкой систем автоматической сегментации снимков. Подобного рода сегментация позволяет обеспечить обработку изображений «на лету» сразу его после получения. Однако, для автоматических систем, такой процесс требует предоставления необходимой точности, чтобы быть полезным в производственной среде.

Эта статья состоит из седьмых частей. Первая часть вводит читателей в суть поставленной задачи. Вторая часть посвящена обзору работ, связанных с проблемой сегментации спутниковых изображений. Третья часть посвящена доступным базам данных спутниковых снимков. В четвертом разделе описываются разработанные архитектуры сверточных нейронных сетей для сегментации изображений. Также в этой части были указаны особенности обучения моделей. В пятом и шестом разделах представлены результаты численных экспериментов для разработанных алгоритмов на разных базах данных. И наконец, в заключении подводятся итоги исследования, а также делаются предположения о возможном улучшении классификаторов.

II. ОБЗОР РАБОТ

В последние годы были предложены различные архитектуры сверточных нейронных сетей, которые могут осуществлять сегментацию спутниковых снимков. Один из самых успешных алгоритмов сегментации основан на полностью сверточных сетях (FCN). Основная идея этого подхода заключается в использовании сверточной нейронной сети для извлечения необходимых признаков с заменой полностью связанного слоя слоем свертки с выводом результатов классификации [6]. Такой подход позволяет обучать сверточные нейронные сети для сегментации изображений разных размеров.

Следуя этим соображениям, авторы [7] в своей работе представляют архитектуру сверточной нейронной сети подготовленные веса на наборе данных COCO. Этот алгоритм состоит из двух сетей: сети предложений регионов (RPN) и FCN. Первая модель принимает на вход входное изображение и выводит преобразованную картинку с предположениями о местоположении объектов на фотографиях. Вторая модель использует эту информацию и выполняет сегментацию преобразованного первым алгоритмом изображения. Применение такого метода позволило точно определить здания на фотографиях спутников Inria Aerial Images Dataset [4]. Наилучшая точность алгоритма составила 92,49%.

Вскоре подход по использованию FCN был дополнен, и стал известен как U-Net. В статье [8] представлена архитектура сверточной нейронной сети U-Net – это специфический тип FCN, который получил свое широкое распространение в задачах сегментации медицинских изображений [8, 9]. Позже эта модель оказалась также очень эффективной и для пиксельной классификации спутниковых снимков [10]. Архитектура U-Net использует skip-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 [10] получают значение меры Серенсена равным 0,75.

Авторы статьи [11] придерживаются аналогичного метода для решения задачи сегментации спутниковых изображений. Они разработали архитектуру U-Net, использующую значения весов, посчитанные для сети ResNet-34. Построенный алгоритм показывает отличные результаты обнаружения дорог на спутниковых снимках из базы данных DeepGlobe [12].

Классические алгоритмы глубокого машинного обучения на основе сверточных нейронных сетей, такие как DeconvNet также очень популярны для решения подобных задач [13]. Такого рода модели обычно инициализируются значениями весов, предварительно посчитанными для других сетей (например, ImageNet или ResNet) на большом наборе данных. Такой подход позволяет значительно повысить эффективность учебного процесса для некоторых практических приложений, таких как обнаружение рака или безопасность движения, где точность моделей играет решающую роль.

Данная статья является продолжением работы [14], посвященной обнаружению объектов и их сегментации на спутниковых снимках.

III. БАЗЫ СПУТНИКОВЫХ СНИМКОВ

Базы данных являются важной частью обучения сверточной нейронной сети, оценки эффективности модели, а также сравнительного анализа различных алгоритмов машинного обучения. На данный момент в открытом доступе есть немало баз данных спутниковых снимков.

База данных DeepSat [15] содержит два набора аннотированных снимков: 500000 изображений со спутника Sat-4, разделенных на 4 класса («земля», «деревья», «поля» и класс, который состоит объектов, отличных от указанных трех) и 405000 изображений со спутника Sat-6 , разделенных на 6 классов («земля», «деревья», «поля», «дороги», «здания» и «водные объекты»). Все образцы базы имеют размер 28 × 28 пикселей при пространственном разрешении 1 м/пиксель и содержат 4 цветовых канала (красный, зеленый, синий и инфракрасное излучение). Хотя этот набор данных полезен для предварительной подготовки моделей, решающих более сложные задачи (например, сегментации изображений), он не позволяет предпринять дальнейшие шаги для детального анализа детектированных объектов. Примеры изображений из базы данных DeepSat показаны на рис. 1.

База данных Inria [4] содержит цветные спутниковые снимки, покрывающие территорию 10 городов 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. Этот набор данных отлично подходит для оценки обобщающей способности методов сегментации изображений. Примеры изображений из базы данных Inria показаны на рис. 2.

База данных DeepGlobe [12] содержит цветные снимки размером 1024x1024 пикселей со спутника DigitalGlobe. В обучающем наборе данных каждый образец содержит также маску с соответствующей дорожной разметкой в полутоновом формате. При этом в большинстве случаев небольшие дороги внутри сельской местности не были размечены. Обучающая и тестовая выборка этого набора данных содержит 6246 и 1243 фотографий соответственно. Примеры изображений из базы DeepGlobe показаны на рис. 3.

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.

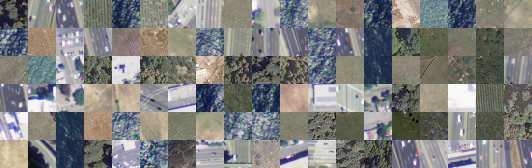


Рис. 1. Примеры изображений из базы данных DeepSat

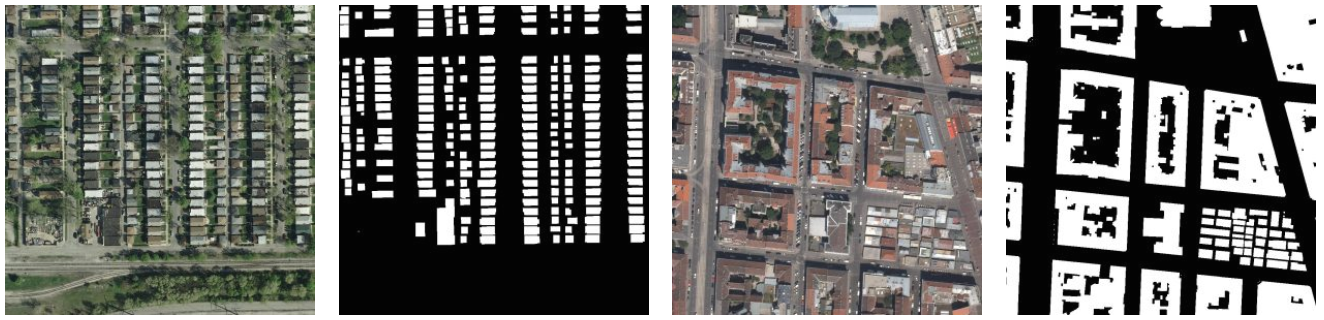


Рис. 2. Примеры изображений из базы данных Inria



Рис. 3. Примеры изображений из базы данных DeepGlobe

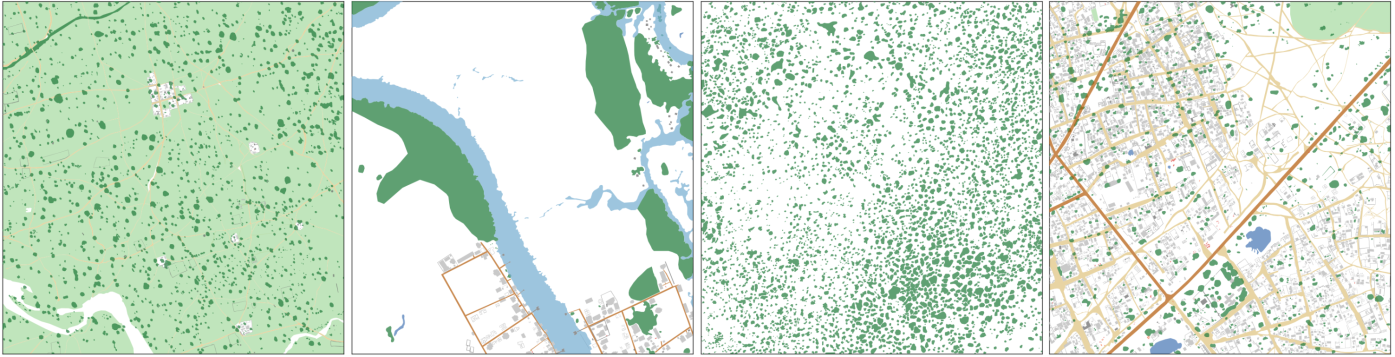


Рис. 4. Примеры изображений из базы данных DSTL

Помимо снимков из базы DSTL, для исследования работы алгоритмов обнаружения объектов также были использованы изображения со спутника Landsat-8 [20] и данные PlanetScope, предоставленные Агентством космических систем России. Изображения Landsat-8 имеют пространственное разрешение 30 м/пиксель. Это самое высокое разрешение среди спутниковых снимков из баз данных открытого доступа. Снимки же из PlanetScope схожи с изображениями со спутника Landsat-8, но при этом имеют более качественное разрешение: 3 м/пиксель.

IV. СВЕРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ

В данной статье представлены модели, основанные на использовании сверточной нейронной сети - специальной архитектуры нейронной сети, направленной на быстрое и качественное обнаружение различных объектов [1]. Сверточные нейронные сети тесно взаимосвязаны с алгоритмами глубокого машинного обучения, которые на сегодняшний день очень популярны для решения большинства задач компьютерного зрения.

Для проведения сравнительного анализа различных алгоритмов сегментации спутниковых снимков были созданы три модели нейронных сетей, основанных на архитектурах UNet [18], SegNet [19] и LinkNet [20]. Изучение работы этих моделей продолжает исследование, которое было представлено в статье [14]. Для построения архитектур сетей использовалась библиотека Keras с использованием фреймворка Tensorflow в качестве бэкэнда. Keras - это библиотека с открытым исходным кодом, написанная с использованием языка программирования Python. Она содержит множество реализаций широко используемых структурных блоков нейронной сети, а также готовых инструментов для предварительной обработки изображений и текстовых данных. Keras предлагает набор инструментов высокого уровня для разработки моделей глубокого обучения, что позволяет значительно упростить процесс разработки сверточных нейронных сетей [21]. Кроме того, эта библиотека позволяет обучать модели на GPU.

Как показано на рис.5, UNet 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 в случае с сетью UNet, SegNet состоит из кодировщика, декодировщика и классифицирующего слоя. Архитектура этой модели показана на рис. 6. UNet состоит из 27 сверточных слоев, 26 операций пакетной нормализации, 26 функций активации ReLU, а также 10 функций повышения и понижения размерности. Все сверточные слои кодировщика соответствуют первым сверточным слоям сети VGG16 для классификации объектов. Перед обучением, SegNet инициализировалась весами этой сети. Каждому блоку кодировщика соответствовал аналогичный блок декодировщика. Таким образом, декодировщик состоял из тех же слоев, что и кодировщик, за исключением применения оператора maxpooling, который были заменен оператором upsampling. Классифицирующим слоем сети SegNet выступал слой с функцией softmax, который позволил независимо предсказывать вероятности принадлежности каждого пикселя к тому или иному классу.

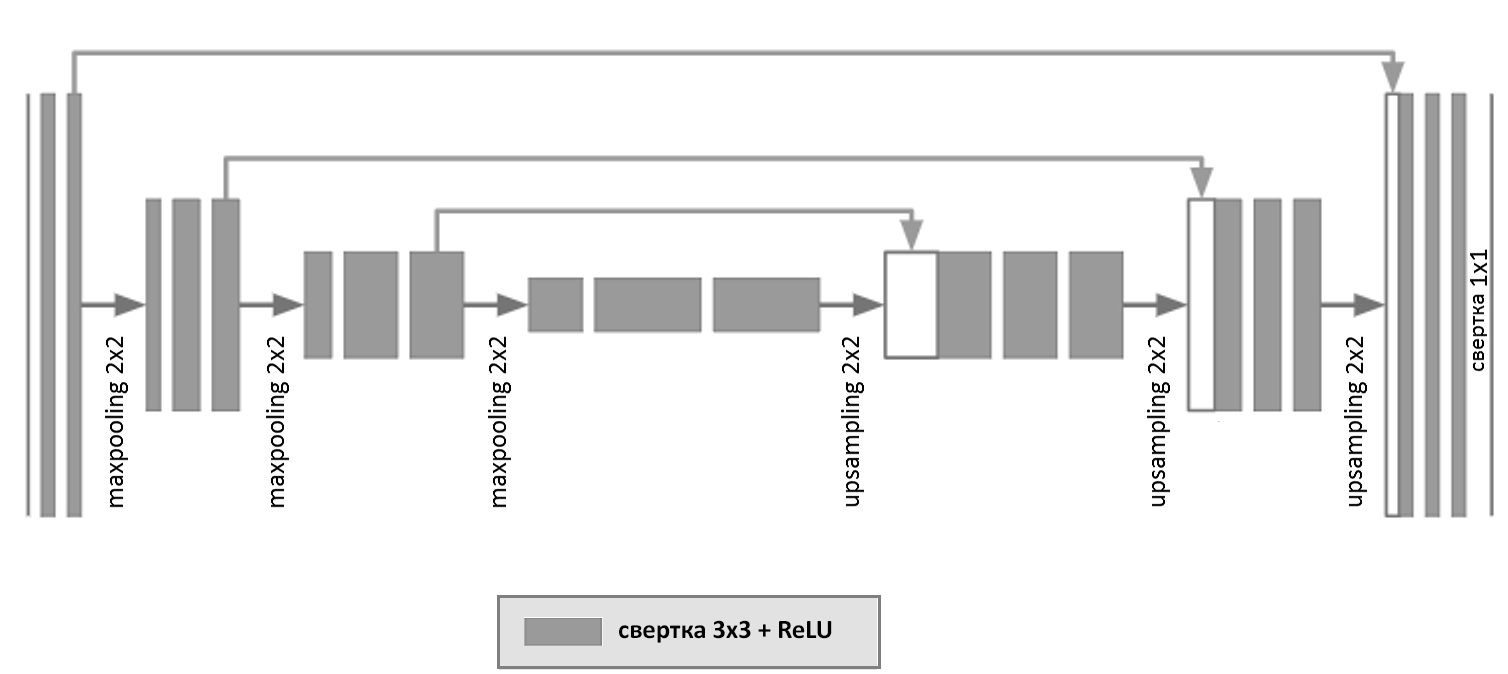


Рис.. 5. Архитектура сети UNet

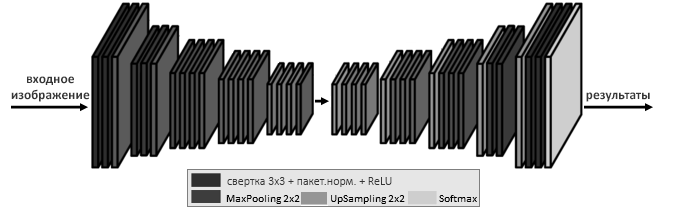


Рис. 6. Архитектура сети SegNet

Также была разработана сверточная нейронная сеть TLinkNet с LinkNet-подобной архитектурой на основе модели из [19]. Разница между созданной TLinkNet и сетью из [19] заключается в отсутствии одного блока у кодировщика и декодировщика. Этот факт объясняется разницей в размере изображений, поступающих на вход соответствующим сетям. Как и описанные выше алгоритмы, TLinkNet имеет две части: кодировщик и декодировщик. Обе подсети состоят из 3 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 применения оператора maxpooling, который был заменен на операцию повышения частоты дискретизации с помощью оператора upsampling. Архитектура сети TLinkNet показана на рис. 7.

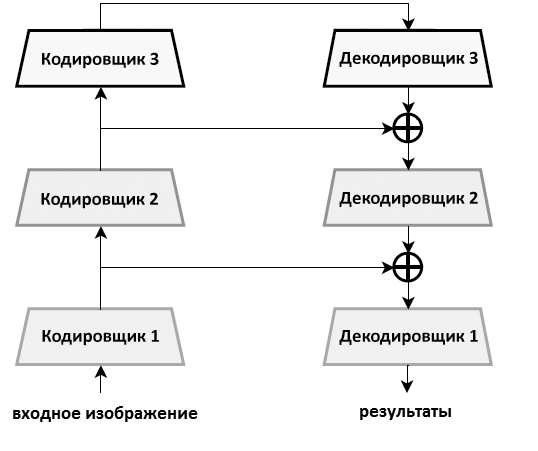


Рис. 7. Архитектура сети TLinkNet

Подход, основанный на использовании сверточных нейронных сетей, требует немалых вычислительных ресурсов. Для ускорения работы нейронных сетей, 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 [21].

Скорость обучения была равной . В качестве алгоритма численной оптимизации был выбран оптимизатор адаптивной оценки моментов (Adam). Эта функция использует значения моментов градиентов для поддержания скорости обучения, что повышает точность работы алгоритма машинного обучения [22]. В качестве функции потерь была выбрана бинарная кросс-энтропия [23]. На каждой итерации обучения модель обновляла свои веса после прогона через сеть 64 образцов из обучающей выборки. Классификатор заканчивал свое обучение после выполнения 256 эпох.

V. ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ НА БАЗЕ DSTL

Для осуществления численных экспериментов из исходного набора данных базы DSTL были извлечены изображения меньшего размера. Такой подход позволяет получить большее число картинок. Соответствующие эти фрагментам маски были получены из общего файла формата csv. В результате подобных преобразований обучающая выборка содержала 3955 фотографий, а тестовая - 600 снимков. Образцы размером 160 × 160 пикселей не содержали одинаковых изображений в сформированных выборках. Для поставленной задачи учитывались только 3 класса: «деревья», «сельскохозяйственные угодья» и «вода».

Запуск алгоритмов проводился на суперкомпьютере NVIDIA DGX-1. Процесс обучения длился около 1 ч. Доля правильных ответов (А) алгоритма машинного обучения рассчитывалась по формуле:

где - количество правильно классифицированных изображений, а – размер тестовой выборки [24]. Результаты численных экспериментов на образцах из тестовой выборки показаны в табл. I.

ТАБЛИЦА 1. Результаты численных экспериментов

|  |  |
| --- | --- |
| **Алгоритм** | **Доля правильных ответов (A)** |
| SegNet | 93,59% |
| TLinkNet | 94.53% |
| U-Net | 94.66% |

Согласно полученным результатам, все алгоритмы продемонстрировали высокую точность. Этот факт можно объяснить небольшим размером объектов рассматриваемых классов. Таким образом, использование одной этой меры недостаточно для измерения эффективности работы сетей.

Как правило, качество работы алгоритмов сегментации оценивается другими специальными метриками, сравнивающими предсказанные и экспертные маски изображений. Для оценки разработанных моделей использовался коэффициент подобия Серенсена (DSC). Этот показатель принимает значения из отрезка [0, 1] и фактически является двоичной мерой сходства двух множеств. Коэффициент Серенсена рассчитывается по следующей формуле:

,

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

, ,

где - значения вероятностей принадлежности пикселей к тому или иному классу экспертной разметки и предсказаний соответственно. Графики зависимости значения DSC от количества проделанных обучающих эпох (E) показаны на рис. 8.

Согласно результатам тестирования, представленным в табл. 2, худшие результаты сегментации спутниковых снимков показал алгоритм SegNet, тогда как лучшие результаты были получены с помощью UNet. Этот факт можно объяснить сложностью структуры сетей. Архитектуры TLinkNet и UNet осуществляют передачу признаков от кодировщика к декодировщика в отличие от SegNet. Эта особенность позволяет использовать дополнительную полезную информацию из поступающих на вход данных.

ТАБЛИЦА 2. Результаты численных экспериментов

|  |  |
| --- | --- |
| **Алгоритм** | **Коэффициент Серенсена (DSC)** |
| SegNet | 0.45 |
| TLinkNet | 0.68 |
| U-Net | 0.75 |

VI. ЧИСЛЕННЫЕ РЕЗУЛЬТАТЫ НА БАЗАХ LANDSAT И PLANETSCOPE

Перед исследованием работы модели U-Net на другом наборе данных, он был вручную размечен 3 независимыми экспертами в веб-приложении «Supervise» [26]. Новая база изображений содержала 30 спутниковых снимков. Каждое изображение из обучающей выборки представляет собой фрагмент спутникового снимка шириной 26 км (8600 пикселей) и высотой 17,5 км (5800 пикселей).

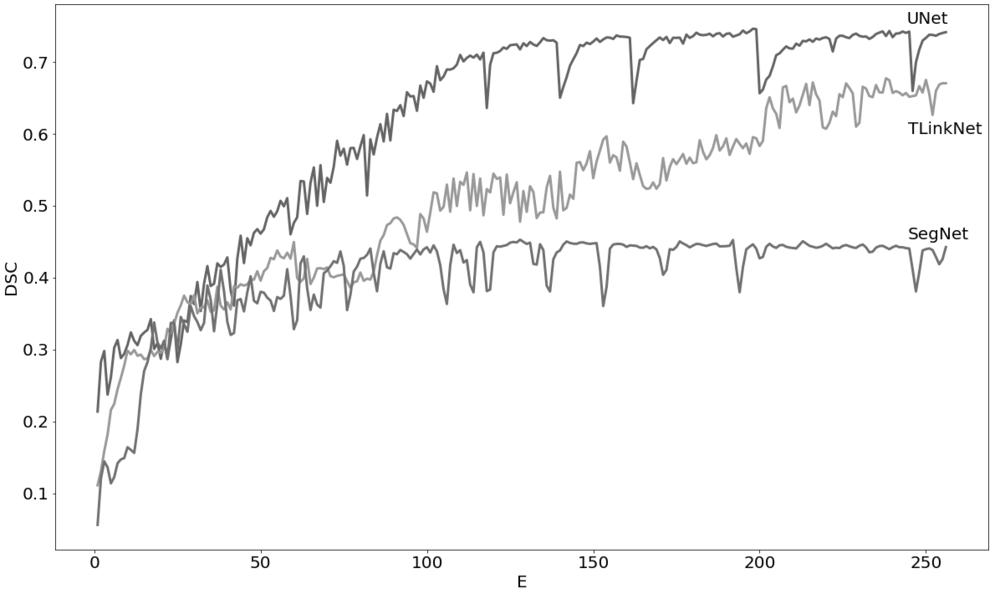


Рис..8. Зависимость значения коэффициента Серенсена от количества обучающих эпох для изображений из тестовой выборки

По ходу выполнения численных экспериментов возникла проблема низкой точности обнаружения объектов класса «лес» на спутниковых снимках базы данных PlanetScope. Это было связано с увеличением деталей кроны деревьев и более четкой границей лесов. Средняя точность обнаружения детектором объектов класса «лес» составила всего 73,84%. Чтобы улучшить эту ситуацию, алгоритм машинного обучения был переобучен на новом наборе снимков, который содержал изображения из баз данных Landsat-8 и PlanetScope, перемешанных между собой и разбитых для обучения на группы из 300 изображений размером 224x224 пикселей. Для чистоты эксперимента структура сети не изменялась, а обучение проводилось в 60 эпох. Подход по перемешиванию изображений разных баз позволил создать универсальный детектор, который отлично работает на наборах данных Landsat-8 и PlanetScope. Тестовые результаты работы сети U-Net для 3 классов объектов представлены в таблице 3.

ТАБЛИЦА 3. Тестовые результаты работы сети U-Net

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Клас** | **Landsat-8** | | **PlanetScope** | |
| **Точность** | **«Средний» процент пересечений** | **Точность** | **«Средний» процент пересечений** |
| Forest | 89.03% | 92.54% | 92.14% | 93.21% |
| Water | 90.87% | 81.64% | 90.98% | 82.14% |
| Agriculture | 94.35% | 96.32% | 96.52% | 96.88% |

В результате объекты класса «лес» имели на спутниковых снимках базы PlanetScope точность детектирования 92,14%, что на 18,3% больше, чем у алгоритма, который был обучен в первый раз. Для объектов других классов такой подход также позволил повысить точность детектирования из-за более четких границ объектов и снижения цены ошибки за счет уменьшения площади заснятой территории, которую покрывали изображения, с 30 м до 3 м. Примеры работы алгоритма обнаружения объектов на спутниковых снимках из баз данных Landsat-8 и PlanetScope показаны на рис. 9.

VII. ЗАКЛЮЧЕНИЕ

Результаты численных экспериментов по оценке эффективности разработанных алгоритмов, выполненных для спутниковых снимков из базы данных DSTL, показывают, что использование более сложных нейронных сетей позволяет повысить качество сегментации. Несмотря на высокие значения точности (A) для каждой модели, именно коэффициент Серенсена (DSC), показывает разницу работы различных алгоритмов. Наибольшее значение DSC равное 0,75 удалось получить с помощью сети U-Net.

Для исследования работы модели U-Net на другом наборе данных он был вручную размечен и переподготовлен. Разметка спутниковых снимков была проведена 3 независимыми экспертами в веб-приложении «Supervise». Средняя точность обнаружения детектором объектов класса «лес» на спутниковых снимках PlanetScope составила всего 73,84%. Для разрешения возникшей ситуации алгоритм машинного обучения был переобучен на новом наборе снимков, содержащем изображения из баз данных Landsat-8 и PlanetScope, перемешанных между собой и разбитых для обучения на группы из 300 изображений размером 224x224 пикселей.

В результате применения такого подхода, объекты класса «лес» обнаруживались с точностью 92,14%, что на 18,3% больше, чем у алгоритма, обученного в первый раз. Для объектов других классов также удалось повысить точность детектирования из-за более четких границ объектов и снижения цены ошибки за счет уменьшения площади заснятой территории, которую покрывали изображения, с 30 м до 3 м.

|  |  |
| --- | --- |
| a) | b) |
| c) | d) |

Рис. 9. Примеры работы сети U-Net по обнаружению объектов класса «лес» (слева) и «сельскохозяйственные угодья» (справа) для спутниковых снимков из баз данных a), b) Landsat-8 и c), d) PlanetScope

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.

СПИСОК ЛИТЕРАТУРЫ

[1] Y. Goodfellow, Y. Bengio, A. Courville, “Deep Learning”, *The MIT Press*, 2016, 800 p.

[2] A. Van Etten, “You Only Look Twice: Rapid Multi-Scale Object Detection In Satellite Imagery”, Web: https://arxiv.org/abs/1805.09512.

[3] “LandSat Database”, Web: https://landsat.visibleearth.nasa.gov.

[4] E. Maggiori, Y. Tarabalka, G. Charpiat, P. Alliez, “Can Semantic Labeling Methods Generalize to Any City? The Inria Aerial Image Labeling Benchmark”, *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. 2017.

[5] “SpaceNet Database”, Web: http://explore.digitalglobe.com/spacenet.

[6] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation”, *CoRR, abs/1411.4038*, 2014.

[7] S. Ohleyer, “Building segmentation on satellite images”, Web: https://project.inria.fr/aerialimagelabeling/files/2018/01/fp\_ohleyer\_compressed.pdf.

# [8] 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.

[9] J. Patravali, S. Jain, S. Chilamkurthy, “2D-3D Fully Convolutional Neural Networks for Cardiac MR Segmentation”, Qure.AI, 2017.

[10] G. Chhor, C.B. Aramburu, “Satellite Image Segmentation for Building Detection using U-net”, Web: http://cs229.stanford.edu/proj2017/final-reports/5243715.pdf.

[11] A. Buslaev, S. Seferbekov, V. Iglovikov, A. Shvets, “Fully Convolutional Network for Automatic Road Extraction from Satellite Imagery”, *CoRR, abs/1806.05182,* 2018.

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

[13] H. Noh, S. Hong, B. Han, “Learning Deconvolution Network for Semantic Segmentation”, *ICCV*, 2015, pp. 1520 – 1528.

[14] V. Khryashchev, V. Pavlov, A. Priorov, E. Kazina, “Convolutional Neural Network for Satellite Imagery”, *Proceedings of the 22th Conference of Open Innovations Association FRUCT'22*. Jyvaskyla, Finland, 2018, pp. 344-347.

[15] S. Basu, S. Ganguly, S. Mukhopadhyay, R. DiBiano, M. Karki, R. Nemani, “DeepSat - A Learning framework for Satellite Imagery”, Web: https://arxiv.org/abs/1805.09512.

[16] “DSTL Satellite Imagery Feature Detection”, Web: <https://www.kaggle.com/c/dstl-satellite-imagery-feature-detection>.

[17] “Landsat8”, Web: https://landsat.usgs.gov/landsat-8.

[18] 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: pp. 234-241, 2015.

[19] V. Badrinarayanan, A. Kendall, R. Cipolla, “SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39(12), 2017, pp. 2481 – 2495.

[20] A. Chaurasia, E. Culurciello, “LinkNet: Exploiting Encoder Representations for Efficient Semantic Segmentation”, *IEEE Visual Communications and Image Processing (VCIP)*, 2017.

[21] A. Gulli, S. Pal., “Deep Learning with Keras”, *Packt Publishing*, 2017, 320 p.

[22] J. Sanders, E. Kandrot, “CUDA by Example: An Introduction to General-Purpose GPU Programming”, *Addison-Wesley Professional*, 2010, 320 p.

[23] D. P. Kingma, J. Ba, “Adam: A Method for Stochastic Optimization”, Web: https://arxiv.org/abs/1412.6980.

[24] P. T. de Boer., D. Kroese, S. Mannor, R. Rubinstein, “A Tutorial on the Cross-Entropy Method”, *Annals of operations research*, vol. 134(1), 2005, pp. 19-67.

[25] J. VanderPlas, “Python Data Science Handbook: Essential Tools for Working with Data. First Edition”*,* *O’Reilly Media*, 2016, 541 p.

[26] "Cloud platform for computer vision", <http://supervise.ly>.