

УДК 004.93*12

СЕМАНТИЧЕСКАЯ СЕГМЕНТАЦИЯ ДАННЫХ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ ПРИ ПОМОЩИ НЕЙРОСЕТЕВЫХ АЛГОРИТМОВ

Друки Алексей Алексеевич¹,
druki@tpu.ru

Спицын Владимир Григорьевич¹,
spvg@tpu.ru

Болотова Юлия Александровна¹,
julya.bolotova@gmail.com

Башлыков Артём Андреевич¹,
Zagta12345@gmail.com

¹ Национальный исследовательский Томский политехнический университет,
Россия, 634050, г. Томск, пр. Ленина, 30.

Актуальность исследования обусловлена необходимостью разработки методов, алгоритмов и программных средств, обеспечивающих повышение эффективности семантической сегментации данных на снимках дистанционного зондирования Земли. Несмотря на интенсивное развитие современных методов и алгоритмов, зачастую они не обеспечивают требуемое качество работы и надежность, поэтому на сегодняшний день существует потребность в повышении качества семантической сегментации объектов на изображениях.

Цель исследования: разработка алгоритмов, предназначенных для решения задачи семантической сегментации данных на снимках дистанционного зондирования Земли.

Объекты исследования: нейросетевые алгоритмы, обеспечивающие семантическую сегментацию данных на изображениях; методы реализации и обучения искусственных нейронных сетей; алгоритмы обработки изображений.

Методы. Для решения поставленных задач были использованы методы вычислительного интеллекта; методы классификации образов на изображениях; теория искусственных нейронных сетей; методы обучения искусственных нейронных сетей с использованием программного обеспечения Visual Studio; библиотеки глубокого обучения и реализации нейросетевых алгоритмов Caffe.

Результаты. Был произведён обзор методов и алгоритмов, позволяющих осуществлять семантическую сегментацию данных на спутниковых снимках земной поверхности. На основании проведенного анализа был сделан вывод, что для решения поставленной задачи наиболее эффективно использовать нейросетевые алгоритмы. Разработана сверточная нейронная сеть с оригинальной архитектурой, состоящая из шести слоев. Выполнена программная реализация алгоритма, позволяющая строить карту с изображением контуров зданий, дорог и фона, на основе входных данных. Представлено сравнение результатов работы различных алгоритмов обучения для разработанной нейронной сети.

Ключевые слова:

Нейронные сети, семантическая сегментация, классификация образов, обработка изображений, дистанционное зондирование Земли.

Введение

Семантическая сегментация изображений заключается в выделении на изображении локальных областей (сегментов), соответствующих различным классам объектов. Сегментация снимков дистанционного зондирования Земли находит применение во множестве областей: геоинформатика, инжиниринг георесурсов, автоматическое создание карт местности, градостроительство, анализ землепользования и т. д. Несмотря на интенсивное развитие методов и алгоритмов классификации объектов на изображениях, данная задача до сих пор не решена в полной мере и зачастую некоторые этапы данного процесса выполняются операторами вручную, что приводит к большим временным затратам и снижению эффективности. В процессе выполнения работы человеку свойственно уставать, терять концентрацию и эффективность его работы снижается.

На сегодняшний день существует множество методов и алгоритмов классификации объектов на изображениях, в том числе на снимках земной поверхности, обеспечивающих хорошие результаты работы. Однако данным алгоритмам присущи определенные недостатки, которые более подробно описаны в следующем разделе [1]. Таким образом, на сегодняшний день является актуальной задача разработки методов, алгоритмов и программных средств, позволяющих автоматизировать данный процесс и повысить эффективность его работы.

В статье представлен обзор существующих методов и алгоритмов, предназначенных для решения задачи семантической сегментации изображений. Представлены достоинства и недостатки существующих алгоритмов. В процессе исследования было выявлено, что одним из наиболее эффективных подходов к решению данной задачи является применение нейросетевых алгоритмов, а

именно сверточных нейронных сетей, так как они обладают рядом преимуществ: повышенная устойчивость к искажению входных данных; высокая эффективность классификации образов; меньшее количество настраиваемых параметров. На основании этого было принято решение разработать сверточную нейронную сеть для решения задачи семантической сегментации данных на снимках дистанционного зондирования Земли.

Обзор алгоритмов семантической сегментации изображений

Основные подходы к семантической сегментации изображений включают совокупность использования трёх типов алгоритмов: детекторы, дескрипторы и классификаторы, которые определяют основные параметры изображения, выделяют объекты и классифицируют их. В качестве основных параметров изображения могут служить: яркость, цвет, текстура, углы и границы объектов на изображении и тому подобное.

Среди наиболее популярных и эффективных алгоритмов, включающих в себя детекторы и дескрипторы, можно выделить алгоритмы SIFT, SURF, FAST, MSER, HOG [1–6].

Алгоритм SIFT (Scale Invariant Feature Transform) включает в себя детектор и дескриптор. Детектор SIFT основан на использовании масштабируемых пространств – набора всевозможных, сглаженных определённым фильтром версий одного изображения. При использовании гауссового фильтра данное масштабируемое пространство становится инвариантно к сдвигам, вращениям и не смещающему локальные экстремумы масштабу. Определение ключевых точек выполняется по трем параметрам: смещение от точного экстремума при помощи многочлена Тейлора; значение контраста разности гауссианов; нахождение точки на границе объекта при помощи матрицы Гессе. Далее вычисляется ориентация ключевой точки исходя из направления градиентов соседних точек [1, 2].

Работа дескриптора SIFT заключается в определении градиента изображения, его направления и модуля, умноженного на вес, для каждого из четырёх секторов вокруг особой точки. Для каждого сектора составляется гистограмма направлений градиента, каждое вхождение которого взвешено модулем этого градиента.

Алгоритм SURF (Speeded Up Robust Features) является модернизацией детектора SIFT, но вместо функции Гаусса используется её приближение прямоугольным фильтром 9×9 , что ускоряет получение результата алгоритма.

В дескрипторе SURF вокруг точки интереса строится квадратная область и разделяется на квадратные сектора, в которых вычисляются отклики на вейвлеты Хаара, направленные вертикально и горизонтально. Данные отклики взвешиваются и суммируются по каждому из секторов [3].

Алгоритм FAST (Features from Accelerated Segment Test) не требует вычисления производных по яркости, а сравнивается яркость в окрестности от проверяемой точки. Сначала проводится быстрый тест четырёх точек от исследуемой, а затем проверяются остальные. Количество проверок и их последовательность определяются на обучающей выборке [4].

Алгоритм MSER (Maximally Stable Extremal Regions) основан на определении интенсивности пикселей изображения и сравнение их с некоторым порогом (если интенсивность пикселя больше порога, считаем его белым, иначе – чёрным). Таким образом, строится пирамида изображений, в начале которой стоят белые изображения, а в конце – чёрные. Такая пирамида позволяет построить множество связанных компонент интенсивности, которые будут инвариантны к аффинным преобразованиям [5].

Алгоритм HOG (Histogram of Oriented Gradients) – дескриптор ключевых точек, основанный на подсчёте направлений градиента в локальных областях изображения. Изображение разделяется на небольшие связанные области, которые называются ячейками, и для каждой ячейки рассчитывается гистограмма направлений градиентов и направлений краёв для пикселей внутри ячейки. Выходом дескриптора является комбинация этих гистограмм [6].

К достоинствам перечисленных алгоритмов можно отнести высокую устойчивость к различным геометрическим и фотометрическим преобразованиям и изменению масштаба изображения. Недостатком данных алгоритмов является низкая устойчивость работы при изменении углов регистрации, условий освещения и при отражающих поверхностях.

Рассмотрим три класса широко используемых классификаторов: мешок слов, метод опорных векторов и нейронные сети.

Алгоритм мешок слов (Bag of words) относится к числу наиболее распространённых классов алгоритмов классификации изображений [7]. Основные шаги данного алгоритма:

- 1) определить особые (ключевые) точки на изображении;
- 2) построить дескрипторы этих точек;
- 3) провести кластеризацию дескрипторов, принадлежащих к объектам обучающей выборки (то есть заполнить словарь «словами»);
- 4) построить описание каждого изображения в виде нормированной гистограммы встречаемости «слов» (вычисление для каждого из кластеров количества отнесённых к нему особых точек изображения);
- 5) построить классификатор.

Основные недостатки данного классификатора заключаются в большом размере словаря. Также данный алгоритм не учитывает пространственную информацию об объекте, и при наличии схожих особых точек на разных объектах их описания могут совпадать [7].

Метод опорных векторов (SVM – Support Vector Machine) – один из наиболее популярных методов обучения классификаторов. Идея данного метода заключается в нахождении гиперплоскости для разделения объектов на классы. Для того чтобы гиперплоскость более устойчиво разделяла классы, она должна находиться от точек выборки на максимальном расстоянии. Это достигается, когда норма вектора минимальна. Опорными векторами называются множества, лежащие на границах областей, разделённых гиперплоскостью [8].

К достоинствам данного алгоритма помимо простоты реализации можно отнести возможность применения небольшой обучающей выборки, при этом классификатор может показывать приемлемый результат. Это же является и недостатком – алгоритм использует не всё множество, а лишь небольшую их часть на границах областей.

Искусственные нейронные сети (ИНС) – алгоритмы, моделирующие способ обработки информации человеческого мозга. Представляют из себя распределённый параллельный процессор, состоящий из элементарных частиц обработки информации (нейроны) и связей между ними (синаптические веса), позволяющих накапливать знания из окружающей среды и использовать в процессе обучения [9]. Обработка информации нейроном ведётся посредством умножения всех входных сигналов на соответствующие синаптические связи, их суммирования и обработки функцией активации нейрона, после чего сигнал распространяется дальше по сети. Обучение сети заключается в настройке синаптических весов в зависимости от выхода сети [9].

Основным недостатком классических нейронных сетей для обработки изображений является большой размер входного вектора данных, в следствии этого увеличивается количество нейронов каждого слоя и сеть становится громоздкой и сложной для обучения. Также классические нейронные сети не могут учесть топологию исходного изображения, так как принимают его целиком.

Этих недостатков лишены свёрточные нейронные сети (СНС), которые состоят из специальных свёрточных и подвыборочных слоев. СНС имеют двумерную иерархическую структуру, нейроны сгруппированы в плоскости, а плоскости сгруппированы в слои. От слоя к слою размер плоскостей постепенно уменьшается, таким образом последние слои имеют классический вид многослойного перцептрона.

СНС основаны на принципах локального восприятия и разделяемых весов. Под локальным восприятием подразумевается, что на вход одного нейрона поступают не все выходы предыдущего слоя, а лишь некоторая определённая их часть [10, 11].

Под разделяемыми весами подразумевается, что для большинства связей используется небольшой набор весов, называемых ядрами. Ядро представляет из себя матрицу, которая применяется к вектору данных посредством математической опе-

рации свёртки. Суть этой операции заключается в поэлементном умножении фрагмента изображения на матрицу ядра, суммировании полученных значений и записи результата в аналогичную позицию выходного изображения, которое называется картой признаков. Таким образом, свёрточные нейронные сети обладают гораздо меньшим количеством настраиваемых параметров. Также данный тип нейронных сетей обладает высокой устойчивостью к изменению масштаба, сдвигу и повороту и к другим преобразованиям входных данных [11–13].

Разработка свёрточной нейронной сети

В последние годы СНС показывают высокие результаты при решении задач классификации объектов на изображениях. Эффективность этого подхода объясняется тем, что свёрточные нейронные сети являются гибким инструментом и позволяют адаптировать свою структуру и параметры для решения задачи.

На сегодняшний день не существует четко регламентированных правил реализации структуры СНС: количество и организация слоев, количество и размер карт признаков, размер матриц свёртки, выбор алгоритма обучения. При разработке структуры СНС нужно понимать, что небольшое количество параметров сети может снизить точность классификации. Однако большое количество параметров увеличивает вычислительную сложность сети и не всегда обеспечивает улучшение классифицирующих способностей сети. Таким образом, разработка оптимальной структуры нейронной сети является эмпирическим процессом, который включает в себя большое количество экспериментальных исследований. Основной целью экспериментов являлось построение конфигурации нейронной сети с наименьшим числом параметров.

В процессе проведенных экспериментальных исследований были реализованы СНС различной архитектуры, включающие в себя разное количество параметров. Эксперименты показали, что нейронные сети с упрощенной архитектурой и малым количеством параметров показывали худшие результаты. При последовательном усложнении архитектуры СНС удалось подобрать оптимальную архитектуру, обеспечивающую высокие результаты классификации (рис. 1). Дальнейшие эксперименты по усложнению архитектуры и увеличению количества параметров СНС не обеспечили улучшение качества классификации, но при этом время работы и обучения сети значительно увеличилось.

Данная нейронная сеть состоит из 6 слоев и включает в себя 3 свёрточных слоя, 1 подвыборочный слой и 2 полносвязных слоя. В качестве входных данных используются цветные изображения.

Входной слой имеет размер 64×64 нейрона. Данный слой не выполняет каких-либо преобразований и предназначен лишь для подачи на него входных данных.

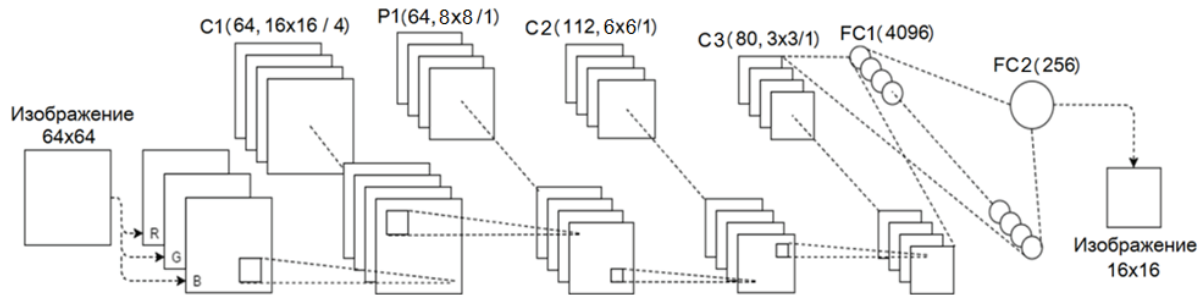


Рис. 1. Архитектура разработанной свёрточной нейронной сети

Fig. 1. Architecture of the developed convolutional neural network

Следом за входным слоем расположен первый скрытый слой C1. Данный слой является свёрточным, содержит 64 карты признаков, каждая из которых имеет размер 16×16 нейронов. Матрица свертки имеет размер 4×4 нейрона. Смещение выполняется на 4 нейрона.

Второй скрытый слой P1 является слоем подвыборки, он состоит 64 карт признаков, каждая из которых имеет размер 8×8 нейронов. Матрица свертки имеет размер 2×2 нейрона. Смещение выполняется на 1 нейрон. Данный слой уменьшает размер предыдущего слоя в два раза.

Третий скрытый слой C2 является свёрточным и состоит из 112 карт признаков, каждая из которых имеет размер 6×6 нейронов. Матрица свертки имеет размер 2×2 нейрона. Смещение выполняется на 1 нейрон.

Четвёртый скрытый слой также является свёрточным и состоит из 80 карт признаков размера 3×3 нейрона. Матрица свертки имеет размер 3×3 нейрона. Смещение выполняется на 1 нейрон.

Пятый скрытый слой FC1 является полносвязным. Данный слой состоит из 4096 нейронов и имеет структуру в виде одномерного вектора.

Шестой скрытый слой FC2 состоит из 256 нейронов и также имеет структуру в виде одномерного вектора.

Первые четыре слоя сети имеют двумерную структуру и предназначены для извлечения признаков на изображении. Последние два слоя имеют структуру одномерного вектора и предназначены для классификации признаков, извлеченных на предыдущих слоях.

На выходе нейронная сеть генерирует вектор из 256 значений, который преобразуется в двумерную матрицу размерами 16×16 пикселей в градациях серого цвета. Значения каждого пикселя выходного изображения находятся в диапазоне от 0 до 255.

Инициализация синаптических коэффициентов сети задавалась случайным образом в диапазоне от 0 до 1.

При разработке структуры нейронной сети также необходимо подобрать функцию активации, которая предназначена для вычисления выходного сигнала искусственного нейрона. Одними из наиболее эффективных функций активации являются

сигмоидальные функции: логистическая и гиперболический тангенс. Данные функции являются похожими в графическом представлении, но имеют различные диапазоны значений [14].

Достоинства сигмоидальных функции заключается в том, что они позволяют усиливать слабые сигналы и не насыщаются при сильных сигналах.

Логистическая функция имеет область значений от 0 до 1 и задается формулой [14]:

$$f(x) = \frac{1}{(1 + e^{-s})},$$

где s – сумма входных значений нейрона.

Функция гиперболический тангенс имеет область значений от -1 до 1 и задается формулой [14]:

$$f(x) = a \tanh(bx),$$

где a и b – константы.

Для решения поставленной задачи была выбрана функция гиперболический тангенс, так как она имеет ряд преимуществ, которые заключаются в следующем:

- является симметричной относительно начала координат и обеспечивает более быструю сходимость, по сравнению с логистической функцией;
- имеет простую производную;
- имеет максимум второй производной при $=1$.

Разработка нейронной сети выполнялась на языке программирования C++ при использовании библиотеки глубокого обучения Caffe [15]. Caffe – это библиотека, позволяющая реализовывать алгоритмы глубокого обучения при использовании архитектуры параллельных вычислений Cuda. На сегодняшний день данная библиотека является одним из самых эффективных инструментов реализации СНС и позволяет обрабатывать миллионы изображений в день на одном компьютере. В режиме CPU вычислительные затраты занимают около 20 мс при обработке изображений в пакетном режиме.

Выбор алгоритма обучения

Обучение нейронных сетей заключается в последовательной корректировке синаптических весов между нейронами. Одним из самых распро-

страненных и эффективных алгоритмов обучения для нейронных сетей является алгоритм обратного распространения ошибки [10, 14]. Своё название алгоритм получил благодаря тому, что ошибка, вычисляемая на каждой итерации, распространяется по ИНС от выхода к входу с целью перенастройки синаптических весов. В процессе обучения сети, при подаче входного вектора, выход сети сравнивается с выходом из обучающей выборки, формируя ошибку [14]:

$$E_k = \frac{1}{2} \sum_{j=0}^N (t_{kj} - x_{kj})^2,$$

где t_{kj} – желаемое значение выхода нейрона; x_{kj} – текущее значение выхода нейрона.

Корректировка синаптических весов производится по следующей формуле [14]:

$$\Delta w_{ij} = -\eta \delta_{kj} x_{ij},$$

где η – коэффициент скорости обучения; x_{ij} – значения входа нейрона; δ_{kj} – ошибка нейрона.

Значение ошибки нейронов сети определяется по формуле [14]:

$$\delta_i^{(q)} = (f_i^{(q)}(S))' \sum_j w_{ij} \delta_j^{(q+1)},$$

где $\delta_i^{(q)}$ – значение ошибки i -го нейрона в слое q ; $\delta_j^{(q+1)}$ – значение ошибки j -го нейрона в слое $q+1$; w_{ij} – вес связи, соединяющей два нейрона; $(f_i^{(q)}(S))'$ – значение производной активационной функции i -го нейрона в слое q .

Достоинствами данного алгоритма обучения являются: простота реализации, возможность использовать множество функций потерь, возможность применения при больших объемах данных. К недостаткам алгоритма можно отнести малую корректировку весов, что ведет к долгому процессу обучения. При этом возникает задача выбора оптимального размера шага. Слишком маленький размер шага приводит к медленной сходимости алгоритма, слишком большой размер шага может привести к потере устойчивости процесса обучения [14].

Для решения данных проблем существуют различные методы оптимизации данного алгоритма. Из множества существующих методов оптимизации для обучения и последующего сравнения их работы были выбраны следующие:

- Nesterov accelerated gradient – метод оптимизации, основанный на идее накопления импульса, т. е. при длительном движении в одном направлении скорость будет сохраняться некоторое время. Для этого необходимо сохранять несколько предыдущих значений параметров и вычислять среднее значение, что занимает слишком много памяти [16].
- AdaGrad (Adaptive gradient) – алгоритм оптимизации, принцип которого заключается в сохранении частоты изменения градиента. Идея заключается в сохранении суммы квадратов обновлений для каждого параметра сети. На основе этой величины регулируются обновления па-

раметров – часто встречаемые обновляются реже, освобождая место для редко встречаемых, тем самым обеспечивается адаптивная скорость обучения сети [17].

- Adam (Adaptive moment estimation) – алгоритм оптимизации, сочетающий в себе принципы накопления импульса и сохранения частоты изменения градиента. Данный метод обладает преимуществами обоих рассмотренных выше методов [18].

Обучение и тестирование разработанного алгоритма

Для обучения и тестирования разработанной СНС использовалась база изображений, состоящая из нескольких тысяч спутниковых снимков штата Массачусетс (США) [19]. Размер каждого изображения 1500×1500 пикселей, с разрешением $1 \text{ м}^2/\text{пиксель}$. Среди изображений присутствуют спутниковые снимки (рис. 2) и соответствующие им карты сегментированных объектов (рис. 3). Все изображения сгруппированы на обучающую, тестовую и валидационную выборки в соотношении 0.7/0.2/0.1.

Как видно на рис. 2, изображения содержат различные классы объектов. Основные объекты, которые представляют интерес для поставленной задачи: здания, дороги и фон.

На рис. 3 представлены изображения (карты) сегментированных объектов. Данные изображения соответствуют исходным изображениям из обучающей выборки и предназначены для обучения СНС.

В процессе работы СНС обрабатывает небольшие участки входных изображений в соответствии с размером входного слоя (64×64 пикселя). Таким образом, происходит последовательное сканирование входного изображения окном размером 64×64 пикселя. В каждом местоположении данного окна нейронная сеть выполняет сегментацию признаков изображения, формируя на выходе карту размером 16×16 пикселей. Такая разница в размерах обусловлена тем, что при выборке небольшого участка изображения зачастую тяжело узнать, что на нём изображено (рис. 4). Увеличенный размер входного участка изображения позволяет сохранить часть данных для более эффективной классификации (рис. 4).

Для того чтобы избежать проблемы переобучения, в пятом полносвязном слое реализован метод *DropOut* [20], который заключается в том, что во время обучения из общей сети многократно и случайным образом выделяется некоторая подсеть и обновление весов происходит только в рамках этой подсети. Нейроны попадают в подсеть с вероятностью 0,5.

С целью регуляризации сети использовалась L2 регуляризация, которая заключается в большом штрафе слишком высокого значения веса и малом при низком значении, что выражается в использовании коэффициента регуляризации. Также при обучении выполнялась минимизация функции потерь при помощи метода Mini-batch градиентного спуска [20].



Рис. 2. Снимки дистанционного зондирования Земли

Fig. 2. Images of Earth remote sensing

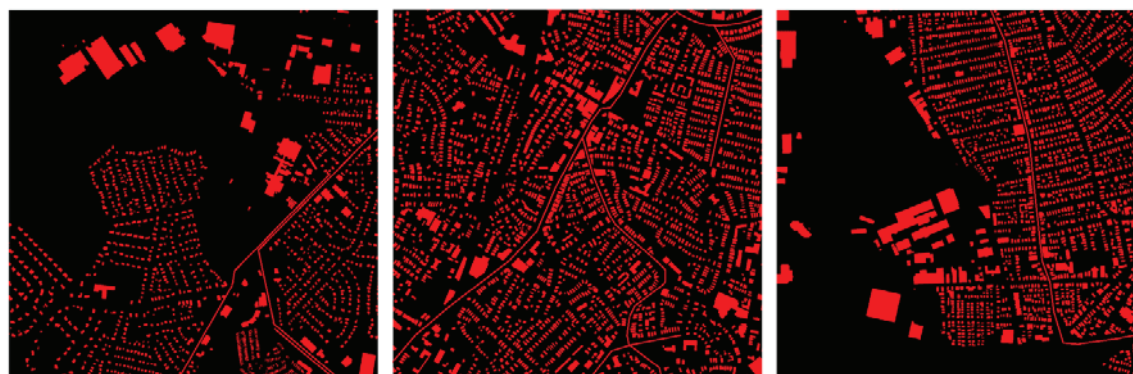


Рис. 3. Карты сегментированных объектов

Fig. 3. Maps of segmented objects

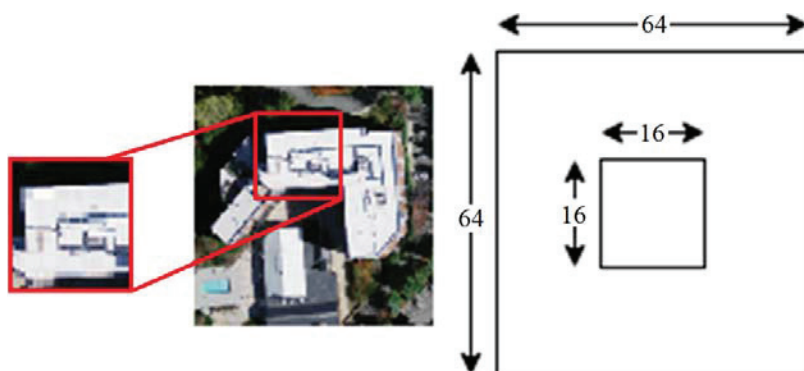


Рис. 4. Размеры входного и выходного изображения

Fig. 4. Sizes of input and output images

Для увеличения обучающей выборки и повышения инвариантности сети к различным поворотам и смене ракурса исходные изображения поворачивались на случайный угол.

При обучении и тестировании применялись следующие параметры нейронной сети:

- коэффициент обучения 0,0005;
- частота изменения коэффициента обучения 10^4 ;
- величина изменения коэффициента обучения 0,1;
- затухание для регуляризации L2 0,0005.

Все эксперименты проводились на следующем оборудовании: процессор Intel Core i5–4690k (4GHz), ОЗУ 8ГБ, видеокарта NVIDIA Quadro FX 4800.

Для обучения разработанной СНС исследовалась работа трех алгоритмов, которые более подробно описаны в предыдущем разделе: Nesterov accelerated gradient, AdaGrad и Adam. Конфигурация сети оставалась неизменной. Количество эпох обучения для каждого случая составило 400. В табл. 1 представлены результаты обучения.

Таблица 1. Результаты обучения СНС

Table 1. Results of convolution neural network (CNN) training

Название алгоритма Name of algorithm	Время обучения Time of training		Точность Accuracy, %
	ч/ч	мин/мин	
Nesterov accelerated gradient [16]	10	46	78,13
AdaGrad [17]		27	77,97
Adam [18]		24	85,31

Из результатов, представленных в табл. 1, видно, что алгоритм обучения Adam даёт лучшие результаты относительно остальных. Время обучения составило 10 ч 24 мин, точность классификации 85,31 %. Под точностью понимается совпадение пикселей выхода сети с соответствующими на тестовой выборке.

На рис. 5 представлены результаты сегментации изображения сверточной нейронной сетью. Слева

представлено исходное изображение (рис. 5, а), справа – результат сегментации (рис. 5, б).

Как видно из рис. 5, СНС обеспечивает достаточно высокую эффективность сегментации. Практически все здания на снимке были точно выделены, однако присутствуют погрешности при выделении дорог. В основном это связано с тем, что в некоторых областях изображения дороги имеют слабый контраст по отношению к остальному фону и плохо различимы. Таким образом, в дальнейшем планируется проведение экспериментов с применением алгоритмов по улучшению качества изображений, улучшению контраста и применение различных фильтров.

Сравнение с существующими аналогами

Результаты работы предложенной СНС были сопоставлены с результатами работы алгоритмов Inria2 и TeraDeep [21]. Данные алгоритмы являются закрытыми, поэтому не представлялось возможным реализовать их и протестировать. Результаты их работы взяты из открытых источников. В табл. 2 представлено сравнение результатов сегментации разработанной СНС с аналогами.

Таблица 2. Сравнение результатов работы СНС с аналогами

Table 2. Comparison of the neural network performance with the analogues

Название алгоритма Name of algorithm	Точность Accuracy, %
Разработанная СНС Developed neural network	85,31
Inria2	93,93
TeraDeep	94,41

Как видно из табл. 2, разработанная СНС уступает по точности сегментации алгоритмам Inria2 и TeraDeep, но тем не менее показывает хорошие результаты на уровне 85,31 %. Как уже было сказано выше, данные алгоритмы являются закрытыми, поэтому сложно сделать однозначные выводы о их достоинствах и причинах эффективности.



Рис. 5. Результаты работы СНС: а) исходное изображение; б) результат сегментации

Fig. 5. Results of the CNN: a) original image; б) segmentation result

В будущем планируется оптимизация структуры СНС, применение новых подходов к обучению, а также расширение обучающей выборки с целью получения лучших результатов сегментации.

Заключение

В процессе выполнения работы была реализована СНС для решения задачи семантической сегментации снимков дистанционного зондирования Земли. Были проведены эксперименты по выбору алгоритма обучения – лучшие результаты были получены с помощью алгоритма Adam.

Сравнение в другими методами показало, что разработанная сверточная нейронная сеть уступает некоторым аналогам, но, тем не менее, выполняет поставленную задачу с достаточно высоким результатом. Разработанная нейронная сеть выполняет семантическую сегментацию для трех классов объектов: здания, дороги и фон.

Работа выполнена в рамках Программы повышения конкурентоспособности ТПУ и при финансовой поддержке РФФИ в рамках научного проекта №18-08-00977 А "Создание интеллектуальной системы детектирования, распознавания и понимания искаженных печатных текстов на изображениях и видео».

СПИСОК ЛИТЕРАТУРЫ

1. Ивашечкин А.П., Василенко А.Ю., Гончаров Б.Д. Методы нахождения особых точек изображения и их дескрипторов // Молодой ученый. – 2016. – № 15. – С. 138–140.
2. Park S., Yoo J.H. Real-time face recognition with SIFT based local feature points for mobile devices // The 1st International Conference on Artificial Intelligence, Modelling and Simulation (AIMS 13). – Malaysia, 2013. – P. 304–308.
3. Tawfiq A., Ahmed J. Object detection and recognition by using enhanced Speeded Up Robust Feature // International Journal of Computer Science and Network Security. – 2016. – V.16. – № 4. – P. 66–71.
4. Tore V., Chawan P.M. FAST Clustering Based Feature Subset Selection Algorithm for High Dimensional Data // International Journal of Computer Science and Mobile Computing. – 2016. – V. 5. – № 7. – P. 234–238.
5. Mammeri A., Boukerche A., Khiari E. MSER-based text detection and communication algorithm for autonomous vehicles // IEEE Symposium of Computers and Communication. – Messina, Italy, 2016. – P. 456–460.
6. Dalal N., Triggs B. Histograms of Oriented Gradients for Human Detection // IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). – San Diego, USA, 2005. – V. 1. – P. 886–893.
7. Mohey D.E. Enhancement Bag-of-Words Model for Solving the Challenges of Sentiment Analysis // International Journal of Advanced Computer Science and Applications. – 2016. – V. 7. – № 1. – P. 244–251.
8. Kecman V., Melki G. Fast online algorithms for Support Vector Machines // IEEE South East Conference (SoutheastCon 2016). – Virginia, USA, 2016. – P. 26–31.
9. Хайкин С. Нейронные сети: полный курс, 2-е изд. – М.: Изд-во «Вильямс», 2006. – 1104 с.
10. Efficient BackProp / Y. LeCun, L. Bottou, G.B. Orr, K.R. Muller // Neural Networks: Tricks of the trade. – Berlin: Springer, 1998. – 44 p.
11. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition / K. He, X. Zhang, S. Ren, J. Sun // Transactions on Pattern Analysis and Machine Intelligence (TPAMI). – 2015. – V. 4. – P. 534–542.
12. Howard A.G. Some Improvements on Deep Convolutional Neural Network Based Image Classification // International Conference on Learning Representations (ICLR). – Banff, Canada, 2014. – V. 10. – № 4. – P. 652–659.
13. Krizhevsky A., Sutskever I., Hinton G. ImageNet Classification with Deep Convolutional Neural Networks // Conference on Neural Information Processing Systems (NIPS). – Nevada, USA, 2012. – P. 27–35.
14. Спицын В.Г., Цой Ю.Р. Применение искусственных нейронных сетей для обработки информации. – Томск: Изд-во ТПУ, 2007. – 32 с.
15. Caffe deep learning framework. URL: <http://caffe.berkeleyvision.org> (дата обращения: 20.10.2017).
16. Botev A., Lever G., Barber D. Nesterov's Accelerated Gradient and Momentum as approximations to Regularised Update Descent // Machine Learning. – 2016. – V. 1. – P. 1–7.
17. Duchi J., Hazan E., Singer Y. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization // Journal of Machine Learning Research. – 2011. – V. 12. – P. 1–39.
18. Diederik P.K., Ba J. Adam: a Method for Stochastic Optimization // The 3rd International Conference for Learning Representations. – San Diego, USA, 2015. – P. 1–15.
19. Road and building detection datasets. URL: <http://www.cs.toronto.edu/~vmnih/data> (дата обращения: 20.10.2017).
20. Dropout: a Simple Way to Prevent Neural Networks from Overfitting / N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov // Journal of Machine Learning Research. – 2014. – V. 15. – P. 1929–1958.
21. TeraDeep. URL: <https://www.teradeep.com> (дата обращения: 20.10.2017).

Поступила 13.12.2017 г.

Информация об авторах

Друки А.А., кандидат технических наук, доцент Инженерной школы информационных технологий и робототехники Национального исследовательского Томского политехнического университета.

Спицын В.Г., доктор технических наук, профессор Инженерной школы информационных технологий и робототехники Национального исследовательского Томского политехнического университета.

Болотова Ю.А., кандидат технических наук, доцент Инженерной школы информационных технологий и робототехники Национального исследовательского Томского политехнического университета.

Башлыков А.А., студент Инженерной школы информационных технологий и робототехники Национального исследовательского Томского политехнического университета.

UDC 004.93'12

SEMANTIC SEGMENTATION OF EARTH SCANNING IMAGES USING NEURAL NETWORK ALGORITHMS

Alexey A. Druki¹,
druki@tpu.ru

Vladimir V. Spitsyn¹,
druki@tpu.ru

Yulia A. Bolotova¹,
druki@tpu.ru

Artyom A. Bashlykov¹,
Zagta12345@gmail.com

¹ National Research Tomsk Polytechnic University,
30, Lenin Avenue, Tomsk, 634050, Russia.

The relevance of the research is caused by the need to develop methods, algorithms and software to improve the effectiveness of semantic segmentation of the Earth scanning images. Today there is a need to improve the quality of semantic segmentation of objects in images, despite the intensive development of modern methods and algorithms, often they do not provide the required quality of work and reliability.

The main aim of the research is to develop the algorithms to solve the problem of semantic segmentation of the Earth scanning images.

Objects: neural network algorithms which provide semantic segmentation of images; methods of implementation and training of artificial neural networks; image processing algorithms.

Methods. To solve the tasks, the authors have used the methods of computational Intelligence; methods of pattern classification on images; theory of artificial neural networks; methods for training of artificial neural networks using Visual Studio software; deep learning framework Caffe for implementation of neural network algorithms.

Results. The authors made a review of methods and algorithms which allow carrying out semantic segmentation of images. Based on the analysis, it was concluded that neural network algorithms provide more efficient results. The authors developed the convolutional neural network with the original architecture consisting of six layers. Software implementation of the described algorithms is implemented. It allows building a map of segmented buildings, roads and background based on input data. The paper introduces the comparison of results of using different learning algorithms for the developed neural network.

Key words:

Neural networks, semantic segmentation, pattern classification, image processing, remote scanning of the Earth.

The research was carried out within the TPU Competitiveness Enhancement Program. It was financially supported by the RFBR within the project no. 18-08-00977A «Development of the intelligent system of detecting, recognizing and understanding corrupted printed texts in the pictures and video».

REFERENCES

1. Ivashechkin A.P., Vasilenko A.Yu., Goncharov B.D. Methods of finding the singular points on image and their descriptors. *Young scientist*, 2016, no. 15, pp. 138–140. In Rus.
2. Park S., Yoo J.H. Real-time face recognition with SIFT based local feature points for mobile devices. *The 1st International Conference on Artificial Intelligence, Modelling and Simulation (AIMS 13)*. Malaysia, 2013. pp. 304–308.
3. Tawfiq A., Ahmed J. Object detection and recognition by using enhanced Speeded Up Robust Feature. *International Journal of Computer Science and Network Security*, 2016, vol. 16, no. 4, pp. 66–71.
4. Tore V., Chawan P.M. FAST Clustering Based Feature Subset Selection Algorithm for High Dimensional Data. *International Journal of Computer Science and Mobile Computing*, 2016, vol. 5, no. 7, pp. 234–238.
5. Mammeri A., Boukerche A., Khiari E. MSER-based text detection and communication algorithm for autonomous vehicles. *IEEE Symposium of Computers and Communication*. Messina, Italy, 2016. pp. 456–460.
6. Dalal N., Triggs B. Histograms of Oriented Gradients for Human Detection. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*. San Diego, USA, 2005. Vol. 1, pp. 886–893.
7. Mohey D.E. Enhancement Bag-of-Words Model for Solving the Challenges of Sentiment Analysis. *International Journal of Advanced Computer Science and Applications*, 2016, vol. 7, no. 1, pp. 244–251.
8. Kecman V., Melki G. Fast online algorithms for Support Vector Machines. *IEEE South East Conference (SoutheastCon 2016)*. Virginia, USA, 2016. pp. 26–31.
9. Haykin S. *Neyronnye seti: polny kurs* [Neural networks: full course]. Moscow, Williams Publ., 2006. 1104 p.
10. LeCun Y., Bottou L., Orr G.B., Muller K.R. Efficient BackProp. *Neural Networks: Tricks of the trade*. Berlin, Springer, 1998. 44 p.
11. He K., Zhang X., Ren S., Sun J. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2015, vol. 4, pp. 534–542.
12. Howard A.G. Some Improvements on Deep Convolutional Neural Network Based Image Classification. *International Conference on Learning Representations (ICLR)*. Banff, Canada, 2014. Vol. 10, no. 4, pp. 652–659.

13. Krizhevsky A., Sutskever I., Hinton G. ImageNet Classification with Deep Convolutional Neural Networks. *Conference on Neural Information Processing Systems (NIPS)*. Nevada, USA, 2012. pp. 27–35.
14. Spitsyn V.G., Tsoy Yu.R. *Primenenie iskusstvennykh neyronnykh setey dlya obrabotki informatsii* [Application of artificial neural networks for information processing]. Tomsk, TPU Publ. house, 2007. 32 p.
15. *Caffe deep learning framework*. Available at: <http://caffe.berkeleyvision.org> (accessed 20 October 2017).
16. Botev A., Lever G., Barber D. Nesterov's Accelerated Gradient and Momentum as approximations to Regularised Update Descent. *Machine Learning*, 2016, vol. 1, pp. 1–7.
17. Duchi J., Hazan E., Singer Y. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. *Journal of Machine Learning Research*, 2011, vol. 12, pp. 1–39.
18. Diederik P.K., Ba J. Adam: a Method for Stochastic Optimization. *The 3rd International Conference for Learning Representations*. San Diego, USA, 2015. pp. 1–15.
19. *Road and building detection datasets*. Available at: <http://www.cs.toronto.edu/~vmnih/data> (accessed 20 October 2017).
20. Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 2014, vol. 15, pp. 1929–1958.
21. *TeraDeep*. Available at: <https://www.teradeep.com> (accessed 20 October 2017).

Received: 13 December 2017.

Information about the authors

Alexey A. Druki, Cand. Sc., senior lecturer, National Research Polytechnic University.

Vladimir V. Spitsyn, Dr. Sc., professor, National Research Tomsk Polytechnic University.

Yulia A. Bolotova, Cand. Sc., associate professor, National Research Tomsk Polytechnic University.

Artyom A. Bashlykov, student, National Research Tomsk Polytechnic University.