УДК 004.8

СВЕРТОЧНЫЕ НЕЙРОННЫЕ СЕТИ ДЛЯ СЕМАНТИЧЕСКОЙ СЕГМЕНТАЦИИ ИЗОБРАЖЕНИЙ ЗЕМНОЙ ПОВЕРХНОСТИ

CONVOLUTION NEURAL NETWORKS FOR SEMANTIC SEGMENTATION ON IMAGES OF EARTH SURFACE

Береснев Алексей Павлович
аспирант отделения информационных технологий,
Национальный исследовательский
Томский политехнический университет
apb3@tpu.ru

Зоев Иван Владимирович
аспирант отделения информационных технологий,
Национальный исследовательский
Томский политехнический университет
ivz3@tpu.ru

Марков Николай Григорьевич
доктор технических наук, профессор,
профессор отделения информационных технологий,
Национальный исследовательский
Томский политехнический университет
markovng@tpu.ru

Аннотация. В настоящее время для анализа изображений различной физической природы все чаще начинают применять сверточные нейронные сети (СНС). Для решения задачи семантической сегментации космических снимков земной поверхности предложены три новые архитектуры СНС подкласса LeNet5. Приводятся результаты исследования эффективности каждой из новых архитектур СНС.

Ключевые слова: свёрточные нейронные сети, анализ изображений земной поверхности, семантическая сегментация изображений.

Beresnev Alexey Pavlovich
PhD Student of Department
Information Technologies,
National Research
Tomsk Polytechnic University
apb3@tpu.ru

Zoev Ivan Vladimirovich
PhD student of Department
Information Technologies,
National Research
Tomsk Polytechnic University
ivz3@tpu.ru

Markov Nikolay Grigorievich
Doctor of Engineering, Professor,
Professor of Information Technology,
National Research
Tomsk Polytechnic University
markovng@tpu.ru

Annotation. Nowadays in images analysis for different tasks are beginning to use convolution neural networks (CNN). For solution of semantic image segmentation task are proposed three new CNN architecture subclass LeNet5. In this article we present results of effectiveness study for each of proposed CNN architectures.

Keywords: convolution neural networks, analysis of earth surface images, semantic image segmentation.

Введение

В последние годы все чаще для анализа различных изображений применяются сверточные нейронные сети (СНС). С помощью СНС можно решать следующие формализованные задачи: семантической сегментации изображений, задачу классификации и локализации объекта на изображении, детектирование объекта (предсказывается класс и оценивается положение каждого из группы объектов) и сегментации экземпляров объектов. Основываясь на результатах ряда исследователей различных архитектур СНС [1, 2], можно сформулировать следующий принцип: для решения каждой из этих задач существует своя наиболее эффективная (в первую очередь, по критерию точности распознавания объектов на изображениях) архитектура СНС.

Для реализации этого принципа применительно к решению задачи сегментации спутниковых снимков предлагаются и исследуются три новые архитектуры СНС, входящие в подкласс LeNet5.

СНС подкласса LeNet5

Ян Ле Кун с коллегами для решения задачи классификации объектов на изображениях предложил СНС с архитектурой LeNet5, которая хорошо себя зарекомендовала и сегодня считается классической [1]. На её основе создан целый подкласс СНС, получивший название LeNet5. Ранее нами решалась *задача обнаружения и классификации объектов на изображениях*, причём выявлялась принадлежность объекта к одному из 10 классов [2]. Для решения этой задачи нами предложена новая архитектура СНС (рис. 1), подобная классической LeNet5. В ней используются свёрточные слои (англ.

convolutionallayers), слои подвыборки (англ. poolinglayers), а в качестве функции активации после каждого слоя свертки применяется оператор ReLU. Назначение класса объекту по результатам работы СНС осуществляется с помощью известной процедуры Softmax.

Архитектура свёрточного слоя задается параметрами: *глубина L* – количество входных/выходных карт признаков; *высота h* и *ширина w* каждого из ядер свёртки; *шаг s*, с которым ядро свёртки двигается по входному слою. На рисунке 1 изображены 3 свёрточных слоя. Первый из них обладает следующими параметрами: *высота* и *ширина* каждого ядра равны 7 элементам, *шаг* равен 1 элементу, *глубина* равна 3 (три входные карты). Архитектурные параметры других двух свёрточных слоёв показаны на рисунке 1. В большинстве СНС конечные слои являются полносвязными. Можно задать параметры для свёрточного слоя таким образом, чтобы получить из него полносвязный слой. Так, полносвязный слой на рисунке 1 можно представить как свёрточный слой с такими параметрами: *высота* и *ширина* ядра свёртки будут равны 1, входные карты признаков будут размером 1×1, а их *глубина* равна 100, количество выходных карт признаков будет равным 10 (число классов объектов).

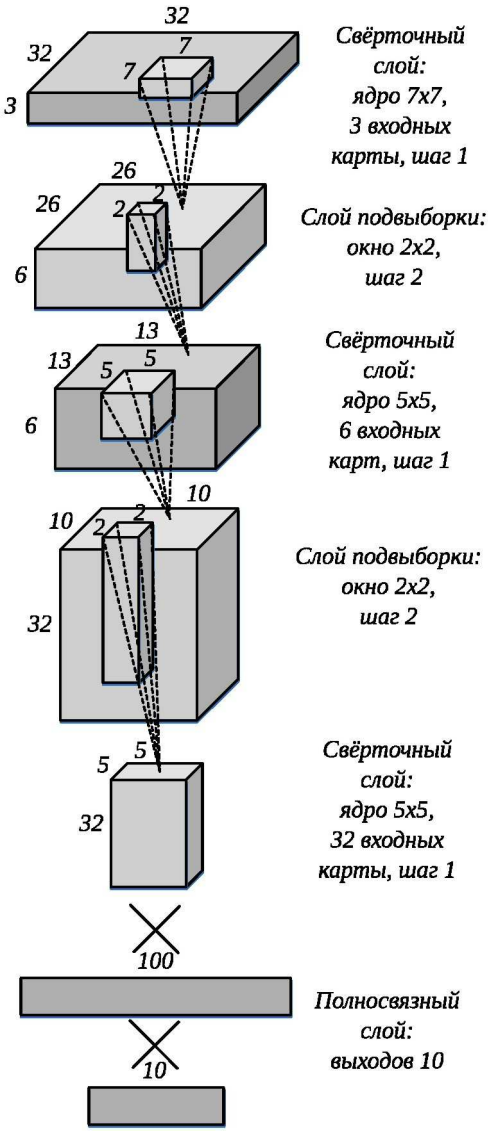


Рисунок 1 – Предложенная архитектура СНС из подкласса LeNet5

Подвыборка (также может называться «пуллинг» от англ. pooling) уменьшает размерность каждой карты признаков, но сохраняет наиболее значимую информацию. Из рисунка 1 видно, что после первого свёрточного слоя, результатом которого является 6 карт признаков размером 26×26 элементов, идёт слой подвыборки с окном 2×2 элемента и шагом 2. Входная карта признаков при подвыборке разбивается на области размером 2×2 элемента. Для каждой такой области выполняется процедура подвыборки, которая проводится по алгоритму выбора максимальных значений элементов (англ. maxpooling). Его использование обусловлено минимальным количеством операций над числами, что немаловажно при аппаратной реализации СНС. После процедуры подвыборки получается 6 карт при-

знаков размером 13×13 элементов. Другой слой подвыборки с таким же окном 2×2 элемента и шагом 2 выполняется после второго слоя свёртки. На выходе этого слоя подвыборки получается 32 карты признаков размером 5×5 элементов.

Предложенная архитектура СНС отличается от классической архитектуры LeNet5 наличием трёх свёрточных слоёв вместо двух свёрточных слоёв у классической архитектуры и наличием только одного полносвязного слоя вместо двух у классической архитектуры. Имеются также отличия в параметрах свёрточных слоёв. Все это ведёт к увеличению количества карт признаков и нацелено на увеличение точности обнаружения и классификации объектов на изображениях. Сравнение предложенной архитектуры с другими известными СНС подкласса LeNet5 позволяет считать её по ряду архитектурных признаков оригинальной.

Для извлечения ключевых признаков весовые коэффициенты процедуры свёртки настраиваются с использованием обучающей выборки. Существует большое число обучающих выборок, собранных для решения различного рода задач.

Архитекуры СНС для решения задачи семантической сегментации

Рассмотрим новые архитектуры СНС подкласса LeNet5 для решения задачи семантической сегментации. Известно, что снимки земной поверхности, получаемые при мониторинге опасных технологических объектов и промышленных предприятий, обычно имеют большие размеры, например 1200×1200 пикселей и более. Задача семантической сегментации таких снимков заключается в определении принадлежности каждого пикселя изображения к тому или иному классу. Решать её можно двумя способами. Первый из них предполагает, что используется предложенная нами и описанная выше архитектура СНС LeNet5 (рис. 1). При этом СНС будет анализировать небольшие участки входного изображения с размером 32×32 пикселя. Иными словами, необходимо сканирование входного изображения окном этого размера. Однако этот способ при его реализации требует от вычислительных устройств высокой производительности. Поэтому перспективным является другой предложенный нами способ, основанный на идее использования энкодера Region Proposal Network (RPN) нейросети Faster-RCNN [3] для анализа изображений разных масштабов. В соответствии с этой идеей первые пять слоёв разработанной нами архитектуры СНС на рисунке 1 модифицируются в энкодер, пригодный для анализа исходного изображения без перемещения по нему окна, то есть в целом. Это позволит значительно сократить время на анализ входного изображения больших размеров.

Пусть необходимо решить задачу сегментирования зданий (два класса: здания и фон) на изображениях 1500×1500 пикселей из выборки Massachusetts Roads Dataset спутниковых снимков коттеджных поселков г. Бостон (США) [4]. Исходным спутниковым изображениям с разрешением 1 м^2/пиксель из выборки соответствуют дополнительные изображения (карты) с выделенными на них зданиями и дорогами, которые можно использовать для тестирования обученных СНС. Для решения этой задачи нами была разработана архитектура СНС, получившая название SegLeNet (табл. 1). В этой архитектуре после первых пяти слоёв – модифицированного энкодера следуют еще два свёрточных слоя с размером ядра 1×1. Так как после первых пяти слоёв энкодера LeNet входное изображение уменьшается в 4 раза (из-за слоёв подвыборки), то карта признаков на выходе СНС интерполируется методом билинейной интерполяции до размеров входного изображения. Чтобы сохранить размеры карт признаков после слоёв свёртки, входные карты признаков дополняются нулевой рамкой (padding).

Таблица 1 – Архитектура СНС SegLeNet

№	Тип слоя	Число ядер свёртки	Размеры ядра/шаг	Размеры входного изображения/ карт признаков
1	Свёрточный	6	7×7/1	1500×1500×3
2	Подвыборки	–	2×2/2	1500×1500×6
3	Свёрточный	32	5×5/1	750×750×36
4	Подвыборки	–	2×2/2	375×375×32
5	Свёрточный	100	5×5/1	375×375×32
6	Свёрточный	10	1×1/1	375×375×100
7	Свёрточный	2	1×1/1	375×375×10
8	Интерполяция	–	–	375×375×2
9	Softmax	2	–	1500×1500×2

В известной СНС Unet [5] для сегментации изображений декодер использует слои транспонированной свёртки (англ. Transposed Convolutional), которые дополнены картами признаков из соответствующих слоёв энкодера сети. Используем идеи слоёв траспонированной свёртки при проектирова-

нии СНС новой архитектуры, названной ULeNet. В ней в качестве энкодера применяются модифицированные первые пять слоёв нашей СНС LeNet5 из рисунка 1. Далее, включены два слоя транспонированной свёртки с размером ядра 2×2 и шагом 2 элемента (для восстановления размера изображения до начального). Слои транспонированной свёртки дополняются картами признаков, соответствующими им по размеру. За этими слоями следует слой свёртки с размером ядра 1×1 и с шагом 1. Такая архитектура СНС приведена в таблице 2.

Таблица 2 – Архитектура СНС ULeNet

№	Тип слоя	Число ядер свёртки	Размеры ядра/шаг	Размеры входного изображения/карт признаков
1	Свёрточный	6	7×7/1	1500×1500×3
2	Подвыборки	–	2×2/2	1500×1500×6
3	Свёрточный	32	5×5/1	750×750×32
4	Подвыборки	–	2×2/2	375×375×32
5	Свёрточный	100	5×5/1	375×375×32
6	Транспонированной свёртки	32	2×2/2	375×375×100
7	Объединение с выходом 3 слоя	–	–	750×750×32 750×750×32
8	Свёрточный	32	5×5/1	750×750×64
9	Транспонированной свёртки	6	2×2/2	750×750×32
10	Объединение с выходом 1 слоя	–	–	1500×1500×6 1500×1500×6
11	Свёрточный	2	1×1/1	1500×1500×12
12	Softmax	2	–	1500×1500×2

В работе [6] рассмотрена СНС DeconvNet, архитектура которой построена с использованием энкодера и декодера, восстанавливающего размер изображения после энкодера до входного.

В разработанной нами архитектуре используется идея из [6], причём предлагается в качестве энкодера модифицированные первые пять слоёв СНС LeNet5 (рис. 1). Далее для восстановления размера изображения до начального, применяются два слоя транспонированной свёртки с размером ядра 2×2 и шагом 2 элемента (табл. 3). После каждого слоя транспонированной свёртки следует свёрточный слой. Такая архитектура СНС получила название DeconvLeNet.

Таблица 3 – Архитектура СНС DeconvLeNet

№	Тип слоя	Число ядер свёртки	Размеры ядра/шаг	Размеры входного изображения/ карт признаков
1	Свёрточный	6	7×7/1	1500×1500×3
2	Подвыборки	–	2×2/2	1500×1500×6
3	Свёрточный	32	5×5/1	750×750×32
4	Подвыборки	–	2×2/2	375×375×32
5	Свёрточный	100	5×5/1	375×375×32
6	Транспонированной свёртки	32	2×2/2	375×375×100
7	Свёрточный	16	5×5/1	750×750×32
8	Транспонированной свёртки	6	2×2/2	750×750×16
9	Свёрточный	2	1×1/1	1500×1500×6
10	Softmax	2	–	1500×1500×2

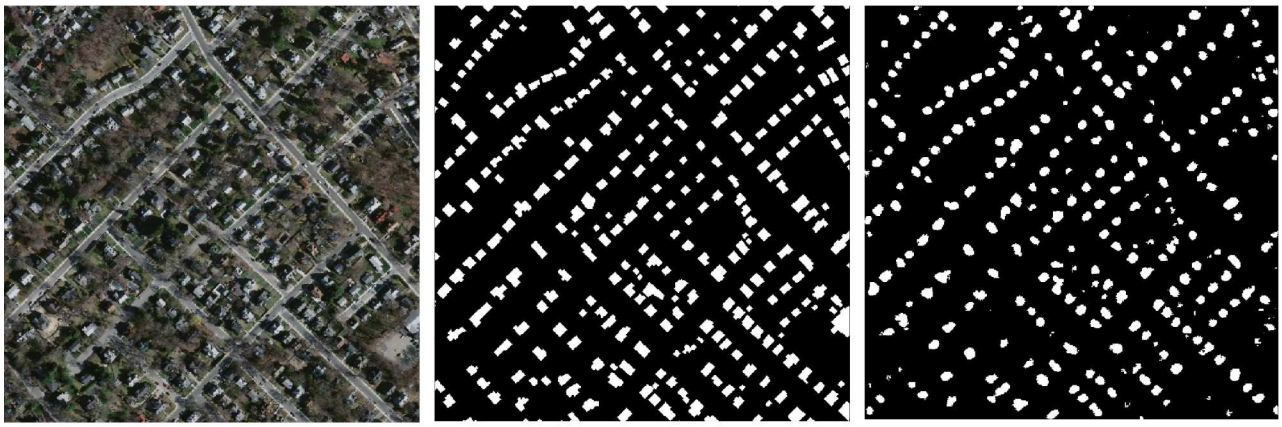
На выходе каждой из трех предложенных архитектур СНС используется операция Softmax. Выход содержит две карты признаков, которые соответствуют двум классам (фон, здание).

Исследование эффективности предложенных архитектур СНС

Проведем исследование предложенных для сегментации изображений трёх архитектур СНС подкласса LeNet5. Программная реализация этих предложенных архитектур СНС осуществлялась на языке программирования Python версии 3.6 с использованием библиотеки PyTorch версии 1.0.0.

Обучение каждой из этих архитектур СНС выполнялось на выборке Massachusetts Road Dataset спутниковых снимков с использованием оптимизатора Adam со следующими параметрами: betas 0,9; 0,9999. Регуляризирующий параметр weightdecay равен 0,00001. Параметр скорости обучения равен 0,00001. Данный параметр уменьшался в 0,99 раз каждые 100 эпох обучения. Обучение проводилось на 8000 эпохах для каждой из трёх предложенных архитектур СНС. Размер батча равен 16 обучающим примерам. Для обучения из каждого изображения обучающей выборки произвольным образом выбиралась подобласть размером 800×800 пикселей. В качестве функции потерь (loss function) использовалась бинарная кросс-энтропия (англ. binary cross entropy). Обучение производилось с использованием видеокарты Nvidia GTX 1080Ti.

Результаты тестирования обученных СНС сравнивались с картами сегментированных объектов, соответствующих исходным спутниковым снимкам и включенных также в выборку Massachusetts Road Datasets. На рисунке 2,*а* приведен пример исходного изображения. На рисунке 2,*в* показаны результаты сегментации этого изображения с помощью СНС архитектуры SegLeNet. На рисунке 2,*б* для сравнения с рисунком 2,*в* показана карта сегментированных объектов, соответствующая исходному изображению. Из рисунка 2,*в* видим, что границы довольно большого числа зданий не совпадают с границами зданий на карте сегментированных объектов на рисунке 2,*б*.



а *б* *в*

Рисунок 2 – Пример результатов сегментации изображения участка земной поверхности:
а – исходное изображение; *б* – карта сегментированных объектов из выборки Massachusetts Road Datasets;
в – результаты сегментации изображения с помощью SegLeNet

Измерение точности сегментации производились путем расчета коэффициента Jaccard (он же Intersectionoverunion) из пакета sklearn. Результаты по точности сегментации и времени, необходимого на анализ одного изображения размером 1500×1500, пикселей представлены в таблице 4.

Таблица 4 – Результаты исследования предложенных архитектур СНС

Наименование СНС	SegLeNet	ULeNet	DeconvLeNet
Jaccard, %	86,92	88,40	88,36
Время анализа одного изображения (мс)	232,61	198,28	159,88

Анализируя таблицу 4, можно сказать, что худшие результаты по точности сегментации изображений и времени, необходимого для анализа одного спутникового снимка, показывает СНС SegLeNet. По точности сегментации наилучший результат дает СНС ULeNet, немного от неё отстает нейросеть DeconvLeNet. Однако по времени на анализ одного изображения СНС DeconvLeNet значительно опережает другие исследуемые архитектуры СНС. По совокупности этих двух показателей предпочтение для дальнейших комплексных исследований эффективности стоит отдать СНС DeconvLeNet.

Заключение

В последние годы актуальным направлением для анализа изображений земной поверхности является использование СНС. Для решения задачи семантической сегментации таких изображений нами предложены три новые архитектуры СНС, относящиеся к подклассу LeNet5. В основе каждой из этих архитектур лежит энкодер, представляющий из себя модифицированные пять первых слоёв разработанной ранее нами СНС подкласса LeNet5.

Проведено исследование эффективности предложенных архитектур СНС, обученных на выборке Massachusetts Road Dataset. Анализ результатов этих исследований показал, что по совокупности двух показателей (точность сегментации изображений и время, необходимое для анализа одного спутникового снимка) предпочтение следует отдать СНС DeconvLeNet.

Исследования были поддержаны грантом РФФИ № 18-47-700010 р_а.

Литература:

1. Le Cun, Y. Gradient-Based Learning Appelied to Document Recognition / Y. Le Cun, L. Bottou, Y. Bengio, P. Haffner // Proc of the IEEE. – 1998. – Vol. 86, Issue 11. – P. 2278–2324. – DOI: 10.1109/5.726791

2. Зоев И.В. Устройство на основе ПЛИС для распознавания рукописных цифр на изображениях / И.В. Зоев, А.П. Береснев, Н.Г. Марков, А.Н. Мальчуков // Компьютерная оптика. – 2017. – Т. 41. – № 6. – С. 938–949. – DOI: 10.18287/2412-6179-2017-41-6-938-949.

3. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. – URL : https://arxiv.org/abs/1506.01497 (дата обращения 22.03.2019).

4. Road and Building Detection Datasets. – URL : http://www.cs.toronto.edu/~vmnih/data/ (дата обращения 22.03.2019).

5. U-Net: Convolutional Networks for Biomedical Image Segmentation. – URL : https://arxiv.org/abs/1505.04597 (дата обращения 22.03.2019).

6. Learning Deconvolution Network for Semantic Segmentation. – URL : https://arxiv.org/pdf/1505.04366.pdf (дата обращения 22.03.2019).

References:

1. Le Cun, Y. Gradient-Based Learning Appelied to Document Recognition / Y. Le Cun, L. Bottou, Y. Bengio, P. Haffner // Proc of the IEEE. – 1998. – Vol. 86, Issue 11. – P. 2278–2324. – DOI: 10.1109/5.726791

2. Zoev I.V., Beresnev A.P., Markov N.G., Malchukov A.N. Fpga-based device for handwritten digit recognition in images // Computer Optics. – 2017. – Vol. 41. – № 6. – P. 938–949. DOI: 10.18287/2412-6179-2017-41-6-938-949.

3. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal. – URL : https://arxiv.org/abs/1506.01497 (date of access 22.03.2019).

4. Road and Building Detection Datasets. – URL : http://www.cs.toronto.edu/~vmnih/data/ (date of access 22.03.2019).

5. U-Net: Convolutional Networks for Biomedical Image Segmentation. – URL : https://arxiv.org/abs/1505.04597 (date of access 22.03.2019).

6. Learning Deconvolution Network for Semantic Segmentation. – URL : https://arxiv.org/pdf/1505.04366.pdf (date of access 22.03.2019).