

Автоматическое обнаружение строений с помощью алгоритмов бинарной семантической сегментации

Е.А. Дмитриев¹, А.А. Бородинов¹, А.И. Максимов¹

¹Самарский национальный исследовательский университет им. академика С.П. Королева, Московское шоссе 34А, Самара, Россия, 443086

Аннотация. В статье рассматриваются алгоритмы бинарной семантической сегментации для автоматического обнаружения строений на изображениях дистанционного зондирования Земли. На основе сравнительного анализа глубоких свёрточных сетей проводятся исследования для нахождения оптимальной модели сети с точки зрения скорости обучения и точности сегментации. Обучение и тестирование проводилось на изображениях Московской области, полученных из открытого картографического источника. В результате экспериментов была найдена оптимальная архитектура нейронной сети для автоматического обнаружения строений.

1. Введение

Задача автоматического обнаружения объектов на изображениях дистанционного зондирования Земли (ДЗЗ) является одной из самых сложных задач. Примером решения рассматриваемой задачи может послужить [1]. В настоящее время одним из самых эффективных подходов является использование алгоритмов семантической сегментации. Другими словами, для каждого пикселя изображения определяется класс объекта, которому он принадлежит.

Сегментация снимков ДЗЗ находит применение во многих отраслях: геоинформатика, создание карт местности, анализ землепользования и т.д. На данный момент многие этапы процесса сегментирования решаются с помощью операторов вручную, что приводит к большим затратам экономических в временных ресурсов, а также к некоторым неточностям в разметке из-за человеческого фактора.

В настоящее время существует множество алгоритмов для сегментирования изображений [2, 3, 4], однако самыми эффективными являются подходы с использованием свёрточных нейронных сетей (СНН) [5]. Почти для всех задач компьютерного зрения свёрточные сети позволяют получить более эффективные результаты по сравнению с другими алгоритмами.

За последние годы были предложены различные подходы по формированию моделей СНН, которые на выходе дают карту сегментации исходного изображения. Один из самых эффективных методов основан на использовании полносвёрточных нейронных сетей [5]. В отличие от свёрточных сетей, которые используются для классификации, в полносвёрточных сетях отсутствует подсеть многослойного перцептрона для классификации.

Архитектуру свёрточных сетей для семантической сегментации можно условно разделить на две части: кодер и декодер. Кодер на выходе выдаёт карты признаков с меньшим размером по сравнению с входным изображением. Для восстановления размера карт признаков используется декодер. В первоначальных версиях моделей полносвёрточных сетей декодер представлял

собой геометрическое преобразование для увеличения размера изображений с различными способами интерполяции [5]. В настоящее время используется подход, где подсеть декодера строится симметрично подсети кодера за исключением слоёв прореживания. Вместо слоёв прореживания могут использоваться транспонированные свёрточные слои [6] или слои с операцией обратной прореживанию [7].

В работе рассматриваются 4 свёрточные сети для обнаружения строений с различными архитектурами кодера и декодера. В качестве критериев эффективности алгоритмов используются время обучения сети и точность сегментации.

Работа выстроена в следующем порядке. Во втором разделе работы даётся описание рассматриваемых архитектур нейронных сетей. В третьем разделе представлены результаты экспериментальных исследований на реальных снимках Московской области. В заключительном разделе подведены итоги экспериментальных исследований и сказано про дальнейшее направление исследований в области алгоритмов семантической сегментации.

2. Описание алгоритмов

В качестве алгоритмов для бинарной семантической сегментации использовались нейронные сети SegNet [7], модель с кодером от сети ResNet-50 [8] и с декодером в виде геометрического преобразования с билинейной интерполяцией, U-net [9], LinkNet [6].

Модель сети SegNet представляет собой классическую архитектуру кодер-декодер. Кодер SegNet состоит из слоёв сети VGG-16 [10]. Архитектура декодера почти симметрична подсети кодера, за исключением слоёв прореживания. Вместо слоёв прореживания используются слои с операцией обратной прореживанию. Модель сети SegNet представлена на рис 1.

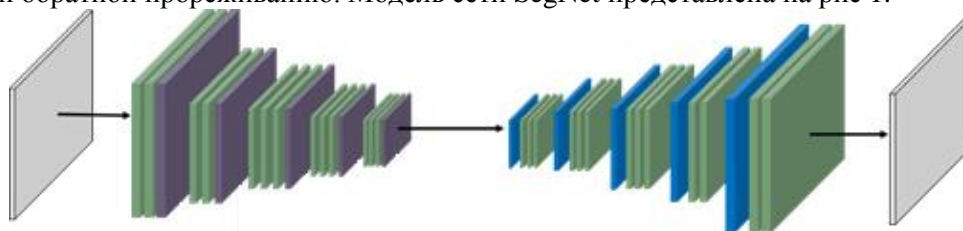


Рисунок 1. Модель сети SegNet.

Также в работе рассматривалась свёрточная нейронная сеть для сегментации с кодером на базе ResNet-50. Особенность сети ResNet-50 заключается в использовании остаточных соединений, которые позволяют эффективно решать проблему затухающего градиента, возникающую с увеличением количества слоёв нейронной сети. Модель сети представлена на рисунке 2.

Следующая рассматриваемая архитектура нейронной сети – U-net. Особенность модели U-net заключается в конкатенации карт признаков на нижних и верхних уровнях нейронной сети. Данный подход очень похож на остаточные соединения в сети ResNet-50, но в случае U-net используются более глубокие соединения. Модель сети представлена на рисунке 3.

Развитием модели U-net является сеть LinkNet. Кодер и декодер сети разбиты на несколько подблоков. LinkNet требует меньше вычислительных ресурсов по сравнению с рассмотренными моделями в силу быстрого уменьшения размера карт признаков. На входе сети уменьшение карт признаков происходит за счёт прореживания и свёртки с шагом равным 2, а в блоке кодера за счёт свёртки вместо прореживания. В декодере для восстановления размера изображений используются транспонированные свёрточные слои. Модель сети представлена на рисунке 4.

В качестве функции потерь выступала перекрёстная энтропия. Согласно [11], в задаче классификации использование перекрёстной энтропии в качестве функции потерь позволяет достичь лучшего локального минимума с точки зрения точности классификации при случайной инициализации параметров алгоритма по сравнению со средним квадратичным отклонением.

Пусть $I(n_1, n_2, n_3)$ – цифровое изображение, подаваемое на вход нейронной сети, где $(n_1, n_2, n_3) \in \mathbf{D}$, $\mathbf{D} = \{(n_1, n_2, n_3) : n_1 = \overline{0, N_1 - 1}, n_2 = \overline{0, N_2 - 1}, n_3 = \overline{0, N_3 - 1}\}$, N_1, N_2 – размер

изображений, а N_3 – количество каналов у входного изображения. Пусть $Y(n_1, n_2, n_3)$ – маска истинной сегментации, размеры которой совпадают с входным изображением, а количество каналов совпадает с количеством классов.

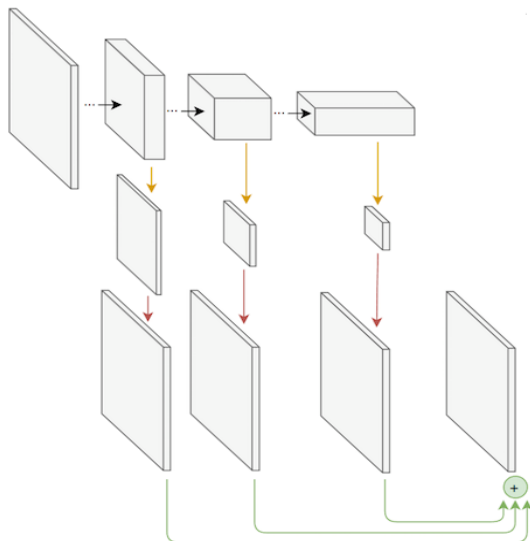


Рисунок 2. Модель полносвёрточной сети на базе ResNet-50.

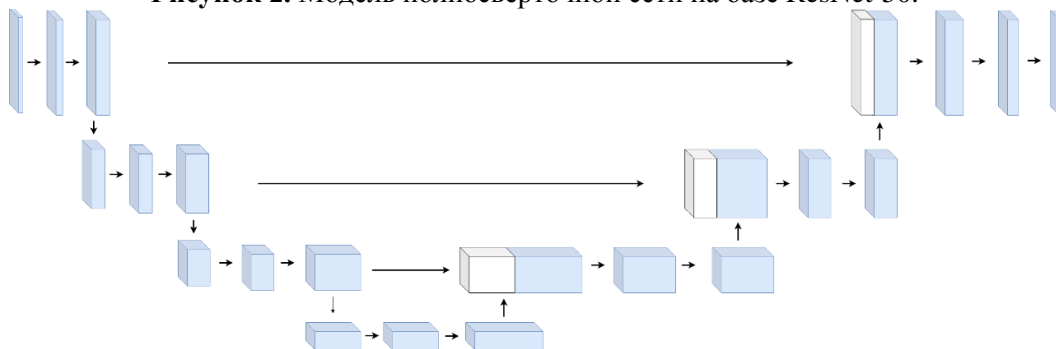


Рисунок 3. Модель сети U-net.

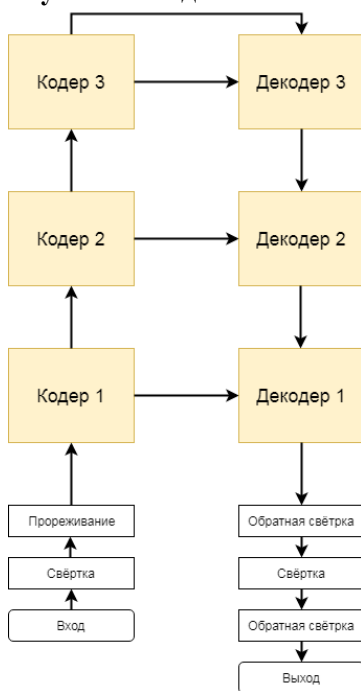


Рисунок 4. Модель сети LinkNet.

Каждому каналу соответствовал определенный класс. В качестве классов выступали строения и фон. Значения $Y(n_1, n_2, n_3)$ в каналах равнялись 0 или 1 в зависимости от класса пикселя на входном изображении. Пусть $O(n_1, n_2, n_3)$ – изображение, полученное на выходе нейронной сети, размеры которого и количество каналов совпадают с размеченным изображением. Пусть $y(n_3), o(n_3)$ – пиксели с одинаковыми позициями на размеченном и выходном изображениях. Тогда функция потерь выглядит следующим образом:

$$H(y, o) = - \sum_{i=0}^{N_3-1} y(i) \log o(i). \quad (1)$$

В качестве целевой функции выступал функционал средней ошибки нейронной сети на обучающей выборке. Пусть X^G – множество с обучающими изображениями, где G – количество элементов, а w – веса нейронной сети. Тогда функционал средней ошибки выглядит следующим образом:

$$Q(w, X^G) = \frac{1}{G} \sum_{i=0}^{G-1} \sum_{j=0}^{N_1-1} \sum_{k=0}^{N_2-1} H(O(i, j), Y(i, j)) \quad (2)$$

Все модели обучались с помощью алгоритма адаптивного стохастического градиента [12]. Во время обучения сети использовалась техника уменьшения коэффициента обучения в том случае, если значение функционала качества сети на валидационной выборке не увеличивалось.

3. Результаты экспериментов

В работе рассматривались снимки населённых пунктов Московской области [13]. На вход сети подавались RGB изображения размером 512×512 . Количество снимков составляло 3323. Отношение количества элементов в обучающей выборке к количеству элементов тестовой выборки составляло 80:20. В роли классов выступали строения и фон. Пример снимка и маски представлен на рисунке 5.



Рисунок 5. Пример снимка и маски части населённого пункта Московской области.

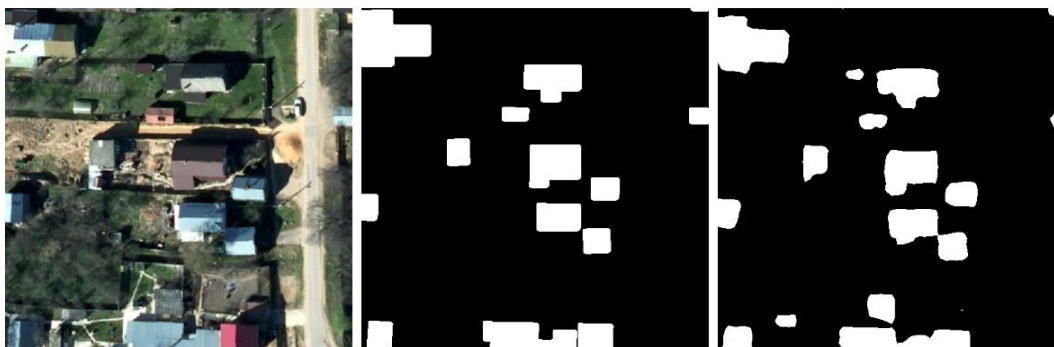
Как видно из рисунка 5, были случаи, когда маска не полностью соответствовала входному изображению. Несмотря на это, включение таких изображений в обучающую выборку позволило увеличить значение используемой метрики на тестовых изображениях с идеальной маской даже без этапа предобработки.

В качестве метрики выступала точность сегментации. Точность сегментации соответствует проценту правильно классифицированных пикселей от общего числа пикселей. Обучение моделей проводилось с помощью видеокарты Nvidia GTX 1080 Ti. Результаты экспериментов представлены в таблице 1.

По результатам экспериментов можно сделать вывод, что подходы с использованием транспонированных слоёв в декодере, используемых в сети LinkNet, и конкатенация верхних и нижних карт признаков, которая используется в сети LinkNet и U-net, позволяют получить более высокие обобщающие способности по сравнению с другими рассматриваемыми архитектурами. Пример выходного изображения сети LinkNet представлен на рисунке 6.

Таблица 1. Результаты рассматриваемых нейронных сетей.

Модель сети	Время обучения, ч	Точность сегментации, %
SegNet	4	96,7
Сеть на базе Res-net 50	1,3	96,2
U-net	2	96,9
LinkNet	0,5	97,2

**Рисунок 6.** Пример тестового изображения, разметки и маски, полученной с помощью сети Linknet.

4. Заключение

В данной статье были исследованы различные архитектуры свёрточных нейронных сетей для обнаружения строений на ДЗЗ.

Была проведена серия экспериментов, в ходе которой выявлена оптимальная архитектура нейронной сети с точки зрения времени обучения и точности сегментации. В дальнейшем планируется исследование по использованию условных случайных полей для повышения качества сегментации.

5. Литература

- [1] Myasnikov, V.V. Method for detection of vehicles in digital aerial and space remote sensed images / V.V. Myasnikov // *Computer Optics*. – 2012. – Vol. 36(3). – P. 429-438.
- [2] Кузнецов, А.В. Сравнение алгоритмов управляемой поэлементной классификации гиперспектральных изображений / А.В. Кузнецов, В.В. Мясников // *Компьютерная оптика*. – 2014. – Т. 38, № 3. – С. 494-502.
- [3] Блохинов, Ю.Б. Разработка алгоритма семантической сегментации аэрофотоснимков реального времени / Ю.Б. Блохинов, В.А. Горбачев, Ю.О. Ракутин, А.Д. Никитин // *Компьютерная оптика*. – 2018. – Т. 42, № 1. – С. 141-148. DOI: 10.18287/2412-6179-2018-42-1-141-148.
- [4] Cortes, C. Support-Vector Networks / C. Cortes, V. Vapnik // *Machine Learning*. – 1995. – Vol. 20(3). – P. 273-297.
- [5] Shelhamer, E. Fully convolutional networks for semantic segmentation / E. Shelhamer, J. Long, T. Darrell // *The Pattern Analysis and Machine Intelligence*. – 2016. – Vol. 324(5). – P. 100-108.
- [6] Chaurasia, A. Linknet: Exploiting encoder representations for efficient semantic segmentation / A. Chaurasia, E. Culurciello // *IEEE Conference on Computer Vision and Pattern Recognition*. – 2017. – Vol. 362(3). – P. 234-247.
- [7] Badrinarayanan, V. Segnet: A deep convolutional encoder-decoder architecture for image segmentation / V. Badrinarayanan, A. Kendall, R. Cipolla // *IEEE Conference on Computer Vision and Pattern Recognition*. – 2015. – Vol. 353(4). – P. 125-145.
- [8] He, K. Deep residual learning for image recognition / K. He, X. Zhang, S. Ren, J. Sun // *IEEE Conference on Computer Vision and Pattern Recognition*. – 2016. – Vol. 123(2) – P. 235-247.

- [9] Ronneberger, O. U-net: Convolutional networks for biomedical image segmentation / O. Ronneberger, P. Fischer, T. Brox // Medical Image Computing and Computer-Assisted. – 2015. – Vol. 345(1). – P. 234-241.
- [10] Russakovsky, O. ImageNet large scale visual recognition challenge / O. Russakovsky, J. Deng, H. Su, J. Krause // IEEE Conference on Computer Vision and Pattern Recognition. – 2014. – Vol. 243(2). – P. 121-136.
- [11] Golik, P. Cross-entropy vs. squared error training: a theoretical and experimental comparison / P. Golik, P. Doetsch, H. Ney // The 14th Annual Conference of the International Speech Communication Association. – 2019. – Vol. 234(4). – P. 100-105.
- [12] Diederik, P. Adam: A Method for Stochastic Optimization / P. Deiderik // The Pattern Analysis and Machine Intelligence. – 2014. – Vol. 723(3). – P. 624-637.
- [13] Региональная географическая информационная система для обеспечения деятельности центральных исполнительных органов государственной власти Московской области, государственных органов Московской области, органов местного самоуправления муниципальных образований Московской области [Электронный ресурс]. – Режим доступа: <https://rgis.mosreg.ru>.

Благодарности

Исследование выполнено при финансовой поддержке РФФИ в рамках научных проектов № 18-01-00748-а.

Automatic detection of constructions using binary image segmentation algorithms

Е.А. Dmitriev¹, А.А. Borodinov¹, А.И. Maksimov¹

¹Samara National Research University, Moskovskoe Shosse 34A, Samara, Russia, 443086

Abstract. This article presents binary segmentation algorithms for buildings automatic detection on aerial images. There were conducted experiments among deep neural networks to find the most effective model in sense of segmentation accuracy and training time. All experiments were conducted on Moscow region images that were got from open database. As the result the optimal model was found for buildings automatic detection.