

УДК 004.932

DOI: 10.46548/21vek-2020-0950-0020

ИНТЕЛЛЕКТУАЛЬНОЕ СЕГМЕНТИРОВАНИЕ СПУТНИКОВЫХ СНИМКОВ С ИСПОЛЬЗОВАНИЕМ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ

©2019

Акинина Александра Владимировна, аспирант

Кузнецов Владимир Владимирович, аспирант

Никифоров Михаил Борисович, кандидат технических наук, доцент кафедры ЭВМ

Рязанский государственный радиотехнический университет

(390005, Россия, г. Рязань, ул. Гагарина, 59/1,

e-mails: alexandra.sokolova00@mail.ru, vla8774@gmail.com, nikiforov.m.b@evm.rsreu.ru)

Аннотация. Для решения задач семантической сегментации, как правило, используются сверточные нейронные сети. Известно, что главным недостатком таких сетей является время работы, поэтому в статье была сделана попытка сократить время работы алгоритма, не теряя при этом в качестве. В статье приведено сравнение трех видов алгоритмов Unet, Enet и Segnet для выполнения семантической сегментации. Был сделан вывод о применении сети Unet с предварительным использованием вегетативного индекса. В архитектуре сети Unet блоки свертки заменены блоками пакетной свертки для ускорения работы сети. В качестве нелинейного слоя используется блок relu. Результаты сегментации оцениваются с помощью метрики Intersection Over Union. Рассмотренный алгоритм показал хорошие временные и качественные характеристики.

Ключевые слова: сверточные нейронные сети, семантическая сегментация, пакетная свертка, unet, enet, segnet, relu.

INTELLIGENT SEGMENTATION OF SATELLITE IMAGES USING CONVOLUTIONAL NEURAL NETWORKS

©2019

Akinina Alexandra Vladimirovna, postgraduate student

Kuznetsov Vladimir Vladimirovich, postgraduate student

Nikiforov Mikhail Borisovich, PhD, associate professor

Ryazan State Radioengineering University

(390005, Russia, Ryazan, Gagarina st., 59/1,

e-mails: alexandra.sokolova00@mail.ru, vla8774@gmail.com, nikiforov.m.b@evm.rsreu.ru)

Abstract. Convolutional neural networks are usually used to solve the problems of semantic segmentation. It is known that the problem is that working on this algorithm does not require any effort. Three types of algorithms Unet, Enet and Segnet are compared to perform semantic segmentation. At the first stage of the algorithm, using the vegetative index, vegetation is distinguished. Box Convolution has replaced convolution blocks to speed up the network. The relu block is used as a non-linear layer. Segmentation results are evaluated using the Intersection Over Union metric. The considered algorithm showed good temporal and qualitative characteristics.

Keywords: convolution neuron network, semantic segmentation, box convolution, unet, enet, segnet, relu.

Введение. За последние годы были получены впечатляющие результаты в решении задач сегментации изображений различными методами, такими как метод нормальных разрезов [1], применение EM-алгоритмов [2], карты Кохонена [3], использование нейронных сетей [4-7]. Особенно хорошие результаты показывают сверточные нейронные сети (CNN) [8]. Однако эти достижения в части точности не обязательно делают сети более эффективными с точки зрения размера и скорости. Во многих реальных приложениях, таких как робототехника, автомобили с автоматическим управлением и дополненная реальность, задачи распознавания образов должны выполняться своевременно на платформе с ограниченными вычислительными возможностями.

Материалы и результаты исследований. Есть набор спутниковых снимков высокого разрешения (B) определенной местности, на которых необходимо выделить области пикселей, относящихся к классам водных объектов, растительности, построек, дорог.

В общем случае задача сегментации сводится к построению разбиения:

$$B = \bigcup B_m^{segm}, m = \overline{1, M},$$

где B_m^{segm} – сегмент такой, что $B_m^{segm} \cap B_n^{segm} = \emptyset$, $\forall m, n \in [1, M], m \neq n, \forall B_m^{segm}$ – связанное подмножество B.

Исходные данные. Данный алгоритм был разработан для спутниковых снимков, обладающих следующими характеристиками:

1) достаточное разрешение, как правило, таким разрешением является 30...60 метров на пиксель;

2) необходимость наличия следующих спектральных каналов:

- синего (B-(англ.) *blue*, длина волны около 450...515 нм),

- зеленого (G-(англ.) *green*, длина волны около 525...605 нм),

- красного (R-(англ.) *red*, длина волны около 630...690 нм),

-нескольких ближних инфракрасных каналов (*NIR*- (англ.) *near-infrared*, длин волны около 750...1400 нм).

К таким спутниковым снимкам относятся данные, полученные от космического аппарата *Landsat 5* и *Landsat 8*.

Существующие методы. Для решения задач семантической сегментации в настоящее время используются сверточные нейронные сети. Для сравнения были выбраны самые распространенные, показавшие высокие результаты на соревнованиях *kaggle*. Реализации моделей этих сетей была опробована с использованием *pyTorch*.

U-Net (2015 год) [9], архитектура которого состоит из сжимающихся (для захвата контекста) и расширяющихся (для точной локализации) направлений. К плюсам этого алгоритма можно отнести использование достаточно небольшой обучающей выборки (и меньших затрат по времени) для достижения хорошего результата.

SegNet (2017 год) [10] отличается от обычной сверточной сети в том, что слои апсэмплинга информационно соединены с пулинг слоями, то есть апсэмплинг слои не обучаются, а лишь получают информацию, как восстановить сжатые изображения от пулинг слоев, которые хранят индексы активиро-

ванных (с наибольшим значением в окне) пикселей.

Enet (2016 год) [11] состоит из блоков *bottleneck*. В отличие от других сетей здесь декодер меньше декодера в связи с тем, что задача выделения признаков более сложная, чем восстановление маски из хороших признаков. Эта модель содержит меньше параметров, что отражается на ее скорости работы.

Реализации моделей этих сетей была опробована с использованием *pyTorch*.

Алгоритм. Для существующих подходов разбиение сразу на несколько классов сильно усложняет работу алгоритма, поэтому было принято решение о первоначальном выделении растительности на спутниковом снимке с помощью вегетационного индекса, который имеет фиксированные значения для растительных объектов:

$$ARVI = \frac{NIR - Rb}{NIR + Rb}, Rb = RED - 0.5 * (RED - BLUE),$$

NIR – отражение в ближней инфракрасной области спектра, *RED* - в красной области, *BLUE* - в голубой.

Спутниковый снимок с выделенной растительностью поступает на вход предварительно обученной нейронной сети, архитектура которой представлена на рисунке 1.

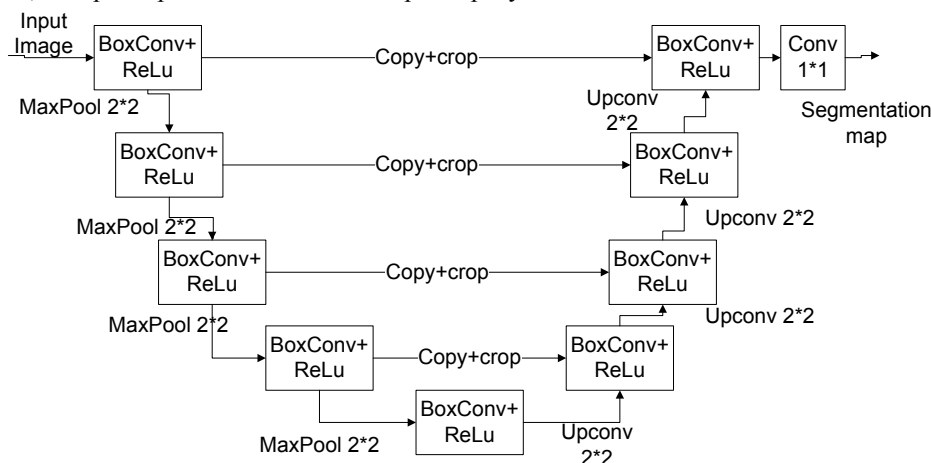


Рисунок 1 - Архитектура нейронной сети

Вместо стандартной свертки применяется пакетная свертка (англ. *Box Convolution*) [12], что значительно сокращает время работы сети. Пакетная свертка основана на «глубинной» свертке (англ. *depthwise convolution*) [13,14] со специальным ядром свертки, которая по некоторому прямоугольнику берет среднее значение. «Глубинная» свертка отличается от обычной свертки тем, что каналы обрабатываются по отдельности, независимо друг от друга. Среднее значение в прямоугольнике вычисляется с помощью интегральной суммы изображения (алгоритм Виолы-Джонса [15]). Для обучения такого прямоугольника нужно всего 4 параметра $x_{min}, x_{max}, y_{min}, y_{max}$, тогда как для обычной свертки 3*3 таких параметров 9. Причем разница между максимальным и минимальными значениями параметров больше одного пикселя $x_{max} - x_{min} > lpxl, y_{max} - y_{min} > lpyl$ (). Параметры инициализируются случайным образом, в процессе

обучения прямоугольник меняет положение, ширину, высоту. Ограничения, наложенные на параметры, гарантируются проективным градиентным спуском при использовании *L2*-регуляризации [16].

В качестве нелинейного слоя используется слой *relu*, благодаря которому обучение большой сети происходит за разумное время. Достоинство этого слоя заключается в исключении «насыщенных» нейронов, которые ведут к затуханию градиента.

К входящему изображению применяется *box*-свертка, затем результат поступает на блок *ReLu* и производится операция максимального объединения для понижения разрешения (*maxpool*). Операции *maxpool* позволяют выделять важные особенности на изображении. В отличие от операций свертки на этом этапе происходит лишь выбор максимального значения из заданного окна. Каналы свойств удваиваются на каждом этапе понижающей дискретизации. Карта

свойств в сужающемся направлении получается обрезанной за счет потери граничных пикселей.

В ходе разворачивания изображения происходят операции, повышающие дискретизацию карты свойств (стрелки *upconv*). За каждой такой операцией следует объединение с соответствующим образом из обрезанной карты свойств (стрелки *copy+crop*), *box*-свертка (уменьшается количество матриц свойств) с нелинейным блоком *relu*.

На последнем слое необходима свертка 1×1 для сопоставления каждого вектора свойств с желаемым количеством классов.

Для обучения сети используется метод градиентного спуска.

Граница разделения на классы вычисляется с использованием морфологических операций. Весовые коэффициенты вычисляются по формуле:

$$\omega(x) = \omega_c(x) + \omega_0 \cdot \exp\left(-\frac{(d_1(x) + d_2(x))^2}{2\sigma^2}\right),$$

где ω_c – карта весов для балансировки частот классов, d_1 – расстояние до границы ближайшей ячейки, d_2 – расстояние до границы второй ближайшей ячейки.

Эксперимент. Для оценки результатов семантической сегментации была использована метрика *Intersection Over Union (IOU)*:

$$IOU = \frac{1}{k} \sum_i \frac{n_{ii}}{t_i - n_{ii} + \sum_j n_{ji}},$$

где k – количество классов, t_i число пикселей в классе, n_{ii} – число верно классифицированных пикселей, n_{ji} – число неверно классифицированных пикселей.

Было произведено экспериментальное сравнение работы существующих систем, результаты сравнения представлены в таблицах 1 и 2. На рисунке 2 изображены исходный спутниковый снимок и результат работы алгоритма.

Экспериментальное исследование осуществлено на вычислительной системе, имеющей следующие характеристики:

- аппаратная часть: *Intel Core i5-4460X* (частота процессора: 3,2 ГГц; 4 вычислительных ядра), *Nvidia GeForce GTX 950*, 8 гигабайт *DDR3* ОЗУ;
- программная часть: операционная система *Linux (Ubuntu)*.

Таблица 1 – Метрики моделей

Модель	Среднее IOU	IOU вода	IOU растительность	IOU дороги	IOU постройки
UNet	0,795	0,91	0,76	0,62	0,89
SegNet	0,78	0,9	0,73	0,6	0,88
ENet	0,792	0,89	0,78	0,61	0,89
NewUnet	0,87	0,91	0,87	0,8	0,9

Таблица 2 – Время обучения и работы в секундах

Модель	Время обучения, с	Время работы, с
UNet	8,0	0,14
ENet	4,2	0,03
SegNet	6,7	0,1
NewUnet	4,1	0,01



Рисунок 2 – Исходный спутниковый снимок и результат работы алгоритма

Заключение. Рассмотренный алгоритм показал лучшие по сравнению с известными алгоритмами временные и качественные характеристики. В дальнейшем планируется работа по изменению внутренней структуры нейронной сети для улучшения качества работы алгоритма.

СПИСОК ЛИТЕРАТУРЫ:

1. Akinina A.V., Nikiforov M.B., Savin A.V. Multiscale Image Segmentation using Normalized Cuts in Image Recognition on Satellite Images // 7 Mediterranean Conference on Embedded Computing (MECO-2018). - Montenegro, Budva. - 2018. - pp. 312-314.
2. Akinin M.V., Akinina A.V., Tarasov A.S., Sokolov A.V. Application of EM algorithm in problems of pattern recognition on satellite images // 6 Mediterranean Conference on Embedded Computing (MECO-2017). - Montenegro, Bar. - 2017. - pp. 175-178.
3. Sokolova (Akinina A.V.), Akinin M.V., Nikiforov M.B., Taganov A.I. Image segmentation algorithm based on self-organized Kohonen's neural maps and tree pyramidal segmenter // 4rd Mediterranean Conference on Embedded Computing (MECO-2015). - Montenegro, Budva. - 2015. - pp. 168 - 170.
4. Акинина Н.В., Акинин М.В., Соколова(Акинина) А.В., Никифоров М.Б., Таганов А.И. Автоэнкодер: подход к понижению размерности векторного пространства с контролируемой потерей информации // Известия ТулГУ. Технические науки. 2016. №9. – С. 3-12.
5. Акинина Н.В., Акинин М.В., Соколова(Акинина) А.В., Никифоров М.Б., Колесенков А.Н. Методы и алгоритмы фильтрации разнородных помех с применением систем искусственного интеллекта в задачах обработки данных дистанционного зондирования Земли // Известия ТулГУ. Технические науки. 2015. – С. 12-19.
6. Кузнецов В.В. Интеллектуальный метод построения

двумерных карт местности, основанный на использовании ансамбля классификаторов // Наука и бизнес: пути развития – 2019 №8 (98). – С. 72-76.

7. Акинин М.В., Никифоров М.Б., Таганов А.И. Нейросетевые системы искусственного интеллекта в задачах обработки изображений. - М.: Горячая линия — Телеком. - 2015. - 154 с.

8. L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs. T-PAMI, 40(4):834–848, 2018.

9. O. Ronneberger, P. Fischer, T. Brox U-Net: Convolutional Networks for Biomedical Image Segmentation // 2015. — Vol. abs/1505.04597. —1505.04597.

10. V. Badrinarayanan, A. Kendall, R. Cipolla SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation // 2016. . — Vol. abs/1511.00561. – 1511.00561.

11. A. Paszke, A. Chaurasia, S. Kim, E. Culurciello ENet: A Deep Neural Network Architecture for Real-Time Semantic Segmentation// 2016. — Vol. abs/1606.02147. —1606.02147.

12. Burkov E., Lempitsky V. Deep neural networks with box convolutions //Advances in Neural Information Processing Systems. – 2018. – С. 6211-6221.

13. F. Chollet. Xception: Deep learning with depthwise separable convolutions. Proc. CVPR, pp.1251–1258, 2017.

14. F. Chollet. Xception: Deep learning with depthwise separable convolutions. Proc. CVPR, pp.1251–1258, 2017.

15. P. Viola, M. J. Jones, and D. Snow. Detecting pedestrians using patterns of motion and appearance. IJCV, 63(2):153–161, 2005.

16. L.-C. Chen, A. Schwing, A. Yuille, and R. Urtasun. Learning deep structured models. Proc. ICML, pp. 1785–1794, 2015.

Статья поступила в редакцию 28.03.2020

Статья принята к публикации 10.06.2020