

ИСПОЛЬЗОВАНИЕ МЕТОДОВ ГЛУБОКОГО ОБУЧЕНИЯ ДЛЯ СЕМАНТИЧЕСКОЙ СЕГМЕНТАЦИИ ИЗОБРАЖЕНИЙ

Н.Д. Никонов*, Т.В. Никонова**, О.Е. Рубаник**, Е.А. Корчевская ***

**Федеральное государственное автономное образовательное учреждение
высшего образования «Санкт-Петербургский национальный исследовательский
университет ИТМО» (Российская Федерация)*

***Учреждение образования «Витебский государственный технологический университет»*

**** Учреждение образования «Витебский государственный
университет имени П.М. Машерова»*

Актуальность рассматриваемого вопроса обусловлена тем, что семантическая сегментация на основе глубокого обучения (Deep learning – DL) применяется в сельском хозяйстве для решения различных задач, таких как составление карт землепользования, определение границ полей, распознавание растений, обнаружение болезней, выявление сорняков, классификация культур или почвенного покрова.

Цель статьи – описание методики, позволяющей значительно повысить эффективность и точность семантической сегментации по сравнению с традиционными методами.

Материал и методы. В качестве материала используется несколько наборов данных снимков со спутников Sentinel-2 и Landsat 8. Авторами применены модели UNet, DeeplabV3plus и их различные модификации, а также модель HRNetv2.

Результаты и их обсуждение. Из результатов, полученных при тестировании на наборе данных снимков со спутников, можно сделать вывод, что модели UNet, DeeplabV3plus и их различные модификации проявляют себя значительно хуже, чем модель HRNetv2. Представляется, что последняя модель показала лучшие результаты в эксперименте по причине того, что сохранение высокого разрешения исходной картинки на протяжении всей сети сильно влияет на точность границ.

Заключение. В работе продемонстрировано, что HRNetv2 с предложенными авторами модификациями имеет преимущества по сравнению с другими рассмотренными сетями. Она соединяет сети высокого и низкого разрешения параллельно, а не последовательно, обеспечивая представление высокого разрешения на протяжении всей сети. Также значительное влияние на итоговую точность оказало использование подхода расчета границ домов и маски домов одновременно, а также применение совместной ошибки по границам домов и маскам домов при обучении сети.

Ключевые слова: дистанционное зондирование, методы глубокого обучения, семантическая сегментация.

USING DEEP LEARNING TEACHING METHODS FOR SEMANTIC SEGMENTATION OF IMAGERY

N.D. Nikonov*, T.V. Nikonova**, O.E. Rubanik**, E.A. Korchevskaya ***

**Federal State Autonomous Higher Education Establishment*

“St. Petersburg National Research University ITMO” (Russian Federation)

***Education Establishment “Vitebsk State Technological University”*

****Education Establishment “Vitebsk State P.M. Masherov University”*

The relevance of the issue is due to the application of semantic segmentation on the basis of Deep learning in agriculture for solving different problems such as compiling land management maps, identifying limits of fields, identifying plants, finding out diseases, weeds, classifying crops and soil cover.

The purpose of the article is a description of the methodology which makes it possible to considerably increase the efficiency and accuracy of semantic segmentation compared to traditional methods.

Material and methods. A number of data sets of Sentinel-2 and Landsat 8 satellite pictures were used as the material. The authors used UNet, DeeplabV3plus models and their various modifications as well as HRNetv2 model.

Findings and their discussion. Findings obtained from testing a set of satellite data pictures makes it possible to conclude that UNet, DeeplabV3plus models and their modifications manifest themselves considerably worse than the HRNetv2 model. The latter

model is supposed to have manifested better results in the experiment due to the fact that preservation of high definition of the original picture all over the whole network influences greatly the precision of the limits.

Conclusion. The paper demonstrates that HRNetv2 with the modifications offered by the authors has advantages compared to other considered networks. It connects networks with high and low definition parallelly but not successively, thus providing high definition picture over the whole network. A considerable impact on the final precision was caused by the use of the approach of the calculation of the limits of houses and house masks at the same time as well as the application of the joint error along house limits and house masks in teaching the network.

Key words: remote sensing, Deep Learning methods, semantic segmentation.

Доступность высокодетальных изображений дистанционного зондирования и развитие методов глубокого обучения (Deep learning – DL) сместили парадигму классификации изображений от методов, основанных на пикселях и объектах, к семантической сегментации на основе глубокого обучения. Архитектуры нейронных сетей, такие как CNN (Convolutional Neural Network) [1] и FCN (Fully Convolutional Networks), значительно повысили производительность и точность вычислений при дистанционном зондировании. DL позволяет автоматически извлекать признаки из больших наборов данных с помощью сложных моделей, уменьшая ошибки классификации.

FCN, в частности, стала успешной архитектурой DL для сегментации городских изображений. Используя CNN в качестве экстрактора признаков и заменяя полностью связанные слои на сверточные слои [2], FCN создает плотные попиксельные выходные карты, что позволяет проводить обучение для задач сегментации. Этот подход показал заметное улучшение точности сегментации по сравнению с традиционными методами. Однако архитектура сети FCN часто модифицируется для решения конкретных задач, возникающих при сегментации городских спутниковых изображений.

В исследованиях в области дистанционного зондирования DL успешно применяется для решения различных задач, таких как составление карт землепользования, определение границ полей, распознавание растений, обнаружение болезней [3], выявление сорняков, классификация культур или почвенного покрова.

Цель статьи – описание методики, позволяющей значительно повысить эффективность и точность семантической сегментации по сравнению с традиционными методами.

Материал и методы. Данные, с которыми проводились эксперименты, представляют собой набор снимков со спутников Sentinel-2 и Landsat 8 [4; 5].

В дальнейшем модель дообучалась и тестировалась на официальном датасете ежегодных соревнований SpaceNet в задаче сегментации данных. В обеих моделях используются стандартные RGB каналы, а также NIR канал. Все снимки были разделены на части размером 512×512, так как исходные изображения были слишком большими, чтобы на них обучать модель.

Результаты и их обсуждение. Карты землепользования (LULC) часто создаются на основе спутниковых снимков среднего разрешения, таких как Sentinel и Landsat. Однако при создании карт LULC для городских территорий необходимо классифицировать конкретные объекты, в частности автомобили, здания и деревья. При извлечении из аэрофотоснимков городских особенностей или информации о земном покрове пространственное разрешение становится более важным, чем спектральное. Высокое пространственное разрешение помогает лучше видеть наземные объекты, особенно в городских районах, что делает коммерческие спутниковые снимки более популярными. С увеличением пространственного разрешения городские объекты видны на спутниковых снимках намного лучше, что побуждает сместить акценты в исследованиях со спектральной классификации изображений, анализа на основе пикселей и анализа на основе объектов на семантическую сегментацию на уровне пикселей.

В начале развития анализа на основе пикселей размер пикселя был недостаточно мелким для точной идентификации объектов на изображениях. По мере улучшения пространственного разрешения спектральный отклик от различных мелких объектов в городских районах становился все более сложным. Эта сложность возникла потому, что многие объекты сделаны из похожих материалов и излучают схожие спектральные отклики. Традиционные методы классификации и попиксельные классификаторы, такие как максимальное правдоподобие, были неэффективны в этих случаях, поскольку они полагались исключительно на спектральную информацию на уровне пикселей и игнорировали пространственную и контекстуальную информацию. Более того, методы анализа на основе пикселей часто приводили к появлению шума «соль с перцем» после классификации. Для преодоления подобных ограничений были разработаны методы анализа изображений на основе объектов OBIA (Object Based Image Analysis).

В отличие от анализа на основе пикселей, OBIA начинается с сегментации изображений на объекты с похожими текстурными, пространственными и спектральными атрибутами. Затем эти объекты анализируются на основе их геометрии, размера, текстуры для классификации признаков. OBIA учитывает контекстную информацию об объектах, в связи с чем признан превосходящим анализ на основе пикселей для классификации изображений очень высокого разрешения (VHR). Для OBIA были разработаны классификаторы с контролируемым и неконтролируемым обучением, в то время как анализ на основе пикселей был не столь полезен до недавнего применения методов глубокого обучения для семантической сегментации изображений VHR на уровне пикселей.

Landsat 8 использует для сбора данных сегменты электромагнитного спектра, представляющие различные цвета, в том числе не видимые человеческому глазу. Каждый снимок состоит в общей сложности из 11 каналов. Красный, зеленый и синий каналы на Landsat 8 обозначены как каналы 4, 3 и 2, соответственно. В сочетании эти полосы создают изображение в истинном цвете. Однако только каналы 1–4 и 8 воспринимают видимый свет, в то время как остальные полосы охватывают невидимые для нас части спектра. Следовательно, изображение в истинном цвете, полученное спутником Landsat 8, представляет собой менее половины того, что он действительно может обнаружить.

Например, канал 5 обнаруживает ближнее инфракрасное излучение (NIR), которое имеет важное экологическое значение. Здоровые растения отражают NIR, поскольку вода в их листьях рассеивает эти длины волн обратно в небо. В процессе сравнения полосы 5 с другими полосами генерируются такие индексы, как NDVI (нормализованный разностный вегетационный индекс), позволяющие более точно оценить состояние растений за пределами видимой зелени. Полосы 6 и 7 захватывают различные участки коротковолнового инфракрасного спектра (SWIR). Они полезны для различения влажной и сухой почвы, а также при геологических исследованиях. Горные породы и почвы, которые кажутся похожими в других диапазонах, часто демонстрируют заметные различия в спектре SWIR. Канал 8, известный как панхроматическая или панорамная полоса, функционирует аналогично черно-белой пленке. Вместо того, чтобы захватывать видимые цвета по отдельности, он объединяет их в один канал. Этот датчик может одновременно улавливать более широкий диапазон света, что обеспечивает самое высокое разрешение среди всех диапазонов на расстоянии 15 метров. Каналы 10 и 11 относятся к тепловой инфракрасной области (TIR), что позволяет им обнаруживать тепло. Вместо измерения температуры воздуха, как это делают метеостанции, эти диапазоны дают представление о температуре поверхности земли. Комбинируя различные полосы, многочисленные комбинации могут раскрыть характерные особенности ландшафта.

Целью работы являлось построение модели сегментации полей и определения их границ. Для решения этой задачи был собран и размечен датасет. У Sentinel Hub есть удобный программный интерфейс приложения, в котором можно выбирать, на сколько процентов снимок покрыт облаками, что значительно помогает в отборе снимков при формировании датасета. В итоге было собрано 3 тысячи снимков, столько же масок полей и их границ. Изначально эксперименты проводились с моделями UNet (рис. 1), DeepLabv3plus (рис. 2) и их различными модификациями.

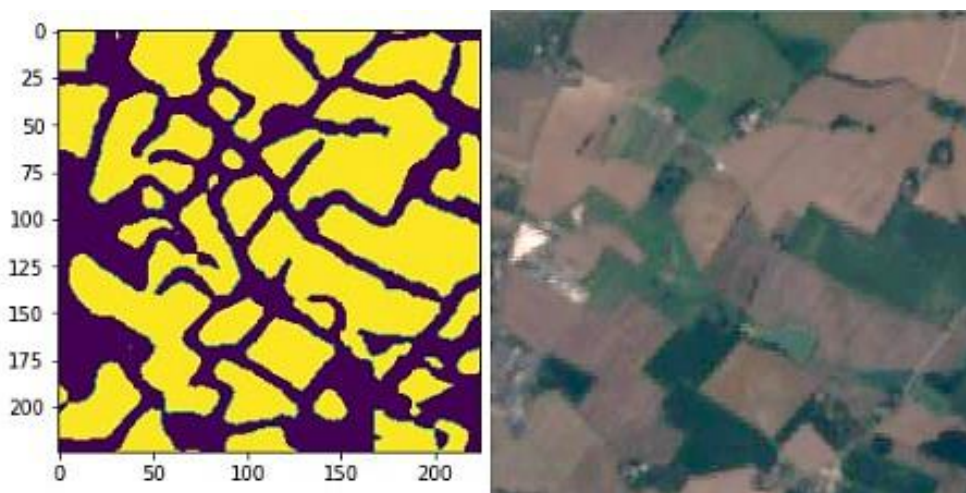


Рис. 1. Результат работы Unet, маски полей



Рис. 2. Результат работы Deeplabv3plus, границы полей

К сожалению, эти модели при предсказывании как масок полей, так и их границ показали низкие значения F1-score. Для дальнейших экспериментов была выбрана модель HRNetv2 (рис. 3).

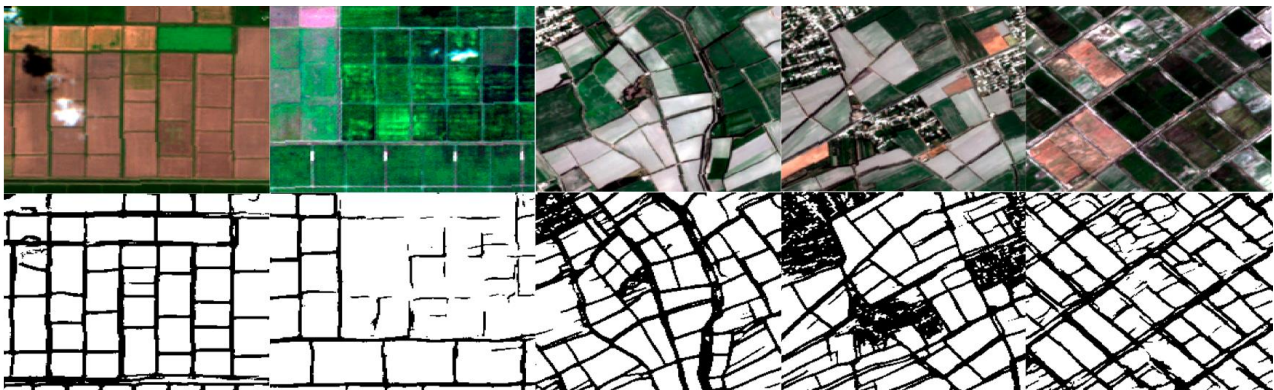


Рис. 3. Результат работы HRNetv2, границы и маски полей

Результаты всех проведенных экспериментов с моделями из библиотеки mmsegmentaion представлены в табл. 1.

Таблица 1

Результаты экспериментов на первом этапе

Metric	UNet	DeepLabv3plus	HRNetv2
mAcc	22.56	72.33	90.57
mIoU	11.28	36.16	45.29
aAcc	20.9	71.6	90.33

Так как более высокие значения метрик показала модель HRNetv2, то далее было решено провести эксперименты с ее гиперпараметрами. Считаем, что представленная модель продемонстрировала лучшие результаты в эксперименте потому, что сохранение высокого разрешения исходной картинки на протяжении всей сети значительно влияет на точность границ. В дальнейшем в экспериментах пробовали использовать классические каналы RGB, RGB+NIR, обучаться только на масках границ и только на масках границ полей или обучать модель предсказывать их одновременно (табл. 2).

По результатам эксперимента видно, что значительно лучший результат дает подход с расчетом ошибки по маскам границ и маскам полей. Так же на точности модели значительно сказалось использование для обучения маски поля и его границ и их последующего объединения. В лучшую модель

на вход подается 4-канальное изображение (RGB+NIR с Sentinel-2), в качестве лэйблов используются маска и границы поля. Соответственно выходная маска является 2-канальной (маска+границы). Кроме этого BCE Loss был изменен на Dice+BCE Loss, а оптимизатор SGD – на Adam. Была предпринята попытка использовать новый оптимизатор Lion, но он способствовал более быстрому переобучению сети. Последние 2 слоя Upsample в головке сегментации FCN изменены на ConvTranspose2D, также были использованы различные аугментации данных во время обучения. Итоговые полученные метрики для оценки производительности модели: Dice mask = 0.8, Dice borders = 0.66.

Таблица 2

Результаты экспериментов на втором этапе

Metric	Границы	Маски	Границы+маски	RGB+NIR
mAcc	53.9	55.7	76	90.57
mIoU	18.3	19.05	33.3	45.29
aAcc	51.1	51.5	74.2	90.33

Для оценки качества модели было принято решение протестировать модель на данных соревнования SpaceNet 2020, основной задачей которого является задача сегментации домов по спутниковым снимкам (рис. 4).

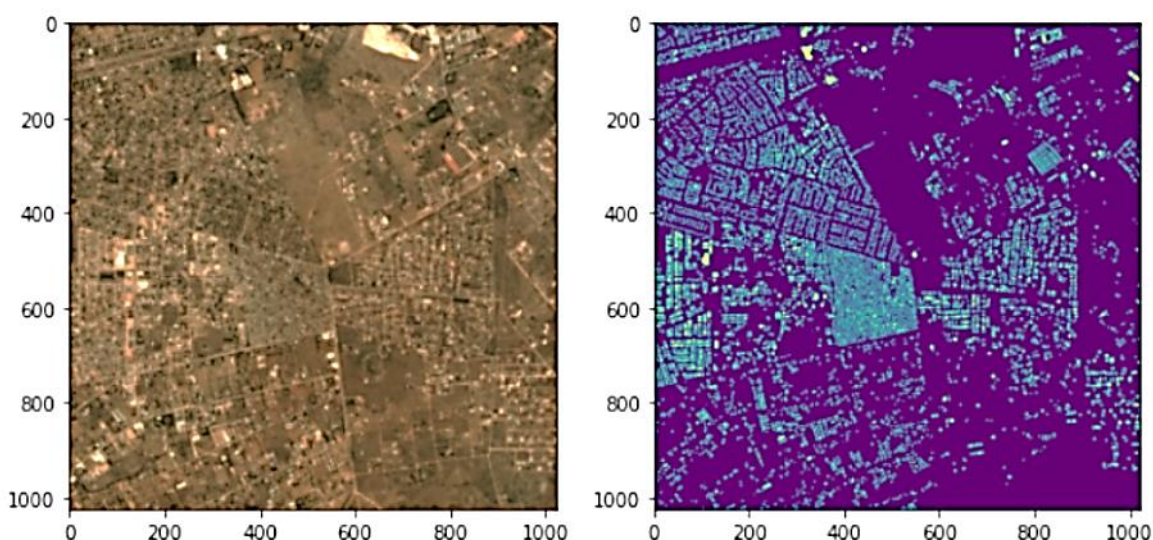


Рис. 4. Результат работы HRNetv2 на данных с соревнования

Для решения этой задачи помимо семантической сегментации можно использовать детекцию объектов [6] и сегментацию экземпляров. Но учитывая, что пересечения между зданиями отсутствуют, а на одном изображении слишком много зданий, сегментация экземпляра требует больше времени на вывод. Семантическая сегментация будет более подходящей, чем сегментация экземпляров. Обнаружение объектов – еще одно возможное решение задачи. Однако ввиду того, что количество целей на каждом изображении велико, а сами цели чрезвычайно малы, необходимо разработать специальный метод обнаружения. Мы использовали HrNet, так как он в данном случае идеально подходит в силу того, что многие здания очень маленькие и требуют более высокого разрешения. К тому же использовалось увеличение изображения до 3-кратного размера для обучения и вывода. Это было сделано потому, что даже если применяется HRNet, в начале сети будет производиться понижающая дискретизация на 1/4. Это неприемлемо для зданий с наименьшим размером 2×2 пикселя. Все обучение и вывод выполняются в 3-кратных условиях. Для изображения в формате .tif масштаб был увеличен в 3 раза и соответственно все координаты были умножены в 3 раза. Чтобы ускорить считывание обучающих данных, для обучающего набора изображения и маски были разбиты на части 512×512.

Постобработка, основанная на маске здания и его границах, значительно увеличила точность, но в этой задаче использование канала NIR не дало столь же сильного прироста в точности.

В соревновании, которое проводилось в 2020 году, первое место заняла модель HrNet, как видно из данных табл. 3, разработанная авторами модификация показывает значительно более высокие значения метрики.

Таблица 3

Результаты экспериментов на данных с соревнования SpaceNet 2020

Модель	Dice mask	mIoU
HrNet	0.41	75.6
HRNetv2+	0.467	82.6
OneFormer	0.432	84.0

Заключение. В работе показано, что HRNetv2 с предложенными авторами модификациями имеет преимущества по сравнению с существующими сетями, такими как DeepLabV3Plus, UNet, стандартным HRNet и OneFormer. Перечисленные сети теряют много необходимой пространственной информации в процессе получения маски высокого разрешения из представления с низким разрешением (процесс upsampling). HRNetv2 соединяет сети высокого и низкого разрешения параллельно, а не последовательно, обеспечивая представление высокого разрешения на протяжении всей сети. Следовательно, итоговая маска сегментации является более точной и пространственно более достоверной, что сыграло очень важную роль в получении точных границ полей и домов. К тому же значительное влияние на итоговую точность оказало использование подхода расчета границ домов и маски домов одновременно, а также применение совместной ошибки по границам домов и маскам домов при обучении сети. Стоит отметить, что использование дополнительного канала NIR значительно повысило итоговую точность сети в задаче определения границ, но дало меньший прирост точности в задаче сегментации домов.

ЛИТЕРАТУРА

1. Корчевская, Е.А. Использование сверточной нейронной сети для решения задачи классификации / Е.А. Корчевская, Л.В. Маркова, Т.В. Никонова // Вестник Віцебскага дзяржаўнага ўніверсітэта. – 2022. – № 2(115). – С. 5–9.
2. Long, J. Fully Convolutional Networks for Semantic Segmentation / J. Long, Ev. Shelhamer and T. Darrell. – URL: <https://arxiv.org/pdf/1411.4038> (date of access: 17.12.2024).
3. Никонов, Н.Д. Использование сверточных нейронных сетей для решения задач классификации в неконтролируемых условиях / Н.Д. Никонов, Т.В. Никонова, О.Е. Рубаник, Е.А. Корчевская // Вестник Віцебскага дзяржаўнага ўніверсітэта. – 2023. – № 2(119). – С. 5–11.
4. Han, L.A. Farmland Parcel Delineation Using Spatio-temporal Convolutional Networks / L.A. Han, U. Burak, B. Marshall // Electrical Engineering and Systems Science. – 2020. – URL: <https://arxiv.org/pdf/2004.05471> (date of access: 17.12.2024).
5. Tao, C. DESTIN: A new method for delineating the boundaries of crop fields by fusing spatial and temporal information from WorldView and Planet satellite imagery / C. Tao, X. Ji, Ya. Gaoxiang // Computers and Electronics in Agriculture. – 2020. – Vol. 178. – P. 249–253.
6. Malcolm, D. Linking Land Policy with Climate Change: A Multi-dimensional Landscape Approach to Territorial Development with a Focus on the Europe and Central Asia (ECA) Region / D. Malcolm, P. Siegel, M. Törhönen // Second Central Asia Climate Knowledge Forum: Moving Towards Regional Climate Resilience, Almaty, Kazakhstan, May 2014. – P. 187–203.

REFERENCES

1. Korchevskaya E.A., Markova L.V., Nikonova T.V. *Vesnik Vitsebskaga dziazhaunaga universiteta* [Bulletin of Vitebsk State University], 2022, 2(115), pp. 5–9.
2. Long, J. Fully Convolutional Networks for Semantic Segmentation / J. Long, Ev. Shelhamer and T. Darrell. – URL: <https://arxiv.org/pdf/1411.4038> (date of access: 17.12.2024).
3. Nikonov N.D., Nikonova T.V., Rubanik O.E., Korchevskaya E.A. *Vesnik Vitsebskaga dziazhaunaga universiteta* [Bulletin of Vitebsk State University], 2023, 2(119), pp. 5–11.
4. Han, L.A. Farmland Parcel Delineation Using Spatio-temporal Convolutional Networks / L.A. Han, U. Burak, B. Marshall // Electrical Engineering and Systems Science. – 2020. – URL: <https://arxiv.org/pdf/2004.05471> (date of access: 17.12.2024).
5. Tao, C. DESTIN: A new method for delineating the boundaries of crop fields by fusing spatial and temporal information from WorldView and Planet satellite imagery / C. Tao, X. Ji, Ya. Gaoxiang // Computers and Electronics in Agriculture. – 2020. – Vol. 178. – P. 249–253.
6. Malcolm, D. Linking Land Policy with Climate Change: A Multi-dimensional Landscape Approach to Territorial Development with a Focus on the Europe and Central Asia (ECA) Region / D. Malcolm, P. Siegel, M. Törhönen // Second Central Asia Climate Knowledge Forum: Moving Towards Regional Climate Resilience, Almaty, Kazakhstan, May 2014. – P. 187–203.

Поступила в редакцию 14.01.2025

Адрес для корреспонденции: e-mail: st.rubon@mail.ru – Никонова Т.В.