УДК 004.85
doi:10.18720/SPBPU/2/id23-106

Махонько Яна Викторовна [1],
техник;
Петряева Александра Андреевна [2],
специалист;
Подмарькова Виктория Александровна [3],
специалист;
Галактионов Игорь Дмитриевич [4],
студент;
Дмитриева Лидия Алексеевна [5],
младший научный сотрудник

ОБЗОР МЕТОДОВ СЕМАНТИЧЕСКОЙ СЕГМЕНТАЦИИ СПУТНИКОВЫХ СНИМКОВ ЗЕМЛИ С ПРИМЕНЕНИЕМ НЕЙРОСЕТЕВЫХ ТЕХНОЛОГИЙ

[1, 2, 3, 5] Россия, Санкт-Петербург, Санкт-Петербургский политехнический университет Петра Великого, Лаборатория «Промышленные системы потоковой обработки данных» Центра НТИ СПбПУ,
[1] mahonko_yav@spbstu.ru, [2] alexandra.petryaeva@spbpu.com,
[3] podmarkova_va@spbstu.ru, [5] lidiya.dmitrieva@spbpu.com;
[4] Россия, Санкт-Петербург, Санкт-Петербургский политехнический университет Петра Великого, galaktionov.id@edu.spbstu.ru

Аннотация. В данной работе рассмотрена проблема экологического мониторинга с помощью космических снимков разной точности на примере использования машинного обучения — нейронных сетей для распознавания разных классов объектов и мониторинга состояния окружающей среды. Многие локальные и

глобальные экологические проблемы могут быть выявлены с использованием спутниковых снимков разной детальности: высокого и низкого разрешения. Снимки высокого разрешения используют для мониторинга локальных проблем, например зоны вырубки вблизи городов, а низкого разрешения — для большего охвата территории, например зоны пожаров. В рамках исследования были проанализированы статьи базы научных публикаций Scopus. Результаты данного исследования могут быть использованы руководителями подразделений при подборе инструмента моделирования для решения задач сегментирования и классификации отдельных участков, учитывая разные классы, задачи и отрасли.

Ключевые слова: нейронные сети, разметка, машинное обучение, экологический мониторинг, космические снимки, датасет.

Yana V. Makhonko [1],
Technician;
Alexandra A. Petryaeva [2],
Specialist;
Victoria A. Podmarkova [3],
Specialist;
Igor D. Galaktionov [4],
Student;
Lydia A. Dmitrieva [5],
Junior Researcher

REVIEW OF METHODS OF SEMANTIC SEGMENTATION OF SATELLITE IMAGES OF THE EARTH USING NEURAL NETWORK TECHNOLOGIES

[1, 2, 3, 5] Laboratory of Industrial Systems for Streaming Data Processing of the SPbPU National Technology Initiative Center for Advanced Manufacturing Technologies, Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia,
[1] mahonko_yav@spbstu.ru, [2] alexandra.petryaeva@spbpu.com,
[3] podmarkova_va@spbstu.ru, [5] lidiya.dmitrieva@spbpu.com;
[4] Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia, galaktionov.id@edu.spbstu.ru

Abstract. In this paper, the problem of environmental monitoring using space images of different accuracy is considered on the example of using machine learning — neural networks for recognizing different classes of objects and monitoring the state of the environment. Many local and global environmental problems can be identified using satellite images of different detail: high and low resolution. High-resolution images are used to monitor local problems, for example, clear-cut areas near cities, and low-resolution images are used to cover a larger area, for example, fire zones. As part of the study, articles from the Scopus database of scientific publications were analyzed. The results of this study can be used by department heads when selecting a modeling tool to solve the problems of segmenting and classifying individual sections, taking into account different classes, tasks and industries.

Keywords: neural networks, markup, machine learning, environmental monitoring, satellite images, dataset.

Введение

В настоящее время технология применения нейронных сетей в отдельных отраслях (транспорт, сельское, лесное и водное хозяйства, промышленность, нефтегазовый комплекс) набирает популярность. Важнейшей особенностью современного уровня развития технологий является обеспечение приёма и обработки изображений Земли [10] из космоса, доступность актуальных снимков различных территорий за определённый период времени, а также наличие инструментов для их обработки.

Благодаря спутникам можно получить данные о погоде, наводнениях, пожарах, осадках, движении песков в пустынях, таянии ледников и движении айсбергов, многих других явлениях антропогенного и природного характера.

Существующие современные технологии космического мониторинга в принципе позволяют решать сформулированные задачи, однако требуют тщательного анализа и адаптации к конкретным задачам с учётом многочисленных природно-промышленных факторов: климатических условий, природных и техногенных факторов.

С использованием технологий обработки космических снимков можно выполнить анализ дымового шлейфа, определить очаги, посчитать расстояние до населенных пунктов, оценить динамику распространения пожара, рассчитать площадь зон горения, выполнить классификацию по характеру происхождения.

Материалы и методы

В рамках исследования был выполнен обзор статей базы научных публикаций Scopus по ключевым словам: «семантическая сегментация», «разметка», «пиксель», «глубокое обучение», «снимки высокого разрешения» и «беспилотные летательные аппараты». Благодаря анализу найденных статьей были выделены методы и модели нейронных сетей, позволяющие улучшить точность семантической сегментации.

Результаты

На основании проанализированных источников были выделены модели, основанные на свёрточных нейронных сетях, FCN (Fully-Convolutional Neural Network, полностью свёрточная нейронная сеть), SegNet (Segmentation network, сеть сегментации) и ResNet (Residual Network, остаточная сеть). В данных статьях были предприняты меры по улучшению точности сегментации.

В статье [1] рассматривался способ повышения точности классификации. Препятствиями для этого в изображениях с высоким разрешением являются маленькая и большая разница внутри классов. Это может привести к ошибке отверстия. Для решения этой проблемы была разработана модель с двумя выходами. Первым этапом этой модели является по-

нижение дискретизации для извлечения признаков, вторым используется многомасштабный модуль для получения результата классификации на уровне области, который может ограничивать взаимосвязь пикселей в одном и том же объекте, чтобы устранить дыры. И потом уже идёт уточнение результата на уровне пикселя, к которой принадлежит пиксель, как только область точно классифицируется, каждый пиксель в ней также классифицируется правильно. Эксперименты на двух наборах данных с разным разрешением продемонстрировали правомерность добавления RO-пути (region-level output, выход на уровне области).

В статье [2] использовалась модель ScasNet (self-cascaded convolutional neural network, самокаскадная нейронная сеть), которая будет обеспечивать более точные результаты маркировки. Предлагается автокаскадная архитектура, приводящая к согласованию результатов маркировки запутанных искусственных объектов. Это достигается с помощью последовательной агрегации контекстов от крупных до мелких, что сохраняет глобальные и локальные контексты. Кроме того, используется стратегия уточнения от грубого к точному, которая постепенно уточняет целевые объекты. Для решения проблемы скрытого остатка прогонки, возникающей при объединении нескольких признаков с разной семантикой, которое может привести к отсутствию информации в процессе слияния, предлагается схема коррекции скрытого остатка. Когда схема остаточной коррекции используется в каждой позиции после слияния многоуровневых контекстов, можно увидеть улучшение в производительности и среднего показателя meanIoU (Расчет пересечения по объединению для оценки модели сегментации изображения) почти на 3 %.

Проблема точности определения границ рассматривается в статье [3]. Для определения унарных потенциалов BNF (Boundary Neural Fields, Граничные нейронные поля) используется грубая сегментация, предсказанная FCN, то есть выделяются силуэты объектов без обеспечения точности на их границах. Сверточные карты признаков этой сети можно использовать для точного предсказания семантических границ. Затем эти границы используются для построения сходства парных пикселей, которые используются BNF в качестве парных потенциалов. Вводится глобальная функция энергии, которая минимизирует энергию, улучшающая начальную сегментацию FCN. Качество прогнозируемых границ оценивается с использованием трех стандартных показателей: фиксированного порога контура (ODS), наилучшего порога для каждого изображения (OIS) и средней точности (AP, average precision). По данным показателям BNF показала наилучшие результаты 0.788, 0.807, 0.851 по показателям соответственно.

Немаловажную роль в задаче сегментации играет функция потерь. Предлагается функция GapLoss в статье [4] для сегментации дорог,

улучшающая соединение дорог и их сегментацию одновременно. При обычной сети сегментации и кросс-энтропией (широко используемая функция потерь) часто выявляются разрывы дорог, где их не должно быть или наоборот появляются новые. В предложенной функции предлагается сначала извлечь векторную линию прогнозируемой дороги, потом отметить конечные пиксели, которым в дальнейшем назначатся веса. Если две точки разрыва находятся рядом и перекрываются, то веса рядом с ними удваиваются. Это нужно для переопределения ошибки. В ходе экспериментов было взято несколько моделей с кросс-энтропией и GapLoss поочерёдно. Предложенная функция потерь в сочетании с разными сетями повысила в среднем mIoU (mean intersection over union, среднее пересечение по объединению) на 2.725 %, точность на 0.2 % и F1(F1-score — метрика, позволяющая определить проблему дисбаланса классов) на 4.85 %.

В задаче семантической сегментации может встретиться проблема дисбаланса классов. В статье [5] предлагается решение вышеуказанной проблемы на задаче сегментации деревьев. Была описана идея взвешивать важность каждого пикселя по критериям важности маркировки его класса и неопределённости маркировки. В ходе экспериментов сравниваются базовые методы и предлагаемый подход с использованием SegNet (SegNet — это модель семантической сегментации. Эта базовая обучаемая архитектура сегментации состоит из сети кодировщика, соответствующей сети декодера, за которой следует уровень попиксельной классификации.) и FCN (Fully-Convolutional Neural Network, полностью свёрточная нейронная сеть) соответственно. $\boldsymbol{\sigma}$ — параметр, который является основным для данного подхода. Он соответствует разбросу неопределённости для функции потерь. При $\boldsymbol{\sigma} = 2$ с FCN мы получаем лучший результат, это можно понять по показателям PA (Producer Accuracy, точность производителя) и IoU (Intersection over Union, пересечение по объединению).

Для семантической сегментации урбанизированной (городская инфраструктура) сцены в статье [6] описана сеть UNetFormer, которая эффективно решает задачу. В его архитектуре использован кодировщик на основе CNN (Convolutional Neural Network) и декодер на основе transformera. Был выбран предварительно обученный ResNet18 (Residual Network, Остаточная сеть), так как он продемонстрировал эффективность и действенность одновременно в широком спектре задач семантической сегментации, в качестве кодировщика для извлечения многомасштабных семантических признаков со значительно низкой вычислительной стоимостью. Кодировщик состоит из четырехэтапных Resblocks, при этом на каждом этапе дискретизация карт объектов понижается с коэффициентом масштабирования 2. Карты объединяются с соответствующими картами признаков декодера с помощью свертки. Семанти-

ческие особенности созданные Resblocks (*Residual Block, Остаточный блок*), объединяются с функциями, созданными GLTB (эффективный глобально-локальный механизм внимания с глобальной ветвью внимания и сверточной локальной ветвью для захвата как глобального, так и локального контекстов для визуального восприятия) декодера с помощью операции взвешенной суммы. Операция взвешенной суммы выборочно взвешивает два признака. Три глобально-локальных блока Transformer и одна функция уточнения в начале для создания легковесного декодера на базе Transformer используется в предложенной архитектуре. Благодаря конструкции декодер способен захватывать глобальные и локальные контексты в разных масштабах при сохранении высокой эффективности. Это нужно для повышения точности. Было проведено несколько экспериментов. Она хорошо справилась с датасетом Potsdam с распознаванием дополнительных поверхностей, зданий, машин. Показатели достигают 92,8 % MeanF1(mean F1-score — метрика, позволяющая решить проблему дисбаланса классов) и 86,8 % mIoU (mean intersection over union, среднее пересечение по объединению) на тестовом наборе. Кроме того, OA (overall accuracy, общая точность), составила 91.3 %. Таким образом, данная сеть лучше всех справилась с задачей сегментации.

В статье [7] рассматривают сеть FCNN для сегментации тонких объектов, таких как дорог. Авторы доказывают, что при правильной настройке она может давать хорошие результаты. Улучшения специально для сегментации дорог могут быть осуществлены путём добавления правил допуска, которые позволяют произойти пространственно небольшим ошибками, не критично влияющих на результат. Было разработано три разные сети на FCNN. Первая — FCN-8 с магистралью VGG-19, которая первая использовалась для решения задач компьютерного зрения. Вторая — Deep Residual U-Net, который продемонстрировал отличную производительность сегментации. Третий — одна из новых архитектур для семантической сегментации DeepLabv3+. В ходе сравнения этих сетей было выявлено, что разработанная версия Deeplabv3+, модифицированная с регрессионной потерей mse (Mean Squared Error, Среднеквадратическая ошибка), перебалансированная в сторону дорожного класса, достигает в среднем 44 % IoU (Intersection over Union, Пересечение по объединению — расчет для оценки модели сегментации изображения), что делает её более подходящим для задачи определения дорог.

В статье [8] ставится задача более точного извлечения информации о зданиях. Для решения поставленной задачи используется подход на основе глубокого обучения, где сетевая архитектура SegNet будет выполнять задачу семантической сегментации. Она состоит из сети кодировщика, сети декодера и последнего слоя классификации на основе пикселей. 13 слоев свертки и слоев с максимальным объединением, ко-

торые обеспечивают инвариантность для небольших изображений со сдвигом в пространстве, входят в состав сети кодировщика. Для пакетной нормализации, которая будет применена к Rectified Linear Unit (ReLU, Выпрямленная линейная единица), и для проведения нелинейного преобразования карты характеристик в систему и создания карт характеристик каждый слой будет свёрнут. Результаты эксперимента показывают, что была получена отличная производительность при обнаружении зданий из неподготовленных мест со средней OA более 90 %.

В таблице 1 представлены результаты аналитического обзора.

Таблица 1

Результаты анализа источников

Номер	**Нейронная сеть/метод**	**Набор данных**	**Полученный результат (точность и тд)**	**Задача**
1	2	3	4	5
1	Сети с добавлением PO и RO-путей (метод)	1) GID 2) Postdam Dataset	Accuracy: 80,67 – 92,21 %	Повышение точности классификации изображений с высоким разрешением
2	ScasNet (нейронная сеть)	1) ISPRS Vaihingen Challenge Dataset 2) ISPRS Potsdam Challenge Dataset	1) OA: 91,10 meanIoU: 83,90 meanF1: 91,6 2) OA: 91,10 meanIoU: 87,08 meanF1: 93,31	Повышение точности маркировки
3	BNF (метод)	1) SBD 2) BSDS 500	ODS: 0,788 OIS: 0,807 AP: 0,851	Точность границ
4	PSPNet+Gaploss (метод)	1) Massachusetts Roads Dataset 2) DeepGlobe Roads Dataset 3) Aerial Image Segmentation Dataset	mIoU: 73,8 F1: 65,4 Accuracy: 96,9	Увеличение точности сегментации дорог
4.1	Unet++ + Gaploss	Massachusetts Roads Dataset	mIoU: 74,6 F1: 69,6 Accuracy: 97,1	
4.2	SegNet+Gaploss	Massachusetts Roads Dataset	mIoU: 71,6 F1: 63,8 Accuracy: 96,1	

Продолжение таблицы 1

1	2	3	4	5
4.3	MUNet+Gaploss	Massachusetts Roads Dataset	mIoU: 74,0 F1: 67,4 Accuracy: 97,2	Увеличение точности сегментации дорог
5	1) FCN+SGD 2) SegNet+SGD(метод)	Urban Tree	1) PA: 90 % IoU: 76 % 2) PA: 83.8 % IoU: 70.5 %	Задача улучшение маркировки растительности
6	UnetFormer (нейронная сеть)	1) UAVid 2) Postdam 3) Vaihingen	meanF1: 90,4 OA: 91,0 mIoU: 82,7	Семантическая сегментация городской территории
7	1) FCN-8 (нейронная сеть) 2) Deep Residual U-Net 3) DeepLabv3+	Собственный набор данных	1) IoU: 43,96 % 2) IoU: 32,13 % 3) IoU: 43,73 %	Сегментация дорог
8	SegNet с подходом глубокого обучения (метод)	1) Inria 2) Postdam	1) OA: 93,42 % MIoU: 85,32 % 2) OA: 95,79 % MIoU: 87,80 %	Более точное извлечение информации о зданиях при сегментации

Были рассмотрены модели и методы по задаче семантической сегментации различных географических объектов [9]. Для повышения точности определения водных объектов предлагается обратить внимание на функцию потерь. Кроме того, на рассмотрение предлагается использовать механизм внимания, добавление путей пиксельного и областного выхода, рассмотреть автокаскадную архитектуру. Точность на границах может быть увеличена с помощью BNF. Для решения проблемы деградации сети подходят остаточные сети и предложенная SR-SegNet. Таким образом, было определено на что стоит обратить внимание для решения проблем, появляющихся в процессе сегментации снимков высокого разрешения и задачи по увеличению её точности.

Заключение

На основании проанализированных статей базы научных публикаций Scopus были выделены три категории внесённых изменений в модели для улучшения точности семантической сегментации: получение карт признаков, классификация на уровне пикселя, внедрение разработанной или подходящей функции потерь. Первая категория используется для точного предсказания границ, решения проблем, связанных со скры-

тым остатком, уточнения объекта и выбор отличительного признака. Классификация на уровне пикселя может предотвратить ошибку отверстия, улучшить точность до пикселя, но недостаток данного подхода заключается в том, что он требует больших затрат времени. Третья категория изменений — функция потерь, которая должна корректировать работу нейронной сети. В задаче сегментации узких объектов она играет большую роль, поэтому важно её подобрать правильно. Проведённое исследование и его результаты могут быть полезны при подборе инструментов для улучшения нейронной сети в задаче семантической сегментации.

Список литературы

1. Kang W., Xiang Y., Wang F., You H. DO-Net: Dual-output network for land cover classification from optical remote sensing images // IEEE Geoscience and Remote Sensing Letters. – 2022. – Vol. 19.

2. Liu Yongcheng, Fan Bin, Wang Lingfeng, Bai Jun, Xiang Shiming, Pan Chunhong. Semantic labeling in very high resolution images via a self-cascaded convolutional neural network // ISPRS Journal of Photogrammetry and Remote Sensing. – 2018. – Volume145. – Pp.78–95.

3. Bertasius Gedas, Shi Jianbo,Torresani Lorenzo. Semantic segmentation with boundary neural fields // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2016. – Pp. 3602–3610.

4. Yuan W., Xu W. GapLoss: a loss function for semantic segmentation of roads in remote sensing images // Remote Sensing. – 2022. – Vol. 14 (10). № 242.

5. Bressan P.O., Junior J.M., Correa Martins J.A., de Melo M.J., Gonçalves D.N., Freitas D.M., Marques Ramos A.P., Garcia Furuya M.T., Osco L.P., de Andrade Silva J., Luo Z., Garcia R.C., Ma L., Li J., Gonçalves W.N. Semantic segmentation with labeling uncertainty and class imbalance applied to vegetation mapping // International Journal of Applied Earth Observation and Geoinformation. – 2022. – Vol. 108. – 102690.

6. Wang L., Li R., Zhang C., Fang S., Duan C., Meng X., Atkinson P.M. UNetFormer: a UNet-like transformer for efficient semantic segmentation of remote sensing urban scene imagery // ISPRS Journal of Photogrammetry and Remote Sensing. – 2022. – Vol. 190. – Pp. 196–214.

7. Henry C., Azimi S.M., Merkle N. Road segmentation in SAR satellite images with deep fully convolutional neural networks // IEEE Geoscience and Remote Sensing Letters. – 2018. – Vol. 15 (12). – № 8447237. – Pp. 1867–1871.

8. Boonpook W., Tan Y., Ye Y., Torteeka P., Torsri K., Dong S. A deep learning approach on building detection from unmanned aerial vehicle-based images in riverbank monitoring // Sensors (Switzerland). – 2018. – Vol. 18 (11). – № 3921.

9. Лабутина И.А., Балдина Е.А. Практикум по курсу «Дешифрирование аэрокосмических снимков» / Московский гос. ун-т им. М. В. Ломоносова. – М.: Географический факультет МГУ, 2013. – 175 с.

10. Сутырина Е. Н. Дистанционное зондирование земли : учеб. пособие. – Иркутск: Изд-во ИГУ, 2013. – 165 с.