

СЕГМЕНТАЦИЯ СПУТНИКОВЫХ ИЗОБРАЖЕНИЙ АЛГОРИТМИЧЕСКИМИ МЕТОДАМИ И МЕТОДАМИ МАШИННОГО ОБУЧЕНИЯ

С.В. Окуньков^{1,2}, М.А. Барулина^{1,3}

¹Институт проблем точной механики и управления РАН

²Саратовский национальный исследовательский государственный университет имени

Н. Г. Чернышевского

³Пермский государственный национальный исследовательский университет

sergeiokunkov@yandex.ru

Автоматизированное распознавание объектов на местности по спутниковым снимкам является актуальной и перспективной задачей на сегодняшний день. Ее решение открывает множество возможностей в синоптической метеорологии, геологии, картографии и других отраслях промышленной и научной сфер. Данную задачу можно решать с помощью двух разных подходов: обнаружение объектов на изображении и сегментация объектов на изображении. Причем сегментация является более доступным и дешевым решением, так как требует менее кропотливой разметки данных и меньшего количества самих данных, а также позволяет применять “классические” алгоритмы для обработки изображений для решения рассматриваемой задачи, такие как фильтры Робетса, Превитта, размытие по Гауссу и т.д. Однако подход с обнаружением объектов является более точным и не требует дальнейшего анализа масок, полученных с помощью алгоритмов сегментации. **Цель настоящей статьи** - исследование возможностей классических алгоритмов и алгоритмов глубокого обучения для сегментации объектов на спутниковых фотографиях и возможности практического применения данных алгоритмов. Для достижения поставленной цели необходимо выбрать набор данных для обучения, проанализировать его, обучить ряд алгоритмов глубокого обучения, сравнить полученные результаты с результатами классических алгоритмов, сформулировать выводы и конкретные рекомендации по использованию алгоритмов ИИ для рассматриваемой задачи.

В качестве исходных изображений для обучения был использован набор данных DOTA [1]. Изображения в нем получены со спутников Google Earth, GF-2 и JL-1. В качестве масок для изображений был взят набор данных iSAID [2], который является разметкой набора данных DOTA с 15 классами объектов. Итоговый набор данных, на котором проводились эксперименты, представляет собой 1869 пар изображение-маска (Рис. 1). После он был разделен на две выборки (обучающую и валидационную) в соотношении 9 к 1.

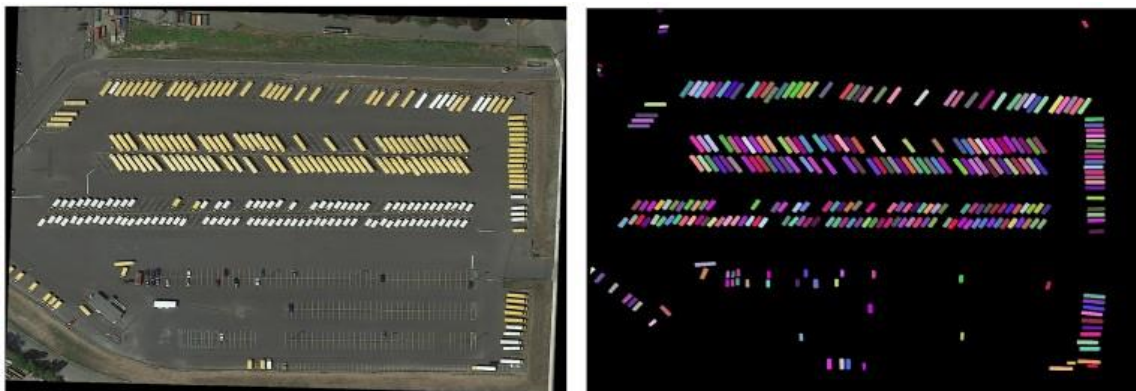


Рисунок 1 - Пример пары исходного изображения и его маски

Так как данных мало, было принято решение свести задачу к бинарной сегментации, переводя изображения в черно-белый формат средствами языка Python и библиотеки для работы с изображениями OpenCV [3].

В качестве стартовой простой модели были использованы ансамбль из классических алгоритмов, представленных в библиотеке OpenCV: применение черно-белого фильтра к исходному изображению, а затем применение глобального фильтра, убирающего шумы, к полученному изображению. Преимущество данной модели заключается в том, что для ее применения ее не нужно обучать, а следовательно, нет необходимости в большом объеме хорошо размеченных данных, однако это преимущество нивелируется качеством предсказания модели, которое сильно ниже, чем у подходов с глубоким обучением. Это связано с тем, что прогнозы модели зависят только от цвета и яркости пикселей, отбрасывая пиксели, которые имеют значение ниже некоторого порога.

В качестве более серьезных моделей с использованием глубокого обучения были использованы классическая реализация архитектуры UNet [4] и предобученные модель PAN [5] и Manet [6] с энкодером в виде ResNet50 [7]. Архитектура UNet на сегодняшний день считается классической. На ее основе строятся все современные архитектуры нейронных сетей для сегментации. Архитектуры PAN и MANet же является просто ее модификациями, использующими механизм внимания, описанный еще в архитектуре трансформеров. Модели с этим механизмом в данный момент считаются самыми лучшими по показателям качества.

В качестве функции потерь для обучения была выбрана функция бинарной кросс-энтропии. В качестве оптимизатора был использован алгоритм AdamW [8] с параметром скорости обучения 0.0002.

В качестве целевых метрик для оценки качества прогноза моделей были выбраны метрики Precision, Recall, F1-score и Intersection over Union (IoU), так эти метрики позволяют хорошо оценить модели в задачах бинарной сегментации. Значения этих метрик для вышеописанных моделей приведены в Таблице 1. Пример работы модели PAN показан на рисунке 2.

Таблица 1 – Значения метрики после обучения моделей

Название модели	Значение Precision	Значение Recall	Значение F1	Минимальное значение IoU	Максимальное значение IoU	Значение IoU
Алгоритмический подход	0.16176	0.3129	0.2133	0	0.4903	0.0499
UNet	0.4926	0.5373	0.5139	0	0.9085	0.5287
PAN	0.4509	0.6142	0.52	0	0.8332	0.5235
MANet	0.486	0.6026	0.538	0	0.8739	0.5572

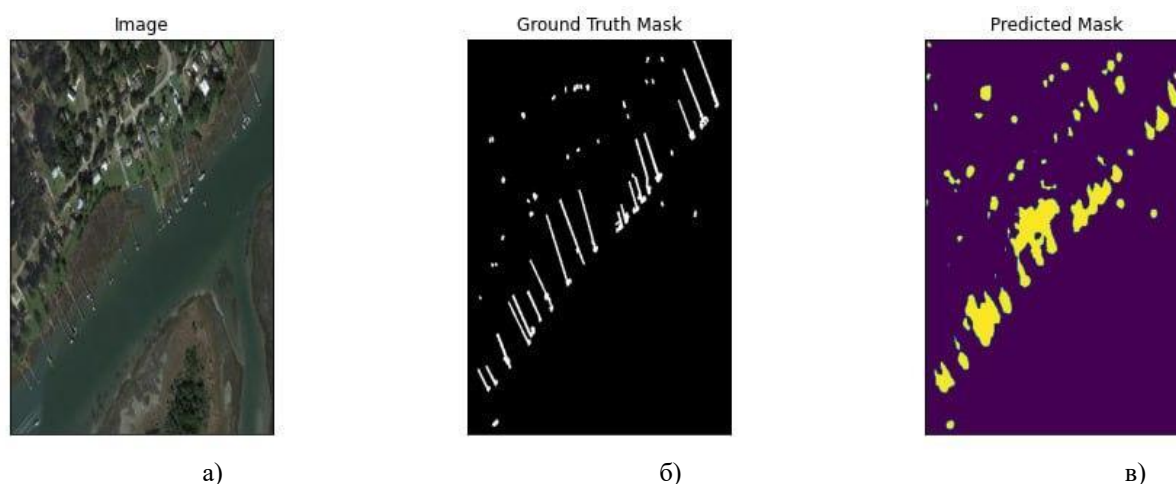


Рисунок 2 - Результат работы модели PAN:
исходное изображение (а); сегментированное вручную (б); сегментация с помощью модели (в)

Исходя из полученных результатов можно сделать следующие выводы. Алгоритмический подход работает намного хуже алгоритмов глубокого обучения, и плохо подходит для практического использования. У всех рассмотренных моделей значения recall (полнота) выше, чем значение precision (точность). Это значит, что модели больше ошибаются с предсказанием лишних пикселей, что в данной задаче менее критично, чем отсутствие предсказания необходимых объектов. Минимальное значение IoU у всех моделей 0, причем это значение было получено на изображениях, на которых присутствует небольшое количество очень маленьких объектов (например контейнеров с нефтью). Лучшее поведение модели показывают себя на изображениях с большим количеством больших объектов (например, машин). Следовательно, обнаружение объектов со спутниковых снимков с помощью сегментации алгоритмами машинного обучения на сегодняшний день хорошо работает только на больших объектах на изображении.

Список литературы:

1. DOTA: [сайт]. URL: <https://captain-whu.github.io/DOTA/dataset.html> (дата обращения: 02.06.2023).
2. iSAID: [сайт]. URL: <https://captain-whu.github.io/iSAID/index.html> (дата обращения: 02.06.2023).
3. OpenCV: [сайт]. URL: <https://opencv.org/> (дата обращения: 03.06.2023).
4. Ronneberger O., Fischer P., Brox T. U-net: Convolutional networks for biomedical image segmentation // Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18. Springer International Publishing, 2015. С. 234-241.
5. Ma Z., Li M., Wang Y. PAN: Path integral based convolution for deep graph neural networks // arXiv preprint arXiv:1904.10996. 2019.
6. Multiattention network for semantic segmentation of fine-resolution remote sensing images / R. Li [et al.] // IEEE Transactions on Geoscience and Remote Sensing. 2021. Т. 60. С. 1-13.
7. Deep residual learning for image recognition / K. He [et al.] // Proceedings of the IEEE conference on computer vision and pattern recognition, 2016. С. 770-778.
8. Loshchilov I., Hutter F. Decoupled weight decay regularization. // International Conference on Learning Representations, 2017.