

СЕГМЕНТАЦИЯ ИЗОБРАЖЕНИЙ ПОДВОДНОГО МУСОРА И МОРСКИХ ОБИТАТЕЛЕЙ

Пономарев Евгений Сергеевич

Аспирант

Вычислительные системы и анализ данных в науке и технике, Сколковский институт науки и технологий, Москва, Россия

E-mail: evgenii.ponomarev@skoltech.ru

Научный руководитель — Иван Валерьевич Оселедец

Океаны Земли населяет множество различных обитателей - рыб, рачков, моллюсков и др. Однако за последний век к ним присоединилось огромное количество различного мусора, загрязняющего их естественную среду обитания. С применением глубоководной техники становится возможным лучше изучить как естественную фауну, так и места загрязнений. Автоматическое распознавание и сегментация несомненно полезны для автоматизации исследований.

Иллюстрации



Одно из изображений датасета TrashCan.

Японское агентство морской науки и техники о Земле (JAMSTEC) собрало обширную электронную библиотеку глубоководных изображений J-EDI. На базе этой библиотеки исследователями [1] был построен набор данных TrashCan - это датасет для сегментации подводного мусора и животных. Он состоит из 7212 изображений, размечен двумя способами - instance и material. Первый способ

предполагает разметку мусора по типу объекта - например "контейнер" а обитателей по виду - "угорь" и т.п. Всего 22 класса. Второй фокусируется на материале мусора (напр. "пластик") и содержит 16 классов. Аннотации в этом наборе данных имеют стандартный формат для задачи сегментации (instance segmentation): маски, отмечающие, какие пиксели изображения содержат каждый объект.

К сожалению, вместе с датасетом был приведен только результат для одного алгоритма сегментации объектов - Mask-RCNN[2]. Для предоставления более широкого представления о возможной точности сегментации и детекции мы обучили четыре современных алгоритма: Mask-RCNN[2], YOLACT[3], Cascade Mask-RCNN[4], Hybrid Task Cascade(HTC)[5] для двух видов разметки.

В задачи сегментации (segm) объект считается правильно распознанным, когда отношение пересечения маски предсказания и маски разметки к их объединению (IoU - intersection over union) больше порогового. Используется усреднение для набора порогов от 50% до 95% с шагом 5%. Для задачи детекции (bbox) используются обрамляющие прямоугольники (bounding boxes) вместо масок. Метрика, использованная для сравнения методов также является стандартной - средняя точность среди всех классов на всех объектах всех изображений (mAP - mean average precision). Отдельно посчитаны mAP для порога IoU=0.5 ($AP_{.50}$), для порога IoU=0.75 ($AP_{.75}$), также для маленьких, средних и больших объектов (AP_S , AP_M , AP_L).

В названиях архитектур r50/r101 значит, что модель построена на базе глубокой нейронной сети ResNet-50/ResNet-101. x101 означает, что на базе ResNeXt-101. YOLACT работает для изображений, сжатых в 320x320 пикселей, остальные методы на изображениях в исходном размере (около 460x300). FPN означает, что использована версия с Feature Pyramid Network. Все алгоритмы построены и обучены с помощью библиотеки MMDetection[6], конфигурация запусков доступна в открытом доступе: https://github.com/evgpsi/mmdetection_trashcan

Видно, что добиться наилучшего качества получилось при помощи Mask R-CNN или Cascade Mask-RCNN с небольшой разницей в результате. К удивлению, это существенно лучше, чем заявленные в исходной статье результаты для той же сети (но реализованной при помощи Detectron 2). Для сегментации (segm): 0.33/0.32 против 0.30/0.28 (для instance и material соответственно) и 0.43/0.38 против 34.5/29.1 для bbox. Особо хочется отметить, что быстрый алгоритм YOLACT позволяет получить не такое плохое качество сегментации

Архитектура	Данные	Задача	AP	AP _{.50}	AP _{.75}	AP _S	AP _M	AP _L
YOLACT r101 320px	instance	segm	0.26	0.49	0.25	0.20	0.30	0.47
		bbox	0.36	0.58	0.39	0.27	0.40	0.52
	material	segm	0.22	0.45	0.19	0.20	0.24	0.35
		bbox	0.29	0.52	0.30	0.28	0.30	0.41
HTC w/o semantic segmentation r50 FPN	instance	segm	0.31	0.55	0.32	0.30	0.33	0.55
		bbox	0.41	0.61	0.44	0.37	0.42	0.64
	material	segm	0.29	0.53	0.27	0.26	0.30	0.41
		bbox	0.35	0.56	0.38	0.31	0.36	0.51
Cascade Mask- RCNN x101 64x4d FPN	instance	segm	0.32	0.57	0.33	0.29	0.36	0.56
		bbox	0.42	0.62	0.44	0.36	0.45	0.65
	material	segm	0.31	0.56	0.28	0.26	0.32	0.42
		bbox	0.37	0.58	0.40	0.29	0.38	0.53
Mask- RCNN x101 32x8d FPN	instance	segm	0.33	0.56	0.33	0.29	0.36	0.56
		bbox	0.43	0.62	0.47	0.32	0.47	0.66
	material	segm	0.32	0.58	0.30	0.27	0.33	0.43
		bbox	0.38	0.61	0.41	0.30	0.38	0.53

Таблица 1: Результаты детекции(bbox) и сегментации(segм) объектов на валидационном наборе данных датасета TrashCan

и может быть запущен в реальном времени на видеопотоке при использовании современных встраиваемых вычислителей.

Работа выполнена при поддержке РФФИ
в рамках научного проекта No. 19-31-90172.

Литература

1. Hong J., Fulton M., Sattar J. TrashCan: A Semantically-Segmented Dataset towards Visual Detection of Marine Debris. // ArXiv:2007.08097
2. He K., Gkioxari G., Dollár P., Girshick R. Mask R-CNN. // IEEE International Conference on Computer Vision (ICCV). 2017
3. Bolya D., Zhou C., Xiao F., Lee Y.J.. YOLACT: Real-time Instance Segmentation. // In ICCV 2019
4. Chen K., Pang J., Wang J., Xiong Y., Li X., Sun S., Feng W., Liu Z., Shi J., Ouyang W., Chen C., Lin . Hybrid task cascade for instance segmentation. // In CVPR 2019
5. Cai Z., Vasconcelos N. Cascade R-CNN: High Quality Object Detection and Instance Segmentation. // IEEE Transactions on Pattern Analysis and Machine Intelligence
6. Chen, K., Wang, J., Pang, J., Cao, Y. and others MMDetection: Open MMLab Detection Toolbox and Benchmark.// arXiv:1906.07155.