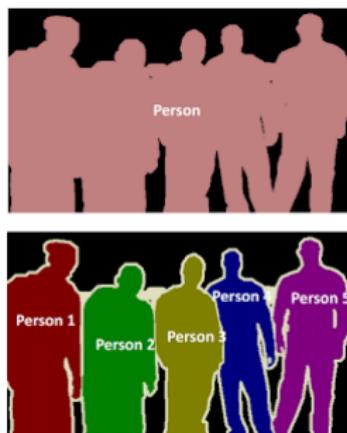


Сегментация экземпляров

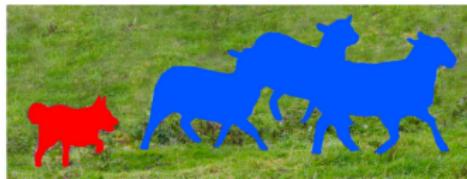
Виктор Китов
victorkitov.github.io



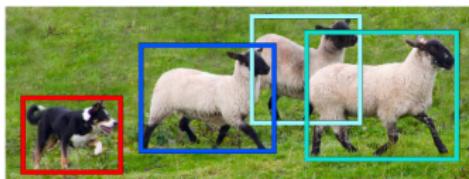
Задачи на изображениях³



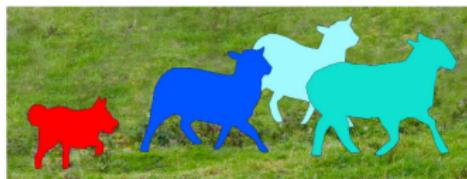
Image Recognition



Semantic Segmentation



Object Detection



Instance Segmentation

- Mask R-CNN: более точная 2x-стадийная instance-сегм-ция¹.
- YOLOACT: быстрая одностадийная instance-сегментация².

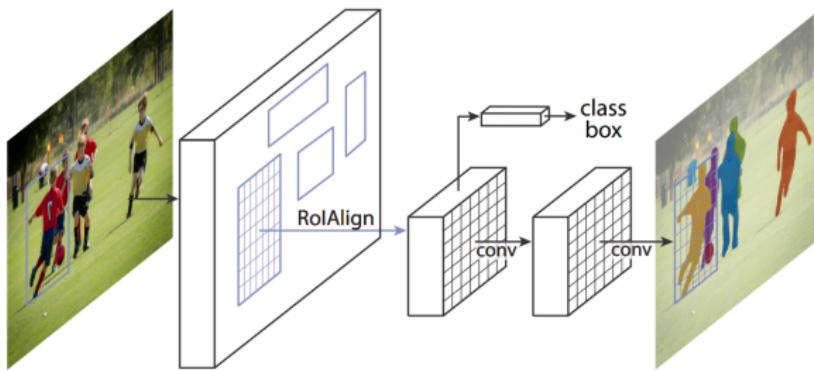
¹<https://arxiv.org/pdf/1703.06870.pdf>

²<https://arxiv.org/pdf/1904.02689.pdf>

³Видео об instance и panoptic сегментации.

Mask R-CNN: архитектура

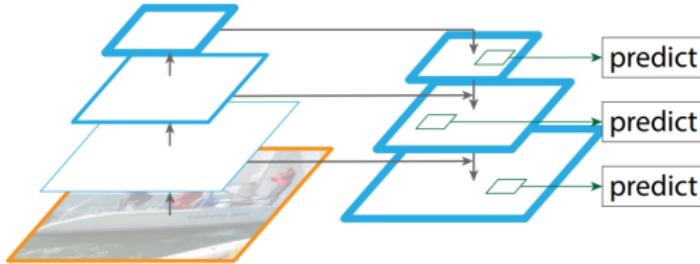
- Mask R-CNN - Faster R-CNN + ветка для предсказания маски объекта внутри каждой рамки.



- Маски выделяются независимо для каждого класса.
 - на выходе маска класса, предсказанного классификатором
- Функция потерь: потери классификации+потери локализации+потери выделения маски (кросс-энтропийные).

Feature Pyramid Network

- Карта признаков Mask R-CNN - из Feature Pyramid Network⁴:
 - в декодировщике все карты пр-ков высокоуровневые.



- Region Proposal Network применяется к каждому ярусу декодировщика.
 - обеспечивает выделение объектов разного размера.

⁴<https://arxiv.org/pdf/1612.03144.pdf>

ROI pool -> ROI align

- В Faster R-CNN регион интереса, извлекается Region-Proposal сетью.
 - Он не точно соответствует (укрупнённой) карте признаков => дискретизация.
 - Также дискретизация на этапе наложения сетки фикс. размера в пирамид. пулинге (ROIpool).
- Дискретизация ↓ точность выходной маски.
- Чтобы такого не было ROIpool заменяется на ROIalign: на регион интереса накладывается сетка фикс. размера, а значения признаков считаются через билинейную интерполяцию.

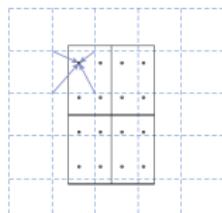
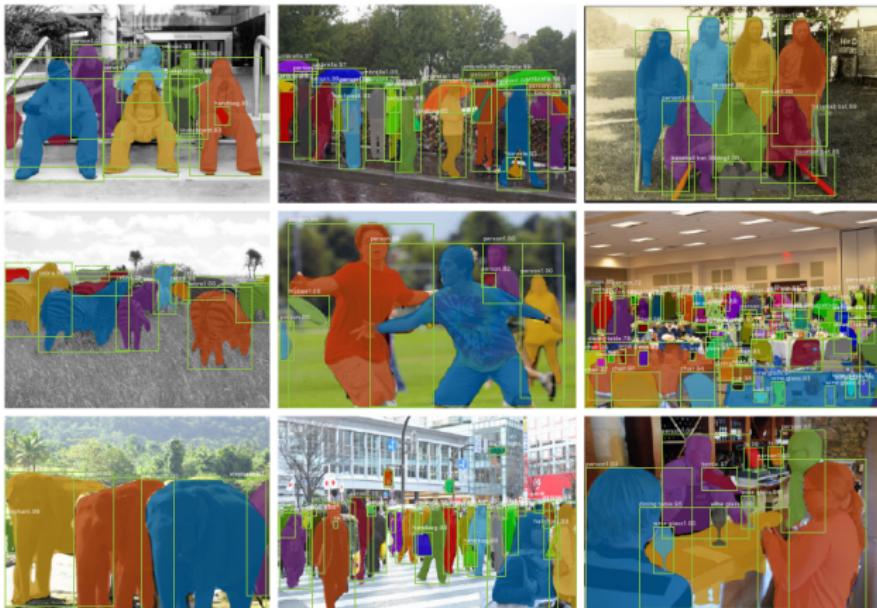


Figure 3. **RoIAlign:** The dashed grid represents a feature map, the solid lines an RoI (with 2×2 bins in this example), and the dots the 4 sampling points in each bin. RoIAlign computes the value of each sampling point by bilinear interpolation from the nearby grid points on the feature map. No quantization is performed on any coordinates involved in the RoI, its bins, or the sampling points.

Результаты работы

Результат - точная, но не real-time, instance-сегментация.



Оценка поз

Добавив к обработке регионов интереса K карт, предсказывающих в one-hot режиме K опорных узлов тела, можем оценивать позы (pose estimation):



YOLOACT

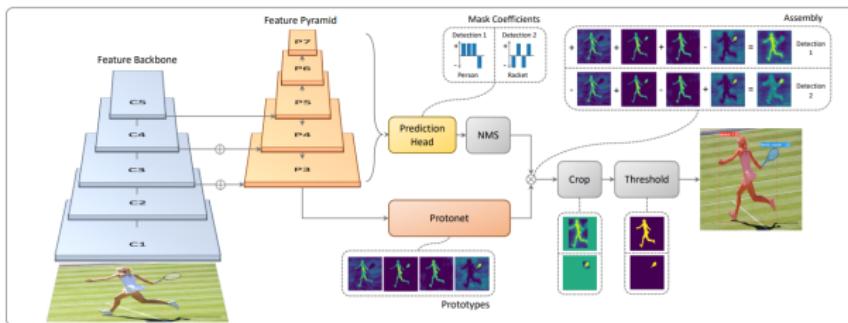


Figure 2: **YOLACT Architecture** Blue/yellow indicates low/high values in the prototypes, gray nodes indicate functions that are not trained, and $k = 4$ in this example. We base this architecture off of RetinaNet [27] using ResNet-101 + FPN.

- Protonet выделяет карты K ($=\text{const}$, гиперпараметр) прототипов
- Prediction head = классификатор & локализация для опорных регионов (anchor boxes) и K коэффициентов (\mathbb{R}^K), с которыми линейно комбинируются прототипы, чтобы получить итоговую маску внутри рамок.

Результаты работы



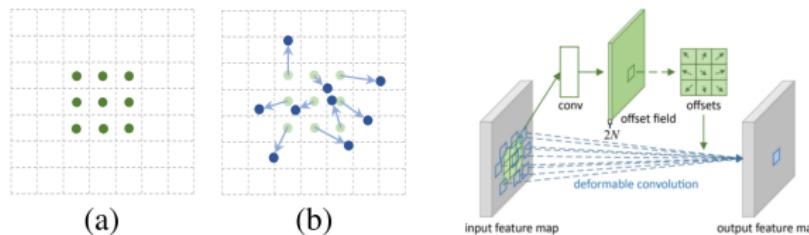
YOLOact не такая точная, как Mask R-CNN, но работает в 3 раза быстрее.

Содержание

1 Деформируемые преобразования

Деформируемые свёртки⁵

- Обычная свёртка (a) и деформируемая свёртка (deformable convolution, b):



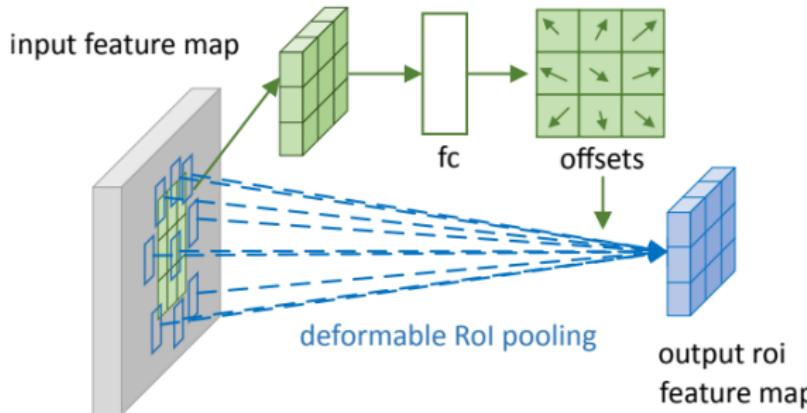
- Применяется обычная dilated свёртка, предсказывающая для каждой позиции смещения
 - Выход: из др. dilated свёртка с предсказанными смещениями, дающей выходную карту признаков.
- Поскольку смещения $\in \mathbb{R}$, то 2ая свёртка применяется к билинейно интерполированным активациям.

⁵<https://arxiv.org/pdf/1703.06211.pdf>

Деформируемый ROI pooling⁶

Деформируемый ROI pooling:

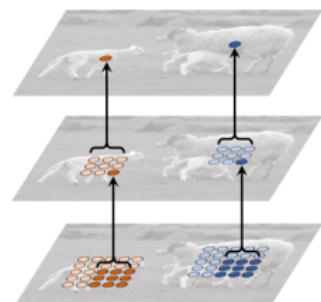
- ① Полносвязный слой, применённый к обычному ROI-пулингу генерирует смещения.
- ② Применяется новый ROI-пулинг к смещенным узлам сетки.



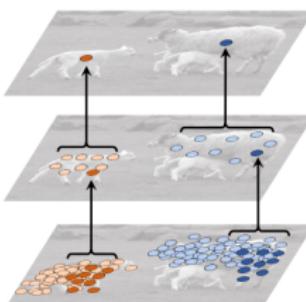
⁶<https://arxiv.org/pdf/1703.06211.pdf>

Результаты

- Замена свёрток и ROI пулингов на деформируемые приводит к ↑ точности сегментации и детекции.
 - лучше акцентируется на объектах



(a) standard convolution



(b) deformable convolution

