**Тема 3.5 Многоэтапные подходы к обнаружению объектов**

**План**

1. Архитектура SPPNet
2. Идея пирамидального пулинга
3. Архитектура Fast-RCNN
4. Архитектура Faster-R-CNN
5. Архитектура MASK-RCNN
6. Архитектура Feature Pyramidal Network

**Архитектура SPPNet**

Один из таких недостатков архитектуры R-CNN – это слишком большое время работы из-за необходимости оценки отдельно каждого из почти 2000 регионов кандидатов. Этот недостаток был решен в архитектуре Spatial Pyramidal Pooling Net или SPPNet. Ранее мы уже обсуждали идею **Spatial Pyramidal Pooling** в отношении задачи семантической сегментации. Основная идеяSPPNetзаключается в выделении признаков для всего изображения, а не для его отдельных частей и использовании пирамидального пулинга для поиска регионов кандидатов. Регионы кандидаты выбираются на каждой карте признаков при помощи селективного поиска.

**Идея пирамидального пулинга**

**Основная идея пирамидального пулинга** – это каждое изображение (регион кандидат) делится на заданное число частей для каждой из которых используется операция maxpooling. Возможно, для поиска объекта достаточно одного значения на весь регион – тогда достаточно глобального пулигна. Если нужна только четверть изображения – это тоже возможно, также мы выделяем 1/8 и 1/16 и так далее. Цель такого пулинга - это учет одного и того же объекта с разным контекстом.

**Архитектура Fast-RCNN**

Развитием идей R-CNN и подхода SPPNet стала архитектура Fast-R-CNN. Так же как и SPPnet cеть использует операции selective search для результата работы кодировщика признаков. Однако, затем вместо пирамидального пулинга сеть использует сжатие полученных регионов кандидатов до одного размера. Эту операцию назвали ROI Pooling или пулинг регионов канатов. По существу**, ROI Pooling** заключается в том, что каждый регион кандидат трансформируется к заданному размеру при помощи сжатия методом макс-пулинга. Подход Fast-RCNN позволил сократить время работы сети в порядка 20 раз по сравнению с R-CNN.

**Архитектура** **Faster-R-CNN**

Следующим шагом стала архитектура Faster-R-CNN. Основная идея Faster-R-CNN - использовать небольшую промежуточную сверточную сеть вместо selective search. Цель этой сети предварительное предсказания результатов по карте признаков. Такую сеть было предложено назвать сетью предложения регионов или **Region proposal network, RPN**. Результаты работы сети RPN – это значения координат и габаритов для регионов кандидатов. После работы RPN для каждого региона кандидата проводится ROIPooling и дальнейшая обработка аналогично Fast-RCNN. В оригинальной работе RPN – это сверточный слой с размером ядра RPN действует по результатам набора карт признаков. **Результат работы RPN** - это так называемые **анхоры** – регионы кандидаты. Для каждого анхора результаты векторизуются и по ним определяются: координаты региона и проводится бинарная классификация на наличие объекта в регионе. То есть оценивается лишь объектность региона, а не класс объекта. Для отсева лишних регионов кандидатов для результатов RPN проводится немаксимальное сжатие. Сеть RPN решает задачу дисбаланса классов – то есть задачу того, что большинство регионов кандидатов оказываются пустыми. В оригинальной работе для каждого предложения RPN предлагалось 9 анхоров отличающихся соотношением размеров сторон – так называемым aspect ratio и масштабом.

**Архитектура** **MASK-RCNN**

Подход Faster-RCNN оказался в несколько раз быстрее предыдущей версии. На основе подхода Faster-RCNN была **одна из наиболее успешных архитектур объектной сегментации – MASK-RCNN**. Главными отличиями MASK-RCNN от Faster-RCNN стали: дополнительная головная часть сегментации объектов в каждом регионе кандидате и замена операции ROI Pooling на **ROI Align**. Суть ROI Align сводится к замене операции макс-пулинга на билинейную интерполяцию. Этот подход более точен для случая небольших объектов, однако требует больших временных затрат.

**Архитектура Feature Pyramidal Network**

Одной из модификаций для данных архитектур стал подход пирамиды признаков или Feature Pyramidal Network, FPN. Основная идея FPN: использовать карты признаков, полученные на разных уровнях кодировщика.

**Резюме**

Архитектуры Faster-RCNN и MASK-RCNN как и использованиемFPN, так и без него до настоящего времени остаются одними из наиболее популярных решений задач обнаружения объектов и их сегментации соответственно. Такие архитектуры хорошо работают с объектами небольшого размера и с объектами в большом количестве на изображении. Однако, оба подхода имеют достаточно высокое время работы. Как правило для обоих подходов используют заранее предобученный кодировщик признаков на основе resnet.