

# АЛГОРИТМЫ ДЕТЕКЦИИ ОБЪЕКТОВ ДЛЯ АНАЛИЗА ИЗОБРАЖЕНИЙ

Хурсов П. С., Искра Н. А.  
Кафедра электронных вычислительных машин,  
Белорусский государственный университет информатики и радиоэлектроники  
Минск, Республика Беларусь  
E-mail: xursovp01@gmail.com, niskra@bsuir.by

*В данной статье рассматриваются актуальные подходы к решению задачи автоматической детекции объектов, проводится анализ их параметров работы и показателей эффективности, делаются выводы о возможности применения подобных алгоритмов в интеллектуальном анализе изображений.*

## ВВЕДЕНИЕ

В последние годы качественный скачок произошёл в области классификации изображений [1]. Дальнейшее развитие области анализа изображений происходит в двух направлениях: с одной стороны, это автоматическая детекция множества объектов с определением их меток и месторасположения [2], а с другой - семантическое описание, которое предполагает интерпретацию взаимосвязей объектов на изображении, например в виде текстового описания [3].

Решение задачи понимания и интерпретации изображения требует тесного взаимодействия этих направлений [4]. Так, например, на основании результатов автоматической детекции можно построить графовую модель, отражающую семантические связи между объектами [5].

В рамках данной работы мы рассматриваем актуальные подходы к автоматической детекции объектов как первого этапа в интеллектуальном анализе изображений [6].

## I. ОСНОВНЫЕ ПОДХОДЫ К ДЕТЕКЦИИ ОБЪЕКТОВ

Среди всех алгоритмов детекции объектов можно выделить три актуальных подхода к решению данной задачи.

Первым и наиболее вычислительно затратным подходом является метод скользящего окна. Основой данного подхода является пошаговый обход всего изображения, при котором извлекается некоторый регион изображения и происходит попытка классификации объекта, который попал в этот регион.

Следующим подходом является использование так называемых “предложений регионов” (Region Proposals). В этом подходе используется комбинация из двух алгоритмов. Первый из них делает множество предположений (регионов, в которых, вероятно, содержится какой-либо объект), а второй осуществляет классификацию данного объекта.

Третьим подходом является “детекция за один проход” (Single Shot Detection), при котором локализация и классификация происходят одновременно, т.е. за один проход алгоритма.

## II. МЕТОД СКОЛЬЗЯЩЕГО ОКНА

В этом классе алгоритмов из изображения последовательно извлекаются регионы прямоугольной формы. Так же для увеличения точности распознавания объектов разного размера может строиться пирамида изображений. После этого к каждому региону применяется классификатор. В зависимости от классификатора можно добиться либо высокой скорости выполнения (метод Виолы-Джонса, каскад Хаара, LBP каскады), либо высокой точности (различного вида нейронные сети).

## III. REGION PROPOSALS

Множество алгоритмов использующее данный подход образуют семейство Regions with CNN (R-CNN [7]). Первой стадией является составление “предложений регионов”. Для этого может применяться как отдельный алгоритм (в том числе и метод скользящего окна), так и часть классификатора (в случае со сверточными нейронными сетями). Например, в R-CNN в качестве алгоритма генерации предположений использовался селективный поиск. Данный алгоритм достаточно затратный, поэтому в Faster R-CNN [8] от него отказались. На смену селективному поиску пришла сеть Region Proposals Network (RPN). Это позволило ускорить выполнение алгоритма, поскольку RPN использует общие с классификатором ресурсы.

Следующей стадией является классификация. Для этого выбираются регионы полученные на первой стадии и происходит их классификация. В R-CNN для этого использовалась последовательность CNN и машин опорных векторов (SVM). Вместе с классификацией происходит уточнение координат объекта. Поскольку последовательность CNN + SVM применялась к каждому предположению, а их в среднем порядка 2000 на изображение, то это очень затратно. Поэтому в следующей версии R-CNN (Fast R-CNN [9]) пришли к концепции “регионы интереса” (Regions of Interests). Основной идеей является обработка всего изображения обученной сверточной нейронной сетью. После этого на выходе получается уменьшенная карта высокоуров-

невых признаков, которые хорошо подходят для классификации объектов. После этого из карты извлекаются предсказанные на первой стадии участки (Regions of Interests) и классифицируются. Это позволило ускорить выполнение алгоритма почти в 25 раз и увеличить среднюю точность.

В Faster R-CNN карта признаков используется так же и для генерации Region Proposals, что позволяет увеличить точность и скорость выполнения. Так же к особенностям Faster R-CNN можно отнести то, что данный алгоритм предсказывает несколько ограничивающих регионов прямоугольной формы разного размера и соотношения сторон (2:1, 1:1, 1:2), что позволяет точнее выделять объект.

В таблице 1 приведено сравнение алгоритмов семейства R-CNN по скорости и точности среднего предсказания для классов, при условии, что детекция считается положительной, если отношение площади пересечения предсказанного региона и региона, содержащего объект, к суммарной площади двух регионов больше 0.5(mAP@0.5).

Таблица 1 – Сравнение алгоритмов семейства R-CNN

Алгоритм	mAP@0.5(%)	Время выполнения(с.)
R-CNN	62.4	50
Fast R-CNN	70.0	2
Faster R-CNN	78.8	0.2

#### IV. SINGLE SHOT DETECTION

Основной идеей данного подхода является попытка одновременно предсказывать и координаты объекта, и его класс. Для этого используются сверточные нейронные сети, поскольку они позволяют уменьшить размерность входного изображения без особой потери информации и с сохранением локализации признаков (Рис. 1). Основными архитектурами использующими данный подход являются YOLO (You only look once) [10] и SSD (Single shot detector) [11].

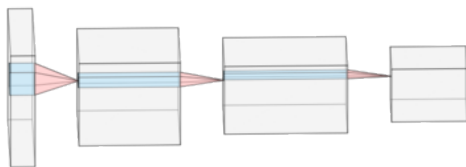


Рис. 1 – Сохранение локализации при свертке

YOLO делит изображение на сетку NxN. Если центр объекта попадает в некоторую ячейку сетки, то эта ячейка ответственна за детекцию данного объекта. Для этого YOLO использует обученную сверточную нейронную сеть.

SSD использует похожий подход, однако использует не только выходной слой сверточной

нейронной сети, но и множество промежуточных слоев для построения предсказаний. Так же SSD использует сверточные слои с ядрами разных размеров, что позволяет лучше распознавать объекты разной величины.

Оба подхода обеспечивают высокую скорость выполнения алгоритма и неплохую точность (Таблица 2).

Таблица 2 – Сравнение архитектур SSD

Алгоритм	mAP@0.5(%)	Время выполнения(мс.)
YOLO	63.7	22
YOLOv3	78.6	15
SSD300	79.6	21
SSD512	81.5	50

#### V. ВЫВОД

Проведя сравнение трех основных подходов, можно сделать вывод о целесообразности использования архитектуры SSD512 для детекции объектов в интеллектуальном анализе изображений, поскольку помимо высокой скорости выполнения она имеет неплохую точность. Однако если скорость исполнения не важна, то можно применить метод скользящего окна и пирамиды изображений, что обеспечит еще большую точность. Так же хорошим выбором может быть и Faster R-CNN, т. к. она позволяет определить не только местоположение объекта, но и сделать более точное предсказание региона, содержащего объект.

#### СПИСОК ЛИТЕРАТУРЫ

1. ImageNet Large Scale Visual Recognition Challenge /O. Russakovsky [et.al.] // IJCV – 2015. – Vol. 115, № 3. – P. 211–252.
2. Recent Advances in Convolutional Neural Networks / J. Gu [et.al.] // Pattern Recognition. – 2018. – Vol. 77. – P. 354–377.
3. A Comprehensive Survey of Deep Learning for Image Captioning / M. Z. Hossain [et.al.] // CoRR – 2018.
4. Johnson, J. DenseCap: Fully Convolutional Localization Networks for Dense Captioning / J. Johnson, A. Karpathy, L. Fei-Fei // CCVPR – 2016.
5. Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations / R. Krishna [et.al.] // IJCV – 2016. – Vol. 123. – P. 32–73.
6. Iskra, N. Neural network based image understanding with ontological approach / N. Iskra, V. Iskra, M. Lukashevich // Open semantic technologies for intelligent systems (OSTIS-2019): materials of IX International.sc.-tech.conf. – Mn.: BSUIR, 2019. – Iss. 3. – P.113-122.
7. Rich feature hierarchies for accurate object detection and semantic segmentation /R. Girshick [et.al.] // CVPR – 2014. – P. 580–587.
8. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks / S. Ren [et.al.] // NIPS – 2015. – P. 91–99.
9. Fast R-CNN / R. Girshick // ICCV – 2015. – P. 1440–1448.
10. You Only Look Once: Unified, Real-Time Object Detection / J. Redmon [et.al.] // CVPR – 2016. – P. 779–788.
11. SSD: Single Shot MultiBox Detector / W. Liu [et.al.] // ECCV – 2016. – P. 21–37.