УДК 004

ОБНАРУЖЕНИЕ ОБЪЕКТОВ НА ИЗОБРАЖЕНИЯХ С ПОМОЩЬЮ НЕЙРОННЫХ СЕТЕЙ

Порохин Ю.М., студент гр. ИТб-221, III курс

Научный руководитель: Сыркин И.С., к.т.н., доцент

Кузбасский государственный технический университет имени Т.Ф. Горбачева, г. Кемерово

Обнаружение объектов (Object Detection) является одним из наиболее востребованных разделов компьютерного зрения. Эти технологии имеют применение во многих сферах жизни и бизнеса, позволяют сделать их проще, дешевле, безопаснее. Область применения обширна, решает задачи, такие как видеонаблюдение, автономное вождение, отслеживание объектов и дополненная реальность. Глубокое обучение, в частности сверточные нейронные сети (CNN) позволило добиться значительных успехов в точности и эффективности обнаружения, в сравнении с классически машинным обучением.

Одним из первых подходов, применяемым для определения наличия объекта на картинке является двухстадийный детектор R-CNN (Region Convolution Network). Архитектура метода состоит из следующих последовательно выполняемых шагов и проиллюстрирована на рисунке 1 [1]:

1. Определение набора гипотез – выделение небольших участков изображения, которые возможно содержат искомые объекты;
2. Извлечение из предполагаемых признаков с помощью сверточной нейронной сети и их кодирование в вектор – каждая гипотеза из предыдущего шага независимо и по отдельности друг от друга поступает на вход сверточной нейронной сети архитектуры AlexNet без последнего softmax-слоя, задачей которой служит кодирование поступающего изображения в 4096-размерное векторное представление. Исходное изображение корректируется под размер 3×227×227 с помощью сглаживания или растягивания входа до нужного размера;
3. Классификация объекта внутри гипотезы на основе шага 2 – вектор признаков, сгенерированный сверточной сетью поступает для обработки в Support Vector Machine (SVM) или машины опорных вектором, которая обучается независимо для каждого класса объектов и на выходе выдает доверительный балл, указывающий на вероятность присутствия объекта в данном участке;

4. Улучшение (корректировка) координат гипотезы – гипотезы, полученные на шаге 1 не всегда содержат правильные координаты, поэтому проводится дополнительная проверка с обработкой линейной регрессией, которая приносит дополнительные 3-4% при прохождении метрик.
5. Все повторяется, начиная с шага 2, пока не будут обработаны все гипотезы с шага 1.

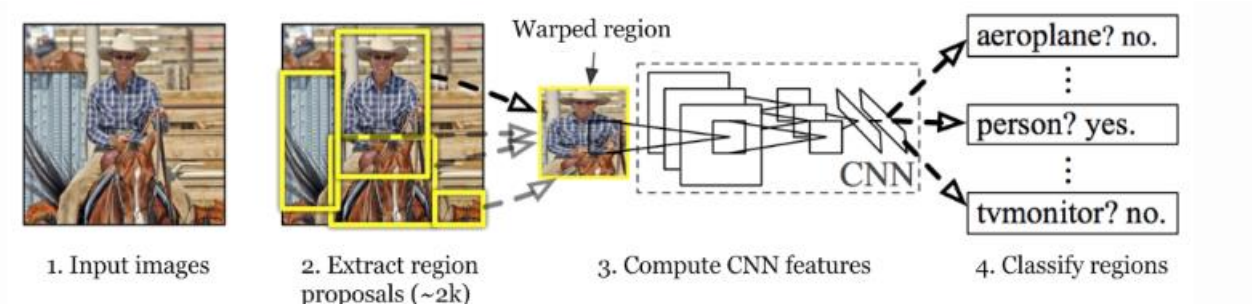

Рисунок 1 – Архитектура R-CNN.

Таким образом, можно выделить следующие достоинства архитектуры R-CNN:

- Хорошее качество обнаружения объектов, в сравнении с традиционными методами;
- Гибкость, для извлечения признаков можно использовать как предобученные сверточные сети, так и самостоятельно обучить их для нужных в конкретной задаче объектов;

В то же время архитектура имеет и недостатки, такие как:

- Медленная скорость работы, для каждой гипотезы требуется отдельный прогон через сверточную сеть, так на одно изображение уходит примерно 2000 прогонов;
- Необходимы большие объемы памяти, т.к. для каждой гипотезы извлекаются признаки и хранятся отдельно;
- Многоэтапный процесс обучения, необходимо обучить CNN, SVM для классификации и регрессионную модель для уточнения координат;
- Плохая масштабируемость, модель не подходит для обнаружения объектов в реальном времени и работы с видео, т.к. имеет слишком низкую скорость обработки.

R-CNN стала первым шагом в использовании глубокого обучения для обнаружения объектов и в последующем претерпела ряд модификаций, так появились более эффективные Fast R-CNN, Faster R-CNN, Mask R-CNN. Объединив глубокое обучение с анализом на основе гипотез, R-CNN установила новый

стандарт в области обнаружения объектов и открыла возможности для различных приложений. Так R-CNN использовалась для обнаружения и классификации различных типов опухолей на снимках МРТ и КТ, что позволило повысить точность диагностики и выявления злокачественных опухолей на ранних стадиях.

В последующем некоторые идеи из двухстадийных детекторов проложили путь к созданию одностадийных детекторов, а именно одной из самых популярных архитектур YOLO.

В 2015 году Joseph Redmon опубликовал статью You Only Look Once: Unified, Real-Time Object Detection. В статье описывается одностадийная архитектура модели для обнаружения объектов с полностью интегрированным подходом, который позволяет проводить обнаружение одним проходом. Такой способ позволяет обрабатывать изображение целиком, что существенно повышает скорость работы и позволяет использовать модель для задач обнаружения объектов в режиме реального времени. Архитектура модели представлена на рисунке 2 [2].

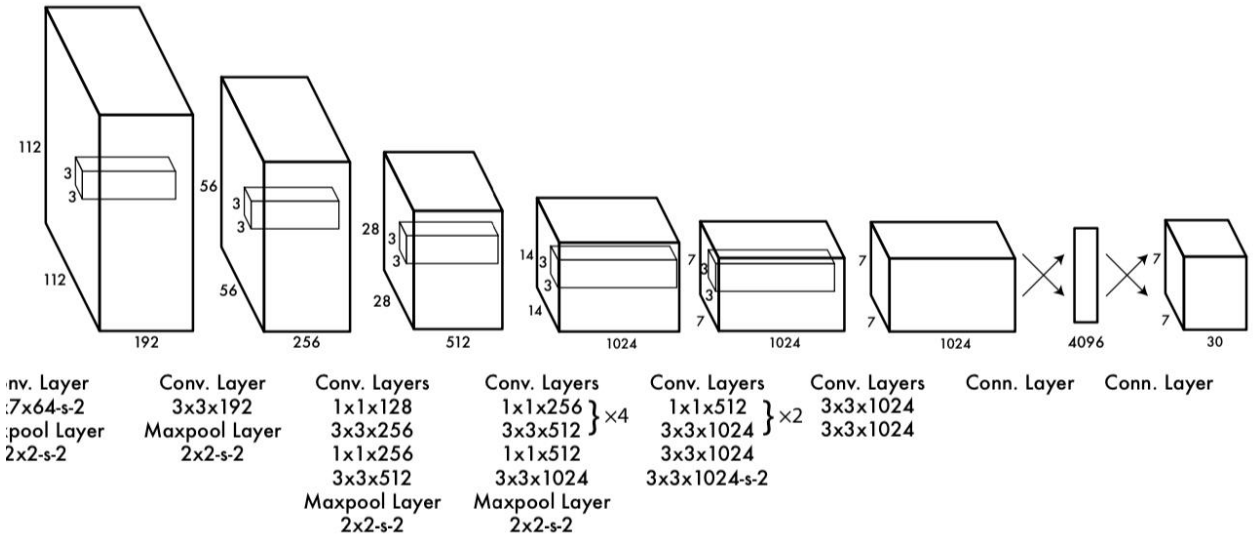

Рисунок 2 – Архитектура YOLOv1.

YOLOv1 использует глубокую сверточную модель, состоящую из 24 сверточных слоев, за которыми следуют 2 полносвязных слоя. Первые 20 слоев представляют из себя классификатор, обученный на датасете ImageNET (1000 classes), вошедший в top-5 Accuracy с 88% при валидации на ImageNET 2012 validation set. Также в модели используются 4 необученных сверточных слоя и 2 полносвязных слоя со случайно инициализированными весами, а также функции потерь для оценки результатов.

В конце прогона модель возвращает тензор, в котором находится информация об ограничивающих рамок (области на изображении с объектом) и соответствующих классах. Такой выход позволяет одновременно учитывать пространственную информацию и принадлежности к классам, что является ключевым моментом в архитектуре YOLO.

В качестве функции потерь при обучении модели используется суммарная квадратичная ошибка, включающаяся в себя:

- Ошибка координат – сумма квадратов разницы между предсказанными и истинными координатами ограничивающих рамок;
- Ошибка достоверности – измеряет разницу между предсказанной оценкой достоверности и истинным значением;
- Ошибка классификации – рассчитывается для ячеек, в которых присутствует объект, отражает расхождение между предсказанными и истинными вероятностями классов.

Таким образом, можно выделить следующие преимущества архитектуры YOLO:

- Одностадийный принцип работы, обнаружение объектов осуществляется за один проход через сеть, что упрощает архитектуру и повышает скорость работы;
- Глобальный контекст, обработка всего изображения позволяет учитывать контекст, снижая число ошибок, связанных с частичными признаками;
- Применимость сети для решения задач в реальном времени, высокая скорость обработки позволяет применять модель для задач, требующих оперативной обработки данных.

В то же время архитектура имеет следующие недостатки:

- Локализация объектов, фиксированное разбиение изображения на сетку может привести к проблемам при обнаружении объектов, расположенных близко друг к другу или имеющих малый размер;
- Ошибка координат, модель может допускать ошибки в определении точных границ объектов, особенно если они занимают несколько ячеек;
- Ошибки обобщающей способности, некоторые виды объектов, особенно редкие или сильно изменяющиеся по форме, могут быть предсказаны менее точно.

YOLOv1 стала поворотным моментом в области обнаружения объектов, предложив новый подход, где задача обнаружения формулируется как задача регрессии. Благодаря своей одностадийной архитектуре модель YOLOv1 и последующие модифицированные модели стали основой для задач обнаружения

объектов в реальном времени. Модели нашли свое применение в видеонаблюдении, робототехнике и беспилотном управлении.

Сравнивая YOLO и R-CNN, можно сделать вывод о том, что модели хороши в разных задачах. R-CNN обеспечивает высокий уровень точности, но требует много времени на обработку, в то время как YOLO имеет малое время обработки, но более низкую точность в некоторых областях применения. Так R-CNN получило применение в медицинской сфере, где точность гораздо важнее быстрого результата, а YOLO стала применяться в системах безопасности, робототехнике и беспилотных системах управления.

Список литературы.

1. R-CNN – Region-Based Convolutional Neural Networks [Электронный ресурс] // URL: https://www.geeksforgeeks.org/r-cnn-region-based-cnns/ (дата обращения 07.03.2025).
2. Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi. You Only Look Once: Unified, Real-Time Object Detection – Текст: электронный [Электронный ресурс] // URL: https://arxiv.org/pdf/1506.02640 (дата обращения 07.03.2025).
3. Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, Li Fei-Fei. ImageNet Large Scale Visual Recognition Challenge – Текст: электронный [Электронный ресурс] // URL: https://arxiv.org/pdf/1409.0575 (дата обращения 07.03.2025).