



Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технический университет
имени Н. Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н. Э. Баумана)

ФАКУЛЬТЕТ «Информатика и системы управления»

КАФЕДРА «Программное обеспечение ЭВМ и информационные технологии»

РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

К НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

НА ТЕМУ:

«Сравнение алгоритмов поиска объектов на изображениях с использованием различных модификаций сверточных нейронных сетей»

Студент ИУ7-71Б
(Группа)

(Подпись, дата)

Постнов С. А.
(Фамилия И. О.)

Руководитель НИР

(Подпись, дата)

Кузнецова О. В.
(Фамилия И. О.)

2024 г.

СОДЕРЖАНИЕ

ОПРЕДЕЛЕНИЯ	4
ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ	5
ВВЕДЕНИЕ	6
1 Описание предметной области	7
1.1 Задача поиска объекта на изображении	7
1.2 Сверточные нейронные сети для поиска объектов на изображениях	7
1.2.1 YOLO	9
1.2.2 R-CNN	11
1.2.3 Fast R-CNN	13
1.2.4 Faster R-CNN	15
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	17

ОПРЕДЕЛЕНИЯ

В настоящей расчетно-пояснительной записке применяют следующие термины с соответствующими определениями.

Нейронная сеть — математическая модель, а также её программное или аппаратное воплощение, построенная по принципу организации биологических нейронных сетей.

ОБОЗНАЧЕНИЯ И СОКРАЩЕНИЯ

В настоящей расчетно-пояснительной записке применяют следующие сокращения и обозначения.

МИ — Медицинские изображения

CNN — Convolutional neural network

RPN — Region Proposal Network

YOLO — You Only Look Once

IOU — Intersection Over Union

FPS — Frames Per Second

R-CNN — Regional Convolutional Neural Networks

SVM — Support Vector Machine

ВВЕДЕНИЕ

Технологии компьютерного зрения и искусственного интеллекта находят применение в разных сферах человеческой деятельности. Важным и интересным направлением, где возможно применение данных технологий, является анализ объектов на медицинских изображениях. На сегодняшний день анализ МИ и поиск объектов на них широко применяется в медицинской диагностике – от анализа крови до магниторезонансной томографии. До недавнего времени задачи анализа МИ решались с использованием различных алгоритмов, основанных на использовании гистограмм градиентов, алгоритмов каскадных классификаторов на основе метода Виолы – Джонса, алгоритмов, основанных на методах контурного анализа и др. Традиционные методы анализа МИ и поиска на них объектов достигли своего предела производительности. Аналогично медицинской сфере, подход распознавания объектов с использованием нейронных сетей нашел свое применение и в задачах мониторинга морского дна [1; 2].

Для решения задачи распознавания объектов зачастую выбирают сверточные нейронные сети из-за простоты реализации, минимальных системных требований и хорошего процента распознавания объектов. Сверточная нейронная сеть (CNN) – частный случай искусственных нейронных сетей глубокого обучения. Архитектура сверточных сетей была предложена Яном Лекуном в 1988 году с целью повышения эффективности распознавания образов [3; 4].

Целью работы является сравнение алгоритмов поиска объектов на изображениях с использованием различных модификаций сверточных нейронных сетей. Для достижения поставленной цели необходимо решить следующие задачи:

- 1) провести анализ предметной области алгоритмов поиска объектов на изображениях;
- 2) описать основные подходы к решению задачи распознавания объектов на изображениях;
- 3) сформулировать критерии сравнения применяемых методов и выполнить их сравнение.

1 Описание предметной области

1.1 Задача поиска объекта на изображении

Задача поиска объекта на изображении сводится к решению следующих подзадач [5]:

- 1) сегментация — выделение участков изображения, которые относятся к разным объектам;
- 2) классификация — определение типа объекта, для каждого выделенного сегмента отдельно.

Таким образом, обнаружение объектов — это процесс сегментации и классификации объектов в изображении.

1.2 Сверточные нейронные сети для поиска объектов на изображениях

Сверточные нейронные сети являются наиболее распространенным алгоритмом глубокого обучения, применяющим несколько сверточных слоев и вычислений. Они предоставляют эффективные способы извлечения признаков, а также являются лучшим выбором для решения проблем обнаружения объектов. Текущие подходы с использованием методов глубокого обучения для задач классификации и регрессии объектов можно разделить на две категории [6]:

- 1) двухэтапные методы, которые представлены такими архитектурами, как R-CNN, Fast R-CNN и Faster R-CNN;
- 2) одноэтапные методы, представленные различными версиями YOLO и др.

Описанные методы представлены на рисунке 1.1.

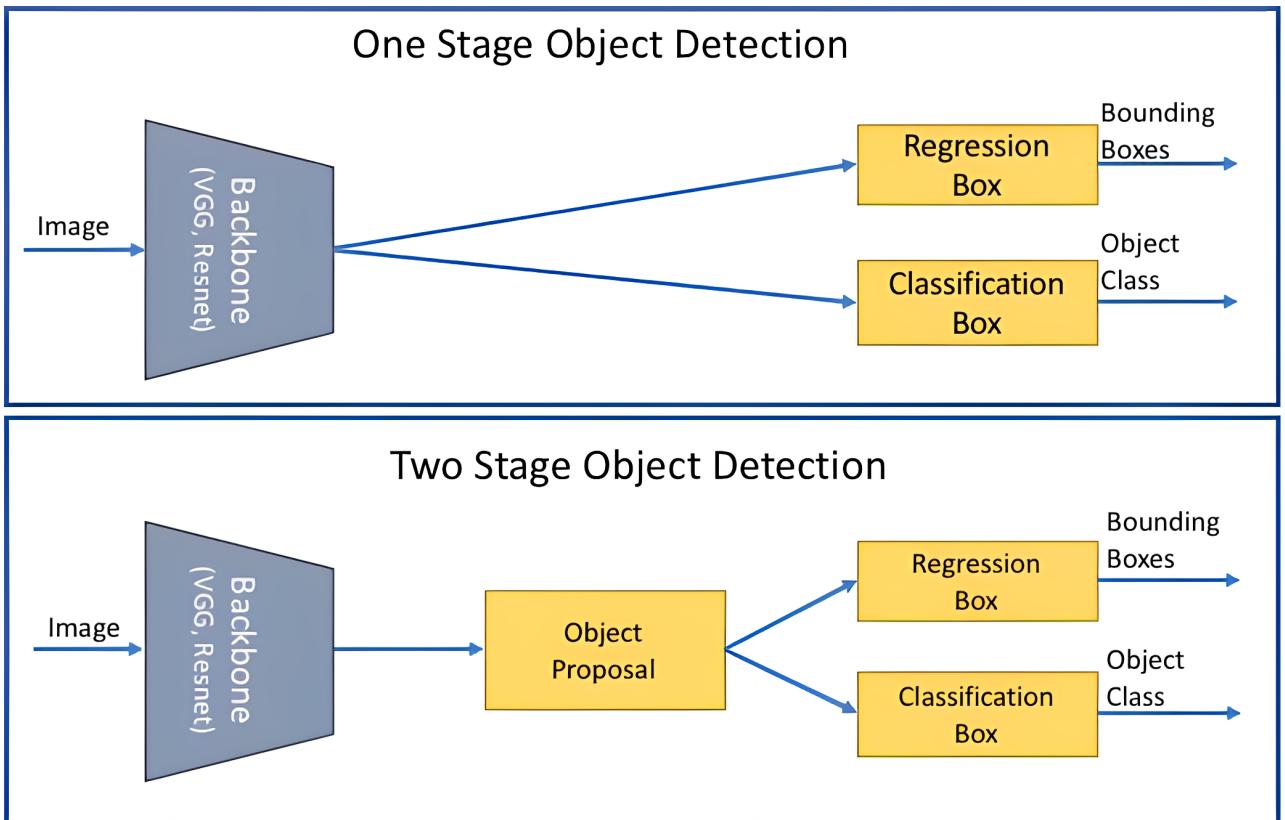


Рисунок 1.1 – Двухэтапный и одноэтапный методы

В двухэтапных методах используется селективный поиск или сеть региональных предположений (англ. RPN) для выделения областей, с высокой вероятностью содержащих внутри себя объекты. Затем, при помощи классификатора, определяется класс объекта, а при помощи регрессора определяются ограничивающие рамки. Данный метод обладает высокой точностью, но при этом ограничен в скорости обнаружения.

Одноэтапные методы не используют отдельную сеть для генерации регионов и основываются на методах регрессии, просматривая изображения целиком. Так как данные алгоритмы не используют RPN, скорость обнаружения выше, но точность выделения, в особенности малых объектов, не такая высокая, как у двухэтапных методов [6].

1.2.1 YOLO

YOLO — сеть, предназначенная для идентификации и распознавания объектов на изображениях в реальном времени. Такой подход к обнаружению объектов называется «Вы смотрите только один раз» (YOLO), что означает распознавание объектов сразу после первого прохода по изображению. Метод YOLO рассматривает обнаружение объектов как задачу регрессии с пространственно разделенными ограничивающими рамками и соответствующими вероятностями классов, которые прогнозируются с помощью одной нейронной сети на основе полных изображений в ходе одной оценки. YOLO быстра по своей конструкции и действительно работает в режиме реального времени, сохраняя большую точность [6; 7].

Базовая модель YOLO также называется YOLO версии 1 (YOLOv1). Она решает задачу обнаружения объектов на изображении как задачу регрессии. Одна сверточная сеть одновременно предсказывает множество ограничивающих рамок и вероятности классов для этих рамок. YOLOv1 разбивает входное изображение на сетку $S \times S$. Если центр объекта попадает в ячейку сетки, эта ячейка отвечает за обнаружение этого объекта. Каждая ячейка сетки предсказывает B ограничивающих рамок, показатели достоверности для этих рамок и вероятности класса C для сетки. Эти прогнозы закодированы в виде тензора $S \times S \times (B \times 5 + C)$. В процессе тестирования YOLOv1 умножает условные вероятности классов и прогнозы достоверности отдельных блоков, которые дают оценки для каждого блока, относящиеся к конкретному классу по формуле 1.1 [7].

$$Pr(Class_i|Object) \times Pr(Object) \times IOU_{pred}^{truth} = Pr(Class_i) \times IOU_{pred}^{truth} \quad (1.1)$$

Оценки отражают вероятность появления этого класса в поле и схожесть поля с объектом. Каждое ограничивающее поле состоит из 5 прогнозов: x , y , w , h и достоверности. Координаты (x, y) представляют центр прямоугольника относительно границ ячейки сетки. Ширина w и высота h рассчитываются относительно всего изображения. Именно поэтому YOLOv1 использует выражение $B \times 5$ для вычисления тензора. Сеть YOLOv1 состоит из 24 сверточных слоев, за которыми следуют 2 полностью соединенных слоя. Вместо начальных модулей, используемых в GoogLeNet, YOLOv1 использует

слой сокращения 1×1 , за которым следуют сверточные слои 3×3 . В Pascal VOC2007 YOLOv1 обрабатывает изображения со скоростью 45 кадров в секунду (FPS), что в 2 — 9 раз быстрее, чем у Faster R-CNN. В частности, Fast YOLO, быстрая версия YOLO, разработанная для расширения возможностей быстрого обнаружения объектов, достигает 155 кадров в секунду (FPS) [6; 7].

Архитектура YOLOv1 CNN представлена на рисунке 1.2.

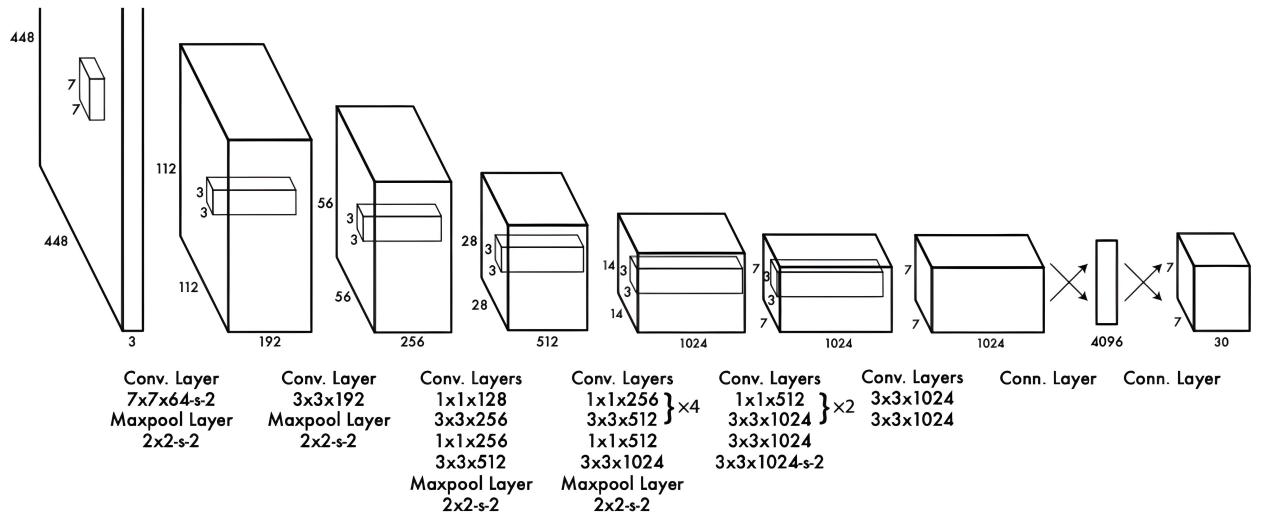


Рисунок 1.2 – Архитектура YOLOv1 CNN

YOLO версии 2 (YOLOv2) — значительно улучшенная модель YOLO, которая сохраняет преимущество в скорости и пытается повысить качество распознавания по сравнению с YOLOv1. Используя новый многомасштабный метод обучения, одна и та же модель YOLOv2 может работать в разных размерностях, предлагая простой компромисс между скоростью и качеством. Алгоритм лучше справляется с небольшими объектами и реагирует быстрее, чем ранее доступные версии. Одноступенчатая архитектура предполагает наличие только одной нейронной сети для прогнозирования ограничивающей рамки и вероятности категории. YOLOv2 имеет множество улучшений по сравнению со своими предшественниками и другими алгоритмами. В первую очередь, YOLOv2 использует Darknet-19 с 19 сверточными слоями и 5 слоями Max-Pooling и вводит якорные рамки (anchor boxes), что позволяет лучше адаптироваться к объектам разных размеров. Вместо абсолютных значений координат предсказываются как смещения относительно якорных рамок, для каждой из которых модель предсказывает параметры координат, уверенность и вероятности классов [8; 9].

Архитектура YOLOv2 CNN представлена на рисунке 1.3.

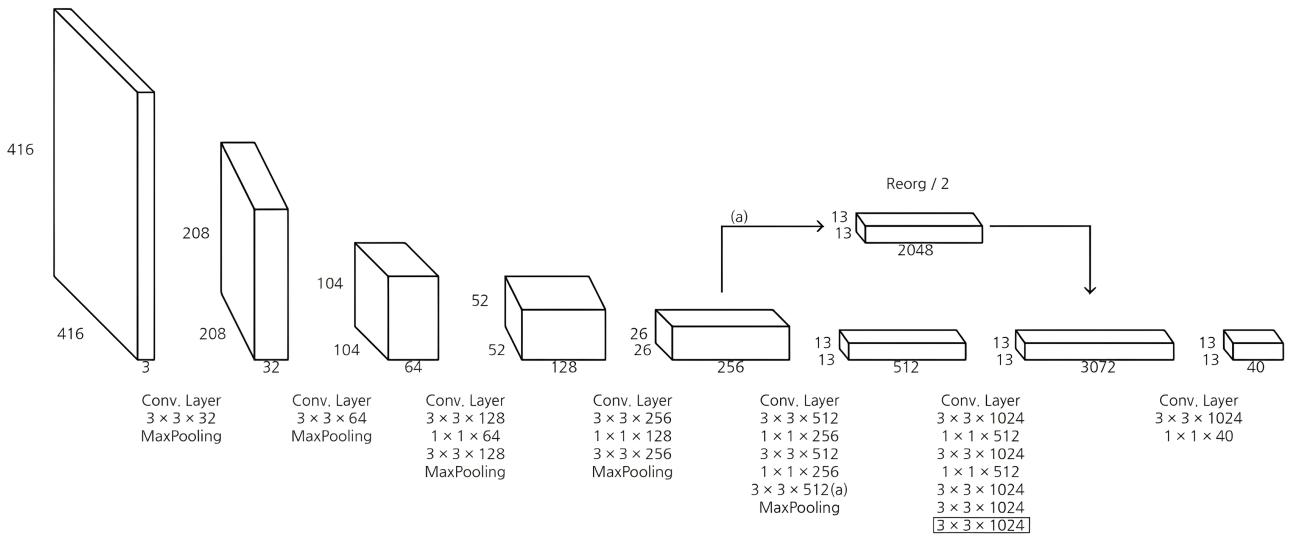


Рисунок 1.3 – Архитектура YOLOv2 CNN

1.2.2 R-CNN

В последнее время для решения задачи поиска объектов на изображении широкое распространение получили алгоритмы, основанные на применении региональных глубоких сверточных нейронных сетей или Regional Convolutional Neural Networks (R-CNN), которые принципиально ориентированы на решение задачи поиска объектов с одновременной их классификацией. Исходная реализация таких моделей базируются на использовании специальных алгоритмов предобработки — алгоритмов region-proposal-function, обеспечивающих предложение так называемых областей внимания, в которых потенциально могут находиться интересующие объекты. Описанный подход предлагает сократить вычислительные затраты, а также позволяют добиться минимального времени определения местоположения объекта и высокой точности его классификации. К настоящему моменту имеется большое количество вариантов реализации подобных алгоритмов, которые достигли хороших показателей по данным критериям [10].

Алгоритм работы R-CNN состоит из следующих основных шагов:

- 1) выполняется генерация областей интереса (region proposals), предположительно содержащих в себе искомые объекты (обычно до 2000 возможных областей) с использованием различных алгоритмов, предназначенных для снижения вычислительной сложности

обнаружения объектов на изображении (например, алгоритмы Edge Boxes, Selective search);

- 2) выполняется формирование карты признаков для исходного изображения путем аффинных преобразований, и каждая область интереса преобразуется в квадрат 227×227 , так как используемая архитектура CNN требует входы фиксированного размера 227×227 пикселей;
- 3) выполняется классификация объектов для каждой области интереса с использованием сформированного вектора признаков на основе метода опорных векторов (SVM).

Для оценки качества классификации, аналогично модели YOLO, используется показатель *IOU*. Считается, что объект обнаружен правильно, если данный показатель превышает некоторый порог, в противном случае считается, что объект не обнаружен [8; 10].

Схема алгоритма работы R-CNN представлена на рисунке 1.4.



Рисунок 1.4 – Схема алгоритма работы R-CNN

Пример распознавания с использованием R-CNN представлен на рисунке 1.5.

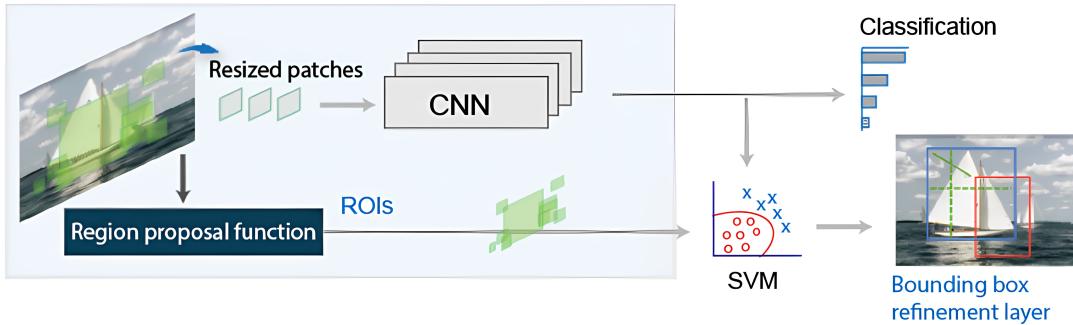


Рисунок 1.5 – Пример распознавания с использованием R-CNN

1.2.3 Fast R-CNN

Модель Fast R-CNN является улучшением классической модели R-CNN, которая известна своей низкой производительностью, особенно при взаимодействии с более глубокими сетями. Также, при использовании R-CNN затрачивается большой объем памяти для хранения данных.

Алгоритм работы Fast R-CNN состоит из следующих основных шагов [8; 10]:

- 1) аналогично R-CNN выполняется генерация областей интереса (region proposals);
- 2) выполняется формирование карты признаков для исходного изображения, но, в отличие от R-CNN, на вход нейронной сети CNN подается полное исходное изображение с последним слоем ROI pooling вместо Max pooling;
- 3) выполняется уточнение границ области интереса при помощи регрессионной модели (Bounding Box Regression);
- 4) выполняется классификация объектов, содержащихся в предполагаемых областях интереса с использованием слоя softmax с $K + 1$ выходами.

Схема алгоритма работы Fast R-CNN представлена на рисунке 1.6.

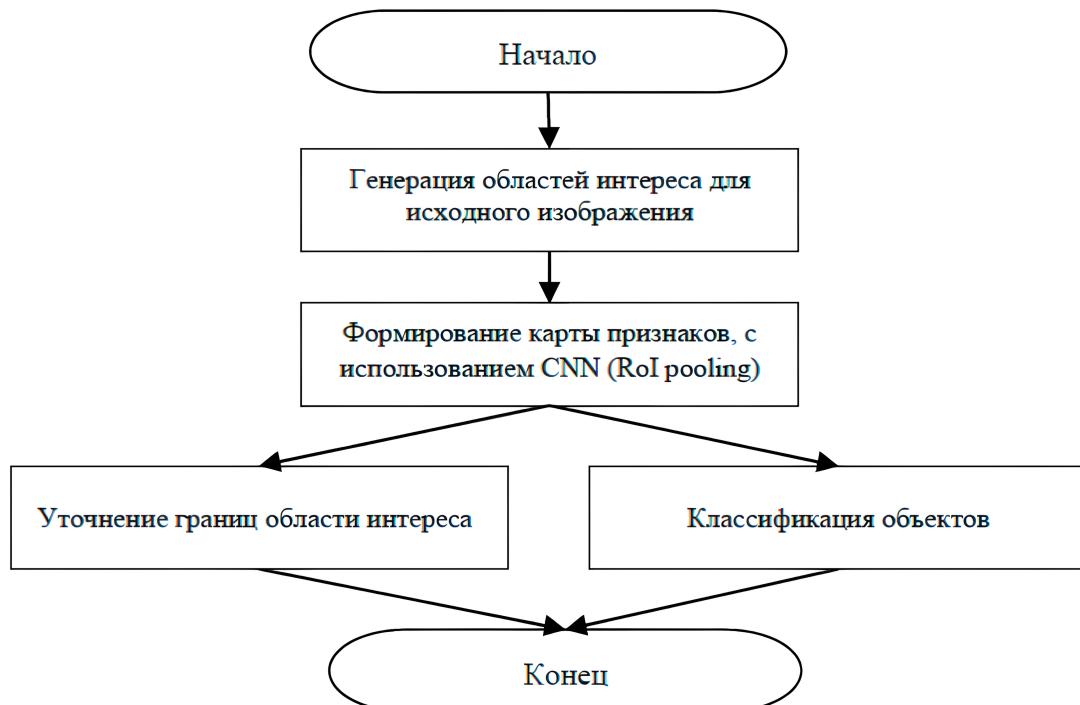


Рисунок 1.6 – Схема алгоритма работы Fast R-CNN

Пример распознавания с использованием Fast R-CNN представлен на рисунке 1.7.

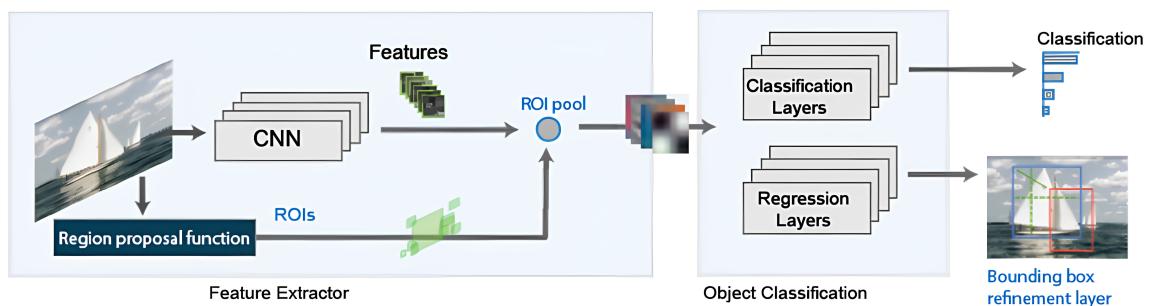


Рисунок 1.7 – Пример распознавания с использованием Fast R-CNN

1.2.4 Faster R-CNN

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. *A. B. Руденко, M. A. Руденко, И. Л. Каширина.* Применение искусственных нейронных сетей для поиска объектов на медицинских изображениях // Моделирование, оптимизация и информационные технологии. — 2024. — С. 480.
2. *B. С. Быкова, A. И. Машошин, A. С. Смирнов.* Способ распознавания назначенного донного объектах. — 2024.
3. *Д. А. Дасаева, В. В. Мокшин.* ПРИМЕНЕНИЕ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ПОИСКА ОБЪЕКТОВ НА ИЗОБРАЖЕНИЯХ // ИНФОРМАЦИОННЫЕ СИСТЕМЫ И ТЕХНОЛОГИИ. — 2021.
4. *И. И. Багаев.* АНАЛИЗ ПОНЯТИЙ НЕЙРОННАЯ СЕТЬ И СВЕРТОЧНАЯ НЕЙРОННАЯ СЕТЬ, ОБУЧЕНИЕ СВЕРТОЧНОЙ НЕЙРОСЕТИ ПРИ ПОМОЩИ МОДУЛЯ TENSORFLOW // Математическое и программное обеспечение в промышленной и социальной сферах. — 2020.
5. *E. Ю. Митрофанова.* Поиск объектов на изображениях с использованием TensorFlow Object Detection. — 2022.
6. *M. C. Тимошкин, A. H. Миронов, A. С. Леонтьев.* СРАВНЕНИЕ YOLOV5 И FASTERR-CNN ДЛЯ ОБНАРУЖЕНИЯ ЛЮДЕЙ НА ИЗОБРАЖЕНИИ В ПОТОКОВОМ РЕЖИМЕ // Международный научно-исследовательский журнал. — 2022.
7. *Juan Du.* Understanding of Object Detection Based on CNN Family and YOLO // Journal of Physics: Conference Series. — 2018.
8. Research on Image Recognition of Tuberculosis Lesions by Minimally Invasive Surgical Robot Based on YOLOv2 / Guo Yu [и др.] // ARTIFICIAL INTELLIGENCE AND ROBOTICS RESEARCH. — 2024.
9. *Joseph Redmon, Ali Farhadi.* YOLO9000: Better, Faster, Stronger. — 2017.
10. *A. A. Сирота, Е. Ю. Митрофанова, А. И. Милованова.* АНАЛИЗ АЛГОРИТМОВ ПОИСКА ОБЪЕКТОВ НА ИЗОБРАЖЕНИЯХ С ИСПОЛЬЗОВАНИЕМ РАЗЛИЧНЫХ МОДИФИКАЦИЙ

СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ // ВЕСТНИК ВГУ, СЕРИЯ:
СИСТЕМНЫЙ АНАЛИЗ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ. —
2019.