НЕЙРОННЫЕ СЕТИ ДЛЯ РАСПОЗНАВАНИЯ ОБЪЕКТОВ ПРИ ВИДЕОНАБЛЮДЕНИИ В РЕЖИМЕ РЕАЛЬНОГО ВРЕМЕНИ

Петровский В.В.
Томский политехнический университет, ИШИТР, аспирант гр. А3-39, e-mail: vvp32@tpu.ru

Аннотация

Нейронные сети играют значительную роль в различных аспектах видеонаблюдения, применяясь для различных задач. Одной из таких задач являются детектирование объектов. В данной статье будут рассмотрены нейронных сетей, которые используются для распознавания объектов при видеонаблюдении: R-CNN (Region-Based Convolutional Neural Network), YOLO (You Only Look Once) и SSD (Single Shot Detector).

Ключевые слова: нейронные сети, распознавание объектов, Faster R-CNN, YOLOv8, SSD.

Введение

Нейронные сети играют значительную роль в различных аспектах видеонаблюдения, применяясь для различных задач. Одними из таких задач являются: детектирование объектов, распознавание лиц, анализа поведения и отслеживания движущихся объектов. Такие нейронные сети будут актуальны в сфере обеспечения контроля и безопасности, как в промышленной, так и в социальной деятельности.

В данной статье будут рассмотрены нейронных сетей, которые обладают высокой эффективность в распознавании объектов на видеоряде: R-CNN (Region-Based Convolutional Neural Network), YOLO (You Only Look Once) и SSD (Single Shot Detector). Проведен анализ популярных и эффективных моделей: Faster R-CNN, YOLOv8 и SSD.

Исследование моделей сверточных нейронных сетей класса R-CNN

Сверточная нейронная сеть R-CNN, архитектура которой представлена на рис. 1 [1], сужает поиск возможных положений объекта при помощи алгоритма Region Proposal (Selective Search).

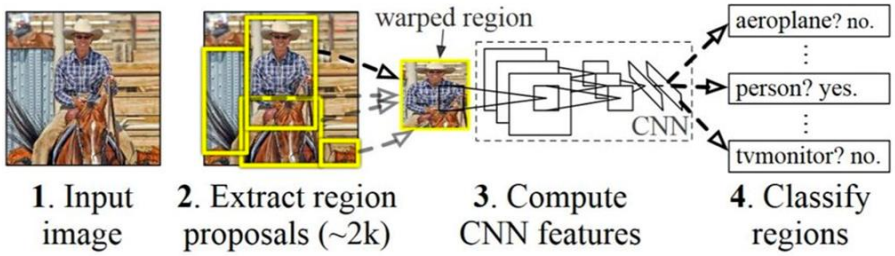

Рис. 1. Архитектура R-CNN

Данный алгоритм получает на вход изображение, а на выходе выдает массив прямоугольников, в которых возможно находится объект. Чтобы распознать конкретный регион запускается классификатор, который разбивает исходное изображение на множество регионов, каждый из которых обрабатывался сверточной нейронной сетью. Такой метод распознавания характеризуется большими временными затратами, и поэтому была разработана улучшенная версия Fast R-CNN, которая представлена на рис. 2 [1].

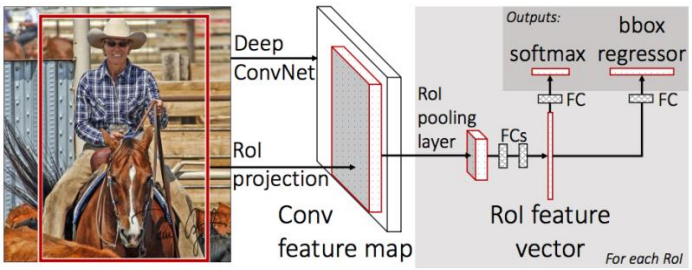

Рис. 2. Архитектура Fast R-CNN

Модификация заключалась в пропускании всего изображения через CNN, где регионы накладываются на полученную общую карту признаков.

Архитектура Faster R-CNN [2], показанная на рис. 3, являлась доработкой предыдущей модели, отличие от которой было в вычислении регионов не по изначальному изображению, а по карте признаков, полученных из CNN. Для этого был добавлен модуль под названием Region Proposal Network (RPN).

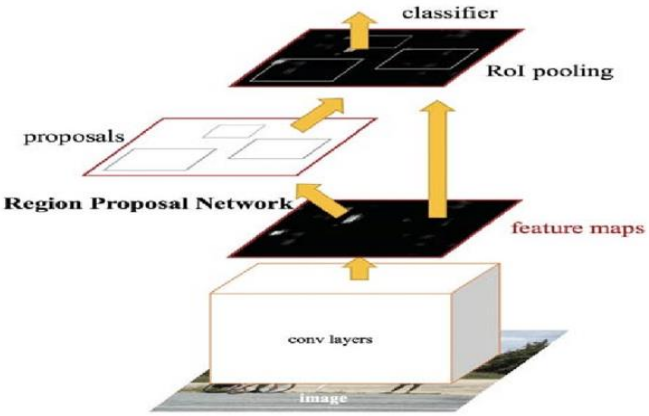

Рис. 3. Архитектура Faster R-CNN

В рамках RPN по извлечённым CNN признакам скользят «мини-нейросетью» с небольшим (3×3) окном. Полученные с её помощью значения передаются в два параллельных полносвязных слоя: box-regression layer (reg) и box-classification layer (cls), как это показано на рис. 4.

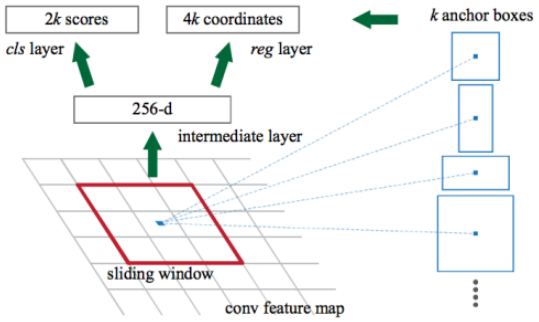

Рис. 4. Схема RPN в Faster R-CNN

Выходы этих слоёв базируются на так называемых anchor-ах: k рамках для каждого положения скользящего окна, имеющих разные размеры и соотношения сторон.

Reg-слой для каждого такого anchor-а выдаёт по 4 координаты, корректирующие положение охватывающей рамки; cls-слой выдаёт по два числа – вероятности того, что рамка содержит хоть какой-то объект или что не содержит. Процесс обучения reg и cls слоёв объединённый; loss-функцию они имеют общую, представляющую собой сумму loss-функций каждого из них, с балансирующим коэффициентом.

В этом эксперименте автор [3] сравнивает производительность и точность архитектур R-CNN, Fast R-CNN и Faster R-CNN. Результаты сравнения показаны на рис. 5.

	R-CNN	Fast R-CNN	Faster R-CNN
Test time per image	50 seconds	2 seconds	0.2 seconds
Speed-up	1x	25x	250x
mAP (VOC 2007)	66.0%	66.9%	66.9%

Рис. 5. Производительность и точность архитектур R-CNN, Fast R-CNN и Faster R-CNN

Для обучения моделей автором использовались 2000 размеченных изображений из датасета PASCAL VOC 2007 разного разрешения.

Из результатов видно, что версия Faster R-CNN имеет высокие показатели по скорости и точности распознавания объекта на изображении по сравнению с предыдущими версиями.

Исследование моделей сверточных нейронных сетей класса YOLO

Архитектура СНС YOLOv8 представлена на рис. 6 [4].

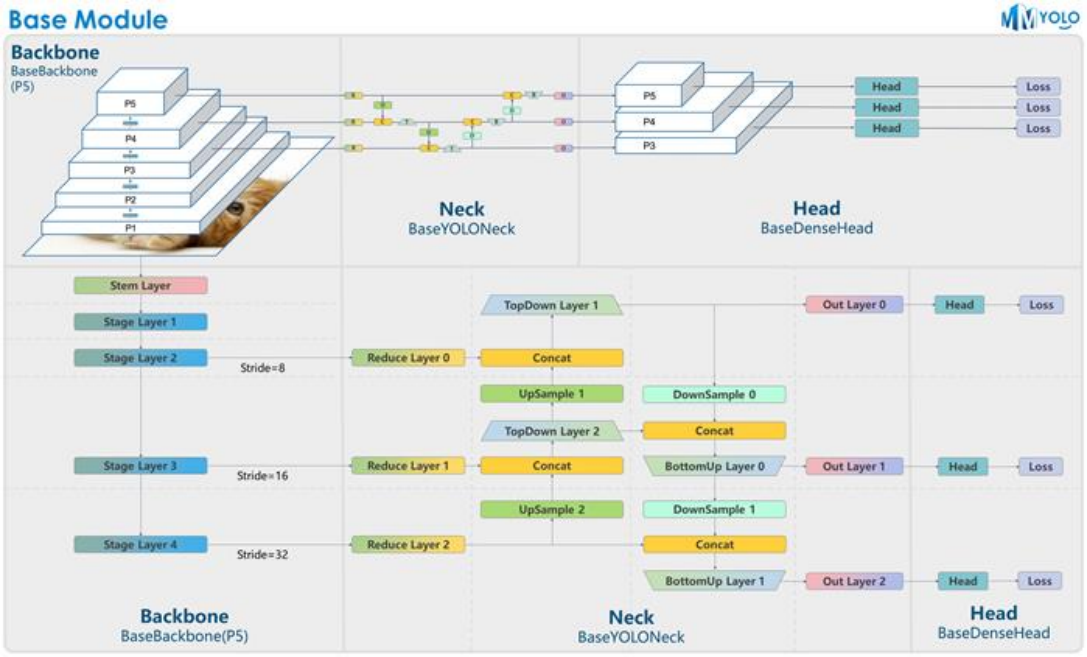

Рис. 6. Архитектура YOLOv8

Принцип работы нейронной сети YOLO заключается в следующем [5]:

1. Изображение делится на квадратную сетку.
2. Для каждой ячейки сети CNN выводит вероятности определяемого класса.
3. Ячейки, имеющие вероятность класса выше порогового значения, выбираются и используются для определения местоположения объекта на изображении.

Нейронные сети с архитектурой YOLO по большей части стали популярны благодаря детектированию объектов при видеонаблюдении в режиме реального времени.

Авторы [6] сравнивают зависимости количества параметров (parameters) и производительности (NVIDIA Latency A100 TensorRT FP16) от точности (COCO mAP 50÷95 %) распознавания объектов на следующих версиях этой архитектуры: YOLO v5, v6, v7 и v8 в соответствии с их сборкой: Nano (n), Small (s), Medium (m), Large (l) и Extra Large (x). Результаты представлены на рис. 7.

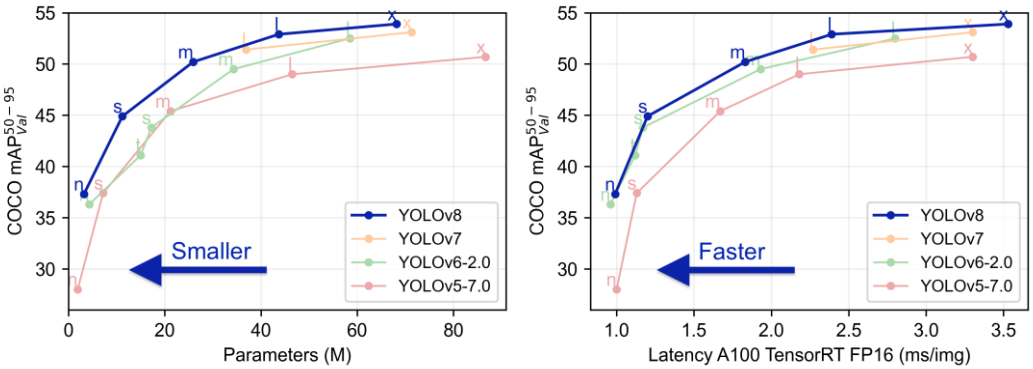

Рис. 7. Влияние параметров и производительности на точность у разных версий YOLO

При обучении архитектур использовался датасет COCO из 2000 размеченных изображений в разрешении 640x640 пиксель.

Все варианты YOLO-v8 обеспечивают лучшую пропускную способность при одинаковом количестве параметров, что указывает на аппаратно-эффективные архитектурные реформы.

Архитектура YOLOv8 является улучшенной версии YOLOv5. В ее основе лежит новый функционал возможностей, показывающий большую точность, гибкость и производительность алгоритма.

Авторы этой статьи [7] сравнивают производительности версий архитектур Small/Large YOLOv5 и YOLOv8, что показано на рис. 8.

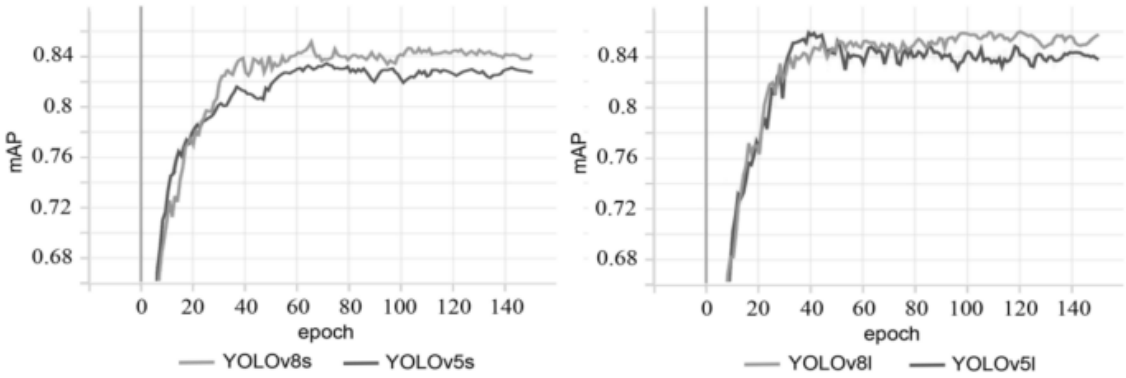

Рис. 8. Изменение средней точности детекции объектов в зависимости от эпохи обучения

Обучение модели происходит на 2231 размеченных изображениях с разрешением 1920x1080 пиксель на платформе Google Colab, используя среду выполнения NVIDIA Tesla T4 с тензорными ядрами.

Обучение архитектур YOLOv8s и YOLOv5s составило 132 минуты и 125 минут соответственно, а обучение архитектур YOLOv8I и YOLOv5I – 247 минут и 195 минут соответственно. При этом улучшение точности между Small и Large версиями эффективнее у архитектуры YOLOv8.

Single Shot Detector

SSD — популярный и эффективный метод обнаружения объектов. Одной из ключевых особенностей твердотельных накопителей является то, что они используют CNN для прогнозирования ограничивающих рамок и меток классов для объектов на изображении. Архитектура данной модели представлена на рис. 9 [8].

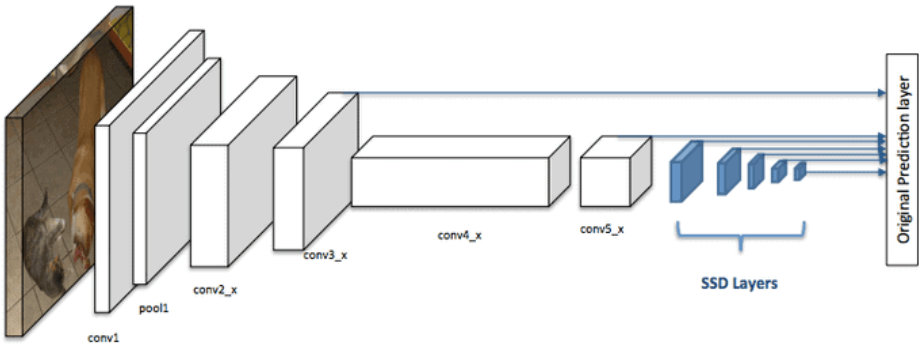

Рис. 9. Применение SSD к сверточной нейронной сети

После основной СНС следуют дополнительные слои, которые добавляются поверх основной сети. Эти дополнительные слои отвечают за обнаружение объектов в разных масштабах и обычно состоят из сверточных слоев и слоев объединения.

В данной статье авторы [9] сравнивают производительность тестовых данных в Faster R-CNN и детекторе одиночных выстрелов (SSD) на метриках: точность (precision), полнота (Recall), показатель

f1 (f1 score) и средний IoU (mean IoU) выходного двоичного изображения модели. Результаты представлены на рис. 10. При обучении было использовано 450 размеченных изображений в разрешении 1152x1152 пиксель из датасета, сделанным вручную из изображений полей, снятых на камеру дрона.

Model	Precision	Recall	F1 Score	Mean IoU	Inference Time of 1152 × 1152 Image in Seconds
Faster RCNN	0.65	0.68	0.66	0.85	0.23
SSD	0.66	0.68	0.67	0.84	0.21

Рис. 10. Сравнение производительности на тестовых данных в Faster R-CNN и SSD

Как видно из опыта, точность, полнота, показатель f1 и среднее значение IoU обеих моделей были одинаковыми, но модель SSD была немного быстрее в исполнении, чем Faster R-CNN.

В статье [10] автор сравнивает производительность YOLOv8 и SSD на NVIDIA GTX 980M.

Сравнение осуществлялось при видеонаблюдении офиса в режиме реального времени, а результаты представлены на рис. 11.

Model	Frames Per Second (FPS)	Accuracy
YOLOv8	40-155	Sacrifices some accuracy
SSD	22-46	Good accuracy

Рис. 11. Сравнения производительности YOLOv8 и SSD

При сравнении производительности YOLOv8 и SSD обе модели имеют свои сильные и слабые стороны. YOLOv8 отдает приоритет скорости, что делает его подходящим для приложений реального времени, но при этом может пожертвовать некоторой точностью. С другой стороны, SSD обеспечивает баланс между скоростью и точностью и может эффективно обрабатывать объекты разных масштабов.

В этой же статье автор сравнивает количество кадров в секунду при распознавании объектов в режиме видеонаблюдении у YOLOv8, SSD и Faster R-CNN, показанных на рис. 12.

Model	Frames Per Second (FPS)
YOLOv8	40-155
SSD	22-46
Faster R-CNN	5-7

Рис. 12. Сравнения производительности YOLOv8, SSD и Faster R-CNN

По рассмотренной информации в таблице 1 приведена эффективность трех моделей распознавания объектов при видеонаблюдении в режиме реального времени.

Таблица 1

Эффективность архитектур

Модель	Влияние изменения размера и разрешения	Скорость детектирования	Точность распознавания
Faster R-CNN	слабое	медленная	сильная
YOLOv8	среднее	быстрая	средняя
SSD	сильное	средняя	сильная

Заключение

На основании проделанной работы по исследованию нейронных сетей для распознавания объектов при видеонаблюдении можно сделать следующие выводы:

1. Точность и быстрота распознавания объектов напрямую зависит от качества и количества изображений датасета при обучении модели;
2. Размер распознаваемого объекта и разрешение изображения влияют на производительность: изображения одинакового разрешения позволяют эффективнее распознать находящийся там объект.

Точность Faster R-CNN меньше всего подвержена влиянию при изменении разрешения изображения и размера распознаваемого объекта в процессе видеонаблюдения. Модели YOLO быстры, но SSD точнее в распознавании. Исходя из различных целей и условий по распознаванию объектов, каждая из моделей может показать хорошие результаты.

Список использованных источников

1. Deep learning concepts and datasets for image recognition // Researchgate.net: сайт. – 2019. – URL: https://www.researchgate.net/figure/Architecture-of-the-Region-based-Convolutional-Neural-Network_fig4_335183353 (дата обращения: 03.04.2024).
2. Faster R-CNN Based Deep Learning for Seagrass Detection from Underwater Digital Images // Researchgate.net: сайт. – 2017. – URL: https://www.researchgate.net/figure/An-overview-of-object-detection-with-Faster-R-CNN-It-is-a-single-deep-network-having-two_fig1_338365099 (дата обращения: 03.04.2024).
3. Introduction to How Faster R-CNN, Fast R-CNN and R-CNN Works // YouTube.com: сайт. – 2017. – URL: https://www.youtube.com/watch?app=desktop&v=v5bFVbQvFRk (дата обращения: 03.04.2024).
4. Raza M. Yolo V8: A Deep Dive Into Its Advanced Functions and New Features. – Текст: электронный // Medium.com: сайт. – 2023. – 16 октября. – URL: https://medium.com/@mujtabaraza194/yolo-v8-a-deep-dive-into-its-advanced-functions-and-new-features-f008599fe604 (дата обращения: 02.04.2024).
5. Model design instructions. YOLO series model basic class // Github.com: сайт. – URL: https://github.com/open-mmlab/mmyolo/blob/main/docs/en/recommended_topics/model_design.md (дата обращения: 02.04.2024).
6. Ultralytics YOLOv8 // Github.com: сайт. – URL: https://github.com/ultralytics/ultralytics (дата обращения 02.04.2024).
7. Филичкин С.А., Вологдин С.В. Сравнение эффективности алгоритмов YOLOv5 и YOLOv8 для обнаружения средств индивидуальной защиты человека / Интеллектуальные системы в производстве: статья // Ижевск: ИжГТУ имени М.Т. Калашникова. – 2023. – Том 21. – № 3. – С. 124–131.
8. Baeldung. Single Shot Detectors (SSDs). The SSDs Architecture. – Текст: электронный // Baeldung.com: сайт. – 2023. – 19 июня. – URL: https://www.baeldung.com/cs/ssd (дата обращения: 03.04.2024).
9. Narenthiran Veeranampalayam Sivakumar A., Li J, Scott S., Psota E., J. Jhala A. Comparison of Object Detection and Patch-Based Classification Deep Learning Models on Midto Late-Season Weed Detection in UAV Imagery / Biological Systems Engineering: журнал // Lincoln: DigitalCommons@University of Nebraska. – 2020. – 3 июля. – С. 13.
10. YOLOv8 vs SSD: Choosing the Right Object Detection Model. Understanding Object Detection Models: Faster R-CNN, YOLO, and SSD. – Текст: электронный // Keylabs.ai: сайт. – 2023. – 23 декабря. – URL: https://keylabs.ai/blog/yolov8-vs-ssd-choosing-the-right-object-detection-model (дата обращения: 03.04.2024).