

**Бабунов Фёдор Андреевич**, магистрант, Тверского Государственного  
Технического Университета, г. Тверь  
**Колачиков Александр Сергеевич**, магистр, Университета информационных  
технологий, механики и оптики, г. Санкт-Петербург

## **СРАВНЕНИЕ ЭФФЕКТИВНОСТИ ОБНАРУЖЕНИЯ ОБЪЕКТОВ СОВРЕМЕННЫМИ СВЕРТОЧНЫМИ НЕЙРОННЫМИ СЕТЯМИ**

**Аннотация.** В данной статье рассматриваются две сверточные нейронные сети, разработанных в 2016 году: YOLO (You Only Look Once) и SSD (Single Shot MultiBox Detector), применяемые для задач обнаружения объектов в режиме реального времени. Описываются основные характеристики каждой из архитектур, их преимущества и недостатки. Обе сети оцениваются по таким критериям, как скорость обработки изображений, точность обнаружения объектов, сложность реализации и эффективность при работе с различными объектами. Для оценки эффективности обеих моделей использовался стандартизированный набор данных (COCO) и единый протокол обучения, включающий одинаковое число эпох и условия тестирования.

**Ключевые слова:** сверточные нейронные сети, YOLO, SSD, датасет

**Annotation.** This article discusses two convolutional neural networks developed in 2016: YOLO (You Only Look Once) and SD (Single Shot MultiBox Detector), used for real-time object detection tasks. The main characteristics of each architecture, their advantages and disadvantages are described. Both networks are evaluated based on criteria such as image processing speed, object detection accuracy, implementation complexity, and efficiency when working with different objects. To evaluate the effectiveness of both models, a standardized dataset (COCO) and a single training protocol were used, including the same number of epochs and test conditions.

**Keywords:** convolutional neural networks, YOLO, SSD, dataset

В нашем мире присутствует огромное количество данных которые наш мозг обрабатывает, однако не все из них он способен проанализировать,

именно по этому в 2016 году учёные задумались о развитии сверточных нейронных сетей для задач компьютерного зрения. Чтобы изменить этот мир или принести в него больше определённости благодаря дополнительному анализу информации. В том же 2016 году были разработаны две популярные и технологичные на тот момент нейронные сети, о которых мы сегодня поговорим.

Первой из них является нейронная YOLO. YOLO (You Only Look Once) — это алгоритм компьютерного зрения, разработанный для детекции объектов в режиме реального времени. Его основная особенность заключается в том, что YOLO способен обнаруживать объекты всего за один проход через изображение (отсюда и название "You Only Look Once").

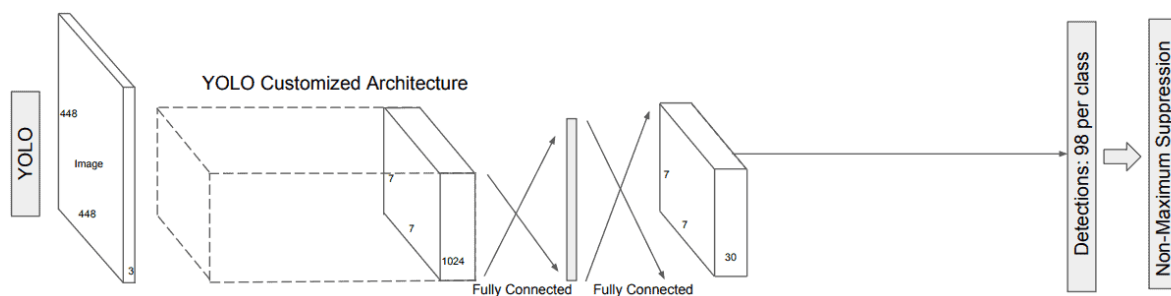
Преимущества нейронной сети YOLO:

- Высокая скорость: благодаря тому, что YOLO обрабатывает изображение за один проход, она способна работать в режиме реального времени.
- Высокая точность для небольших объектов: при помощи деления изображения на небольшие сегменты YOLO обладает способностью хорошо обрабатывать маленькие объекты на изображении.

Недостатки нейронной сети YOLO:

- При наличии сложных сцен, или кадров, в которых находятся объекты, перекрывающие друг друга, YOLO обладает меньшей точностью чем модели семейства Faster R-CNN.
- Проблемы при обнаружении объектов схожих по размерам и формам, происходят из-за фиксированного размера сетки.

Архитектура нейронной сети YOLO



Второй сверточной нейронной сетью была SSD (Single Shot MultiBox Detector) — это нейронная сеть, предназначенная для обнаружения объектов в режиме реального времени. Она относится к классу одноэтапных (single-stage) архитектур, это означает, что процесс обнаружения объектов выполняется за один проход через сеть, в отличие от многоступенчатых подходов, таких как Faster R-CNN.

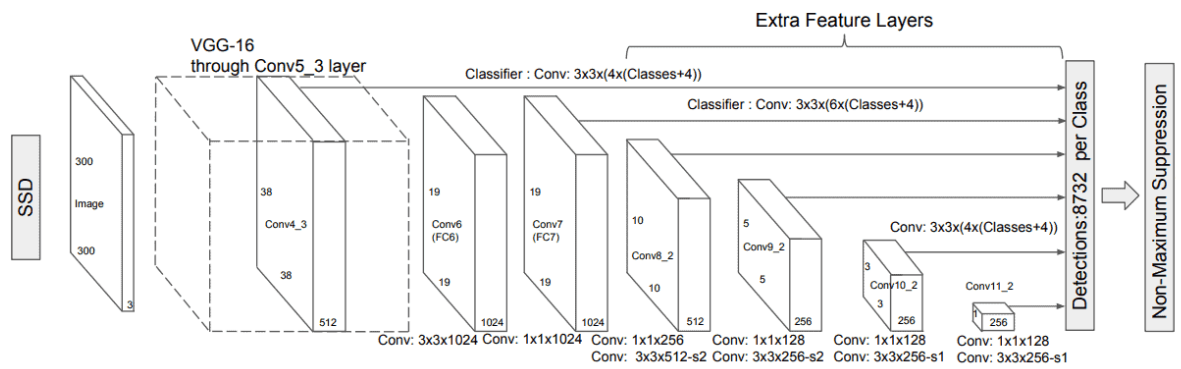
#### Преимущества нейронной сети SSD

- **Скорость:** Одноэтапная архитектура обеспечивает высокую производительность по сравнению с многоступенчатыми.
- **Точность:** Многоуровневый подход позволяет более эффективно обнаруживать объекты разных масштабов.

#### Недостатки нейронной сети SSD

- **Сложность обучения:** относительно других моделей скорость обучения посредственная, а время, затрачиваемое на подготовку к обучению велико.
- **Трудности с маленькими объектами:** у SSD возникают трудности с обнаружением маленьких объектов, так как они занимают небольшую область на высоком разрешении фиче-мапов.
- **Меньшая точность по сравнению с двухэтапными моделями:** по сравнению с двухэтапными моделями скорость SSD выше, однако точность заметно ниже чем у двухэтапных моделей.

#### Архитектура нейронной сети SSD



Так как эти нейросети были самыми популярными на тот момент, под наше сравнение попали именно они. Для обучения нейронных сетей использовался набор данных COCO (Common Objects in Context). Данный набор данных предназначен для распознавания объектов из ряда классов визуальных объектов в реалистичных сценах. Этот набор данных используется для обучения с учителем, так как предоставляет обучающий набор помеченных изображений. Всего набор содержит восемьдесят классов объектов.

Набор данных COCO предоставляет данные для обучения, валидации и тестирования. Набор разделен на 72% для обучения, 25% для тестирования и 3% для валидации. Распределение изображений и объектов по классам примерно одинаково по обучающим тестовым и валидационным наборам.

Данные для обучения состоят из набора изображений, каждое из которых имеет файл аннотации, содержащий координаты ограничивающей рамки и метку класса объекта. Всего обучающий набор содержит 163957 изображений для обучения используется 118287 изображений, для тестирования 40670 изображений, для валидации 5000 изображений. Помимо этого для обучения и использования SSD потребовалось написать на 80% больше строк кода чем для моделей YOLO.

Обучив обе нейронные сети на заранее подготовленном датасете, необходимо было провести тестирование на «незнакомом» наборе данных, чтобы выяснить, насколько сети научились качественно определять объекты на новых изображениях. Все нейронные сети прошли обучение на 150 эпохах

для создания одинаковых условий, так же для валидации и проверки нейронных сетей использовался одинаковый набор изображений и видео.

Модель SSD показала очень слабые результаты на проверку mAP, которая составляет 18,7%, а также низкую скорость обработки кадров, составляющую всего 15 кадров в секунду.

При этом YOLOv11 показала результаты в 53,8% mAP и способность обрабатывать 130 кадров в секунду.

Мы бы могли на этом и закончить с результатами того что YOLO полностью разгромила модель SSD, но это было бы слишком просто, к тому же модель SSD была заброшена, а модель YOLO активно поддерживается и развивается. По этому было принято решение провести сравнение самой популярной версии YOLO с её только вышедшей свежей версией, хотя это не правда ведь разные команды учёных ведут независимые друг от друга разработки используя концепции изначальных версий YOLO.

Обучив YOLOv12 на том же самом датасете она показала результат в 56.1% mAP и способность обрабатывать 100 кадров в секунду.

Если сравнить модели YOLOv11 и YOLOv12, то по способности, в скоростной обработки изображений, YOLOv11 опережает свою старшую версию на 23%, но при этом отстаёт на 4% по mAP и не обладает модулем внимания, благодаря которому ограничивающие рамки дольше находятся на обнаруженных объектах, которые подвержены меньшему количеству ложных срабатываний. Исходя из этого модель YOLOv12 будет более предпочтительно для задач распознавания объектов в реальном времени, но про YOLOv11 тоже не стоит забывать, ведь из-за её скорости она будет более предпочтительна в задачах где важна своевременность получения информации.

## **Список литературы**

1. Модели, поддерживаемые Ultralytics // Ultralytics YOLO Docs: [сайт]/ - URL: <https://docs.ultralytics.com/ru/models>
2. Shaoqing, R. Faster R-CNN: Towards Real-Time Object Detection with Regional Proposal Networks / R. Shaoqing, H. Kaiming, R. Girshick, J. Sun // IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 39, N 6, 2016. – P. 1137-1149.
3. Girshick, R. Fast R-CNN // 2015 IEEE International Conference on Computer Vision (ICCV). – IEEE, 2015. – P. 1440-1448.
4. Liu, W. SSD: Single Shot MultiBox Detector / W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu, A.C. Berg // Computer Vision – ECCV 2016. – Springer, 2016. – P. 21-37.
5. Redmon, J. You Only Look Once: Unified, Real-Time Object Detection / J. Redmon, S. Divvala, R. Girshick, A. Farhadi // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – IEEE, 2016. – P. 779-788
6. Everingham, M. The PASCAL Visual Object Classes (VOC) Challenge / M. Everingham, L. Van Gool, C. Williams, J. Winn, A. Zisserman // International Journal of Computer Vision, Vol. 88, N 2, 2010. – P. 303-338.