

# Использование нейронной сети YOLOv3 для распознавания нарушений техники безопасности

А. И. Султанова, email: sultanovaali@stud.kai.ru <sup>1</sup>

В. В. Мокшин, email: vladimir.mokshin@mail.ru <sup>1</sup>

<sup>1</sup> Институт компьютерных технологий и защиты информации. Казанский национальный исследовательский технический университет им. А.Н. Туполева - КАИ

***Аннотация.** Рассмотрены задачи поиска и классификации объектов на изображениях с использованием свёрточной нейронной сети YOLOv3. Практическая реализация произведена на языке программирования Python под платформу Windows. Представлено описание этапов обучения нейронной сети и оценки качества работы полученной модели. Результаты тестирования нейросетевой модели представлены в виде диаграммы, содержащей среднюю точность по каждому классу и метрику mAP. Полученные результаты могут быть использованы в системах мониторинга безопасности на производственных объектах.*

***Ключевые слова:** распознавание объектов, сверточная нейронная сеть, YOLOv3, Python.*

## Введение

Любая производственная деятельность сопровождается всевозможными рисками – событиями, которые влияют на функционирование всего предприятия. Особое внимание уделяется рискам, связанным с нанесением ущерба здоровью или жизни человека в зонах производственных работ, поскольку уровень безопасности влияет на экономические показатели и статус предприятия. Своевременное обнаружение нарушений техники безопасности может помочь, в предотвращении несчастных случаев на производстве.

На сегодняшний день большую популярность и распространение получили сверточные нейронные сети (CNN). Они прекрасно себя проявили в области компьютерного зрения, как методы детектирования и распознавания объектов на изображениях или в видеопотоке [1-2]. Одним из современных детекторов выступает сверточная нейронная сеть YOLO (You Only Look Once - Ты Смотришь Только Один Раз). Благодаря своей архитектуре, ее достаточно один раз «взглянуть» на

изображение, что значительно сокращает время необходимое для обнаружения объекта в кадре.

Используя технологию сверточных нейронных сетей для обнаружения и распознавания людей, применяющих или не применяющих сигнальные жилеты и защитные каски при производственных работах, можно повысить уровень обеспечения безопасности на производственных объектах.

## 1. Цель исследования

Цель данной научно-технической работы является исследование возможностей применения сверточной нейронной сети YOLOv3 в рамках анализа производственных рисков.

Для достижения цели необходимо выполнить ряд задач:

1. Подготовить обучающую и тестовую выборку изображений, содержащих классы распознаваемых объектов.
2. Обучить модель сверточной нейронной сети YOLOv3.
3. Произвести тестирование обученной модели для определения точности распознавания объектов.
4. Оценить качество полученных результатов.

## 2. Нейронная сеть YOLOv3

Главная особенность нейронной сети YOLO по сравнению с другими состоит в том, что большинство систем применяют CNN несколько раз к разным регионам изображения, а YOLO применяет CNN один раз ко всему изображению сразу. Благодаря этому сеть обладает высокой скоростью и точностью обнаружения объектов [3]. На рис. 1 представлена архитектура YOLO [4].

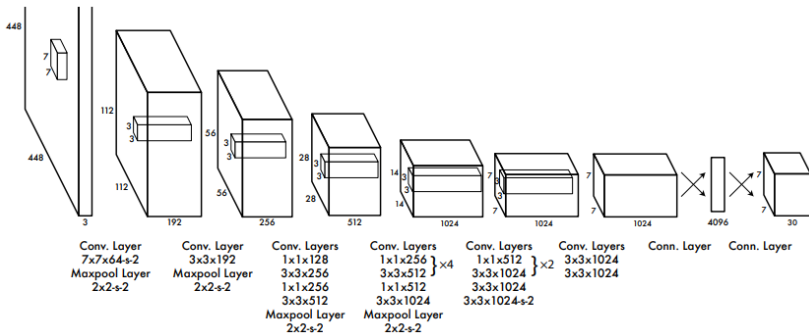


Рис. 1. Архитектура YOLO [4]

Поскольку YOLO смотрит на изображение только один раз, то метод с плавающим окном не подходит для данной сети. Вместо этого, все изображение с помощью сетки разбивается на ячейки размером  $S \times S$ . После этого для каждой ячейки предсказываются два показателя: ограничительные рамки (bounding boxes) и вероятности нахождения в них нужного объекта (confidence).

Размерность и положение ограничительных рамок вокруг объекта на изображении (рис. 2) вычисляются по формулам:

$$b_x = \sigma(t_x) + c_x \quad (1)$$

$$b_y = \sigma(t_y) + c_y \quad (2)$$

$$b_w = p_w e^{t_w} \quad (3)$$

$$b_h = p_h e^{t_h} \quad (4)$$

где  $t_x, t_y, t_w, t_h$  – предсказанные координаты центра рамки  $x, y$ , ширина и высота рамки соответственно,  $p_w, p_h$  – ширина и высота рамки,  $c_x, c_y$  – смещение ячейки от верхнего левого угла изображения [5].

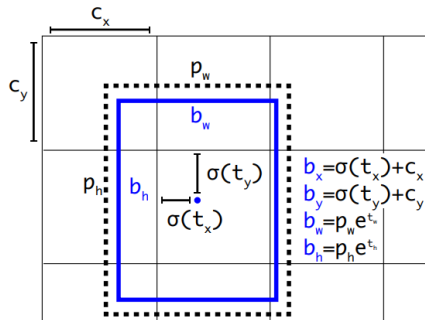


Рис. 2. Схема расчета положения ограничительных рамок [5]

Вероятность попадания искомого объекта внутрь предсказанной рамки рассчитывается по формуле:

$$\sigma(t_o) = \Pr(\text{object}) * IOU(b, \text{object}) \quad (5)$$

где  $\sigma(t_o)$  – «оценка доверия» (вероятность попадания искомого объекта внутрь предсказанной рамки) [5].

После того как были получены все ограничительные рамки и соответствующие вероятности для всех искомым объектов необходимо

избавиться от дубликатов bounding box для каждого объекта. Для этого применяется алгоритм «подавления не-максимумов» (Non-maximum suppression) [6], который ищет ограничительные рамки содержащие объект распознавания и вычисляет для них  $IOU$  (Intersection over Union) по формуле:

$$IOU = \frac{A \cap B}{A \cup B} \quad (6)$$

Модель YOLOv3 сверточной нейронной сети состоит из 106-ти слоев (рис. 3) и лучше детектирует небольшие объекты по сравнению с её предыдущей версией YOLOv2 [7].

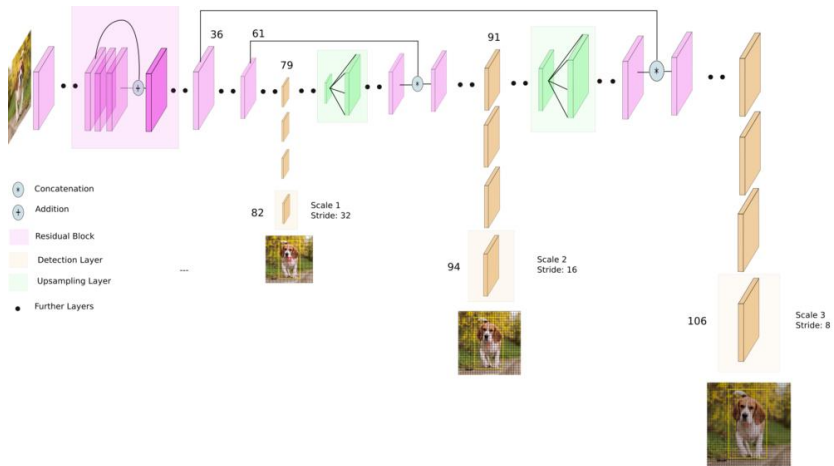


Рис. 3. Архитектура нейронной сети YOLOv3 [8]

Основная особенность YOLOv3 состоит в том, что на выходе есть три слоя, каждый из которых рассчитан на обнаружения объектов разного размера. Для этого используются различные сетки, размеры ячеек которых имеют значения 8, 16 и 32.

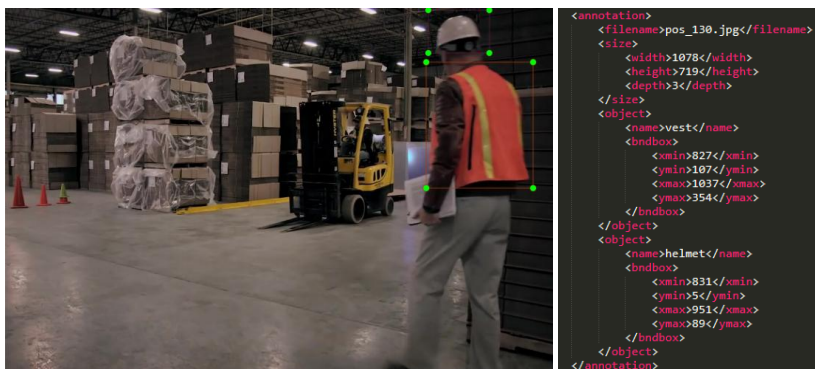
### 3. Обучение модели

Экипировкой рабочего на производстве может быть просто каска, каска и сигнальный жилет, страховочный пояс с жилетом и каской, иногда к этому набору добавляются перчатки. Поскольку во всех случаях участвует каска и сигнальный жилет, выделим два основных класса: каска (helmet) и жилет (vest).

Для обучения нейронной сети был подготовлен собственный набор данных (dataset) в формате Pascal VOC, содержащий изображения и их

аннотации в виде XML-файла. Было подготовлено 300 кадров, которые были разделены на обучающую и тестовую выборку. Для каждого изображения вручную выполнялась разметка объектов и присваивание им меток класса.

На рис. 4 изображен пример разметки элемента обучающей выборки, на нем показано изображение и соответствующая ему аннотация, которая содержит: имя файла изображения, его размер, имя класса и координаты, представляющие локацию рамки вокруг объекта (верхний левый и нижний правый угол).



а – изображение, б – аннотация

Рис. 4. Пример эталонной разметки изображения

Для обучения нейронной сети было использовано «переносное обучение» [6]. Оно предполагает под собой использование в качестве отправной точки веса от предварительно обученной модели на наборе данных COCO.

На вход обучающего алгоритма подается обучающая выборка размером 240 изображений и аннотаций. Затем начинается итерация обучения (эпоха) – изображение проходит через все слои по порядку, а получившийся на выходе результат сравнивается с аннотацией с помощью функции ошибки (рис. 5).



Рис. 5. Общие потери при обучении

С каждой следующей эпохой веса сети в процессе обучения постепенно подгоняются к оптимальному значению, необходимому для точного решения поставленной задачи и общие потери, как видно на таблице, с каждой эпохой уменьшаются.

#### 4. Оценка качества работы

На основе тестовой выборки, содержащей 60 изображений, была проведена оценка качества работы каждой полученной во время обучения модели нейронной сети. Для этого были рассчитаны средние точности распознавания для каждого класса и метрика mAP. Результаты приведены на рис. 6.

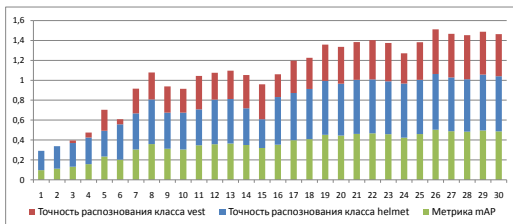


Рис. 6. Точность распознавания объектов и метрика mAP

Исходя из диаграммы, представленной на рисунке 6, был сделан вывод, что модель под номером 6, detection\_model-ex-0045--loss-0017.756.h5 с метрикой mAP = 0,5041 имеет наибольшую точность распознавания объектов на изображении по сравнению с другими сгенерированными во время обучения моделями. Поэтому, в качестве

нейронной сети для дальнейшей работы выбираем данную модель YOLOv3.

Результат работы обученной распознавать сигнальные жилеты и защитные каски, и протестированной нейронной сети представлены на рис. 7.



Рис. 7. Примеры обнаружения объектов на изображениях с помощью сверточной нейронной сети YOLOv3

На некоторых изображениях видны ложные срабатывания и необнаруженные объекты, причиной этому может быть недостаточно точная разметка эталонов или недостаточное количество обучающей выборки.

### Заключение

В соответствии с целью научной работы реализована программа распознавания экипировки на изображении на языке программирования Python для сверточной нейронной сети YOLOv3. Для обучения был сформирован dataset, состоящий из 300 изображений и аннотаций к ним.

Обучение нейронной сети проводилось на 240 изображениях, а для оценки качества было подготовлено 60 изображений. Наилучший результат показала модель `detection_model-ex-0045--loss-0017.756.h5`. Так же во время обучения были выявлены ключевые моменты, значительно влияющие на качество полученных результатов и эффективность работы системы распознавания:

- Выборка изображений для dataset должна быть как можно более полной и качественной;
- Следует максимально точно выполнить аннотацию объектов;
- В начале обучения стоит использовать веса уже обученных моделей.

На базе полученных результатов возможно построение более высокоуровневых моделей для применения в исследовательских целях, а так же для создания систем анализа и мониторинга безопасности на производственных объектах.

### **Список литературы**

1. Распознавание образов транспортных средств на основе эвристических данных и машинного обучения / В. В. Мокшин [и др.] // Вестник Технологического университета. – 2016. – Т. 17, № 5. – С. 130-137.
2. Стадник, Н. А. Разработка информационной системы организации работ производственного цеха / Н. А. Стадник, А. В. Золотухин, В. В. Мокшин // Информационные технологии и нанотехнологии (ИТНТ-2021) : сб. тр. по материалам VII Международной конференции и молодежной школы «Информационные технологии и нанотехнологии» (Самара, 20–24 сентября 2021 г.). – Самара, 2021. – С. 32982.
3. YOLO: Real-Time Object Detection [Электронный ресурс]. – Режим доступа : <https://pjreddie.com/darknet/yolo>
4. You Only Look Once: Unified, Real-Time Object Detection [Электронный ресурс] / J. Redmon [and all] : Computing Research Repository (CoRR), 2016. – Режим доступа : <https://arxiv.org/abs/1506.02640>
5. Redmon J. Yolo9000: Better, faster, stronger. In Computer Vision and Pattern Recognition (CVPR) [Электронный ресурс] / J. Redmon, A. Farhadi : Computing Research Repository (CoRR), 2017. – Режим доступа : <https://arxiv.org/abs/1612.08242>
6. Faster r-cnn: Towards real-time object detection with region proposal networks [Электронный ресурс] / S. Ren [and all] : Computing Research Repository (CoRR), 2015. – Режим доступа : <https://arxiv.org/abs/1506.01497>
7. Redmon J. Yolov3: An incremental improvement [Электронный ресурс] / J. Redmon, A. Farhadi : Computing Research Repository (CoRR), 2018. – Режим доступа : <https://arxiv.org/abs/1804.02767>
8. Kathuria A. What’s new in YOLO v3? [Электронный ресурс] / A. Kathuria : towards data science, 2018. – Режим доступа : <https://towardsdatascience.com/yolo-v3-object-detection-53fb7d3bfe6b>