

## ОСОБЕННОСТИ СЕТИ YOLOV4, ИСПОЛЬЗУЕМОЙ ДЛЯ ОПРЕДЕЛЕНИЯ ПОЛОЖЕНИЯ ИНСТРУМЕНТА

Мисякова В.А., Ковалева И.Л.

Белорусский национальный технический университет  
Минск, Республика Беларусь

В работе [1] для определения положения рабочего инструмента 3D-принтера предлагается использовать сверточную нейронную сеть семейства YOLO. Рассмотрим архитектуру YOLOv4 и основные методы, которые позволяют обеспечивать высокую результативность ее работы.

YOLOv4 [2] представляет собой одноэтапную сеть обнаружения объектов, которая состоит из трех частей: магистрали (backbone), шеи (neck) и головы (head) (рис. 1).

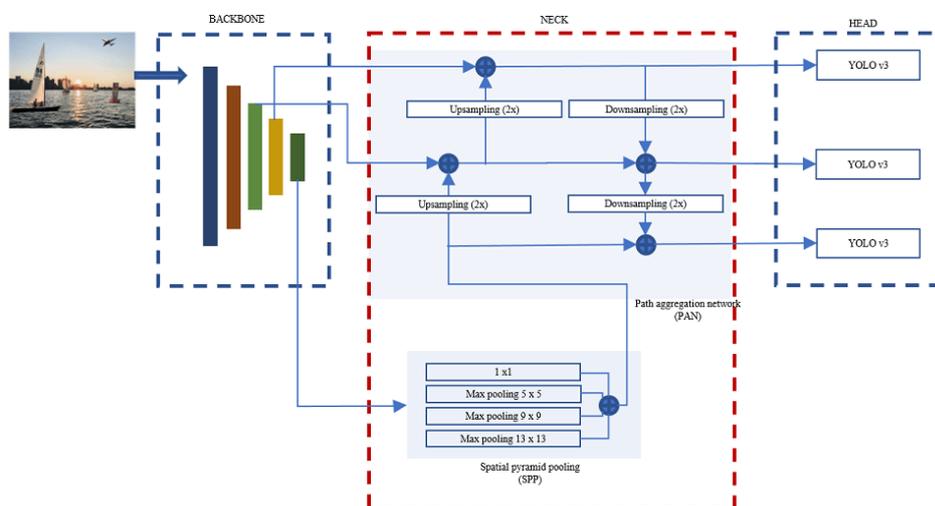


Рисунок 1 – Архитектура YOLOv4

При обучении YOLOv4 используются две группы методов повышения точности детектирования: Bag-of-Freebies (BoF) и Bag-of-Specials (BoS).

В группу BoF входят методы, которые не влияют на вычислительную сложность обученной модели и скорость ее работы. В данной версии разработчики внедрили новые методы мозаичного увеличения данных (Mosaic) и самовредоносного обучения (Self-Adversarial Training, SAT). Mosaic представляет собой новый метод аугментации данных, который объединяет четыре обучающих изображения. Это позволяет обнаруживать объекты вне их обычного контекста. SAT работает в два этапа – прямом и обратном. На первом этапе нейронная сеть выполняет атаку на саму себя, изменяя исходное изображение для создания иллюзии отсутствия на нем желаемого объекта. На втором этапе нейронная сеть обучается обнаруживать объект на этом модифицированном изображении обычным способом. Группа BoS включает методы, которые незначительно

уменьшают скорость работы обученной модели, при этом значительно улучшая точность детектирования.

Сеть YOLOv4 использует предварительно обученную сверточную нейронную сеть CSPDarkNet53 [3] в качестве магистрали для извлечения признаков из входных изображений. Магистральная сеть состоит из пяти модулей остаточных блоков, где каждый блок содержит набор сверточных слоев, а выходные карты признаков из этих остаточных блоков объединяются в шее сети YOLOv4.

Модуль Spatial pyramid pooling (SPP) [4] в шее объединяет выходы подвыборки карт признаков низкого разрешения для извлечения наиболее репрезентативных признаков. Модуль SPP использует ядра размером 1x1, 5x5, 9x9 и 13x13 для операции подвыборки (max-pooling operation). Значение шага установлено равным 1. Объединенные карты признаков из модуля SPP агрегируются с картами признаков высокого разрешения с использованием Path aggregation network (PAN) [5]. Модуль PAN выдаёт набор агрегированных карт признаков, которые используются для предсказаний. Эта агрегация позволяет лучше учесть разнообразие объектов и условий в изображении.

Сеть YOLOv4 выдает карты признаков размерами 19x19, 38x38 и 76x76 для предсказания ограничивающих рамок, оценок классификации и оценок объектности [1].

Рассмотренная архитектура YOLOv4 имеет по сравнению с предыдущими версиями меньшее количество параметров и позволяет работать с относительно небольшим размером мини-пакета, что уменьшает объем вычислений и требования к памяти. YOLOv4 позволяет осуществлять обучение и достигать высокой скорости обнаружения объектов на одном графическом процессоре среднего уровня с объемом графической памяти 8-16 ГБ (например, GTX 1080 Ti или RTX 2080 Ti).

1. Мисякова В.А., Ковалева И.Л. Обоснование использования нейронной сети yolo для определения положения рабочего инструмента 3d-принтера // Новые технологии и материалы, автоматизация производства : материалы международной научно-технической конференции (Брест, 16-17 ноября 2023 г.) / редкол. С.Р. Онысько [и др.] ; Министерство образования РБ, Брестский государственный технический университет. – Брест, 2023. – С. 45-49.
2. Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao. YOLOv4: Optimal Speed and Accuracy of Object Detection / arXiv: journal. – 2020.
3. Chien-Yao Wang, Hong-Yuan Mark Liao, I-Hau Yeh, Yueh-Hua Wu, Ping-Yang Chen, Jun-Wei Hsieh. CSPNet: A New Backbone that can Enhance Learning Capability of CNN / arXiv: journal. – 2019.
4. Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun. Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition / arXiv: journal. – 2015.
5. Shu Liu, Lu Qi, Haifang Qin, Jianping Shi, and Jiaya Jia. Path aggregation network for instance segmentation / arXiv: journal. – 2018.