

DOI 10.36074/logos-13.03.2026.027

MODERN METHODS OF OBJECT DETECTION IN COMPUTER VISION

Vladislav Potapov¹

Scientific supervisor: Liubchenco Valentyn²

1. Student of the Faculty of Information and Analytical Technologies and Management
Kharkiv National University of Radio Electronics, UKRAINE

2. Ph.D., Associate Professor

Kharkiv National University of Radio Electronics, UKRAINE

ORCID ID: 0000-0002-9966-0249

Object detection is a fundamental task in computer vision, which involves determining the presence of objects of specific classes in an image or video frame and localizing their spatial position. The result of the detection algorithm is usually a set of bounding boxes and their corresponding object classes. Object detection methods are the foundation for more complex tasks, such as tracking, action classification, and behavior analysis.

Formally, the object detection task can be presented as finding a set of pairs of the form (1):

$$B = \{(b_i, c_i)\}, i = 1..N \quad (1)$$

where:

b_i - coordinates of the object's bounding rectangle,

c_i - object class. The quality of detection is evaluated by comparing the predicted rectangles with the ground truth labels.

One of the key metrics is the Intersection over Union (IoU) indicator, which is defined as (2):

$$IoU = \frac{S_{\{crossing\}}}{S_{\{association\}}} \quad (2)$$

This metric is used to determine whether the detection is considered correct.

Early object detection methods were based on a sliding window and manually designed features. Such approaches had high computational complexity and limited accuracy.

Modern detection methods are mostly based on deep convolutional neural networks. They are conventionally divided into two large groups: two-stage and one-stage detectors, but within these groups, there are also different approaches to their implementation (Fig. 1-3).

Faster R-CNN is a classic two-stage object detection method in which, in the first stage, a set of candidate regions is formed using a Region Proposal Network, and in the second stage, the coordinates of the bounding boxes are classified and refined [[2]].

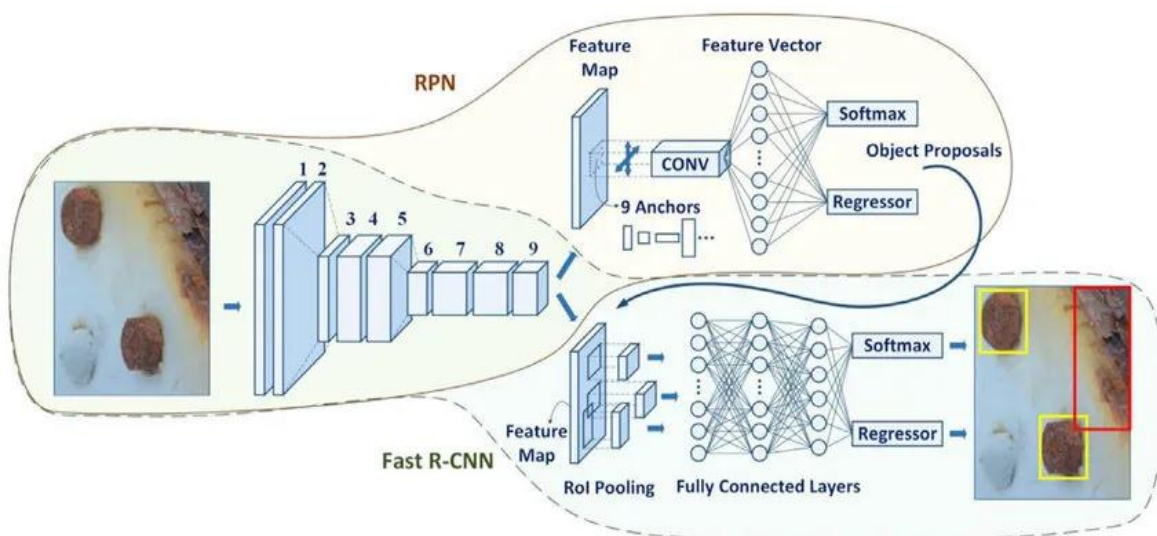


Fig. 1. **Faster R-CNN architecture with RPN and two-stage detection**

The SSD method performs object detection in a single pass through the network using multi-level feature maps, which allows for the effective detection of objects of different scales without a separate region generation stage [[1]].

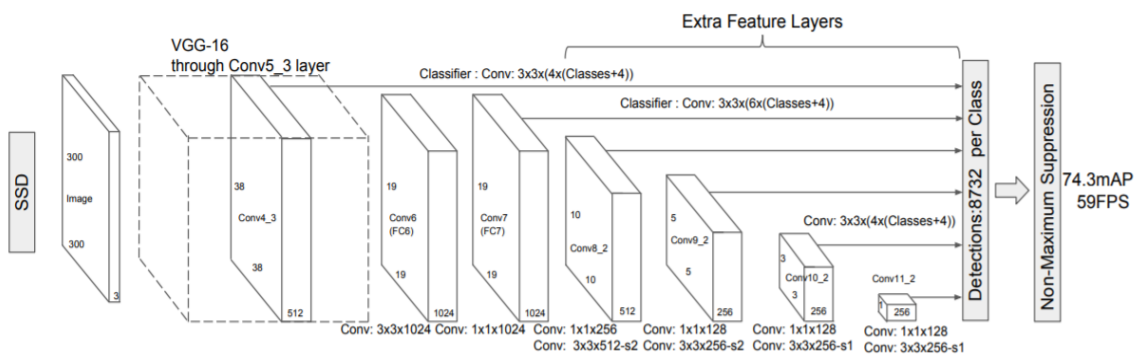


Fig. 2. **SSD architecture with multi-level feature maps and default boxes**

섹션 17.

INFORMATION TECHNOLOGIES AND SYSTEMS

The YOLO (You Only Look Once) method formulates the task of object detection as a direct regression problem. The input image is divided into a regular grid, each cell of which simultaneously predicts the coordinates of bounding rectangles, a confidence score, and the probability of an object belonging to a certain class. This approach allows detection to be performed in a single pass of the neural network, ensuring high performance and suitability for real-time operation.

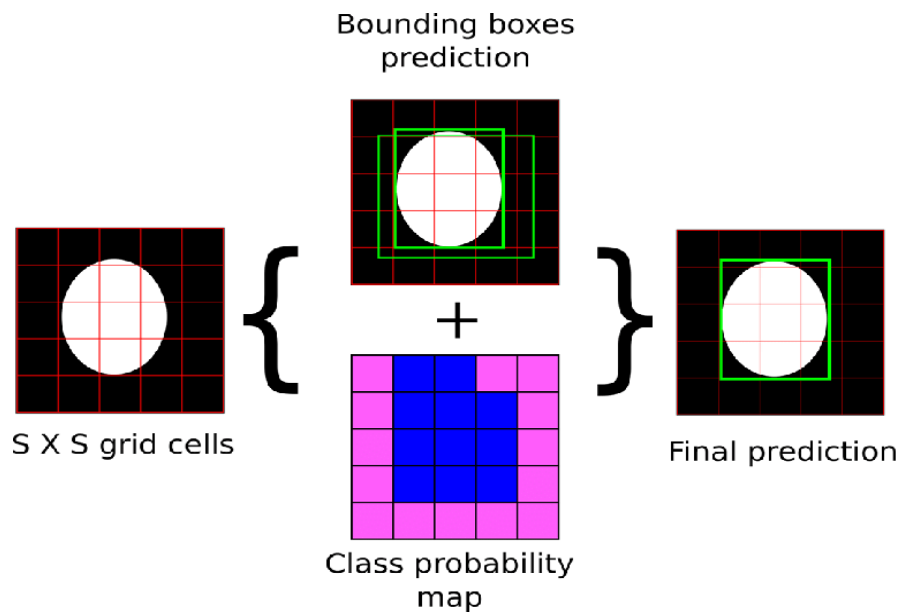


Fig. 3. How YOLO works: grid division and simultaneous prediction of bounding boxes and classes

Table 1

Comparative characteristics of detection approaches

Characteristic	Faster R-CNN	SSD	YOLO
Type of approach	Two-stage	Single-stage	Single-stage
Main idea	Generation of region proposals + classification	Detection at multiple levels of features	Global regression on the grid
Working with object scale	Through region proposals	Explicit multi-scale detection	Through anchors / FPN
Average detection accuracy mAP(50)	Hight	Medium - hight	Medium - hight
Speed	Low - medium	Medium	Hight
Real time	Limited	Limited	Yes
Complexity of implementation	Hight	Medium	Low - medium
Typical areas of application	Research, offline analysis	Embedded systems	Real - time systems

[author's development]

Two-stage approaches provide high localization accuracy, while single-stage models are focused on real-time data processing.

Research overview. Current research focuses on optimizing the balance between accuracy and speed, as well as reducing the number of model parameters without significant loss of quality.

Recent research in object detection aims to further improve the speed and accuracy of single-stage models. Transformer-based approaches such as RT-DETR combine DETR ideas with optimizations for real-time operation. Other modern models, such as RF-DETR and PP-YOLOE, focus on improving feature selection mechanisms, optimizing anchor-free detection, and increasing training stability, which allows them to achieve competitive results in practical applications [[3], [4]].

Conclusions. Object detection methods are a key component of most computer vision systems. The use of deep learning has significantly improved detection quality, but the choice of a specific approach depends on accuracy and speed requirements.

The analysis of modern object detection methods shows that there is no universal approach that can simultaneously ensure high accuracy and speed. Two-stage methods, such as Faster R-CNN, demonstrate high accuracy but are inferior in terms of speed, while single-stage approaches (SSD, YOLO) are focused on real-time applications.

REFERENCES:

- [1] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & Berg, A. C. (2016). *SSD: Single Shot MultiBox Detector*. Proceedings of the European Conference on Computer Vision (ECCV), 21–37. <https://arxiv.org/abs/1512.02325>
- [2] Ren, S., He, K., Girshick, R., & Sun, J. (2015). *Faster R-CNN: Towards real-time object detection with region proposal networks*. Advances in Neural Information Processing Systems (NeurIPS), 28, 91–99. <https://arxiv.org/abs/1506.01497>
- [3] Xu, C., Wang, Y., Zhang, H., & Sun, J. (2022). *PP-YOLOE: An evolved version of YOLO*. arXiv preprint. <https://arxiv.org/abs/2203.16250>
- [4] Zhao, Y., Lv, Y., Xu, S., Wei, S., & Liu, J. (2023). *RT-DETR: Real-Time Detection Transformer*. arXiv preprint. <https://arxiv.org/abs/2304.08069>

