UDC: 004.932’1

Rodiuk A. I., higher education student,
Nikolyuk P. K., Sc. D. (Physical and Mathematical Sciences), professor,
professor at the department of information technologies,
Vasyl’ Stus Donetsk National University

OBJECT DETECTION AND RECOGNITION BASED ON YOLOv11

Abstract. The paper is devoted to the problem of object detection and recognition using the YOLOv11 model. The paper discusses technical improvements to the architecture and analyzes the effectiveness of the model in different conditions and on complex images.

Keywords: YOLOv11, object detection, object recognition.

Introduction. Modern computer vision systems play a key role in the development of artificial intelligence, providing high-precision detection and recognition of objects in real time. The relevance of the research is determined by the need for effective algorithms capable of working in complex dynamic conditions, with a large number of objects and variable lighting.

The aim of this work is to investigate the capabilities of the YOLOv11 model for improving the accuracy and speed of object recognition. Particular attention is paid to analyzing the architectural improvements of the model and evaluating its effectiveness in different environments, which makes it possible to determine the prospects for applying YOLOv11 in practical computer vision tasks.

Main text. The latest development in the YOLO series of object identification models is YOLOv11, created by the Ultralytics team. The model demonstrates exceptional performance in dynamic scenarios, challenging environments, and the detection of small objects [1]. Thanks to the implementation of complex technical modules such as channel-to-pixel space attention (C2PSA), spatial pyramid pooling fusion (SPPF), and C3k2 blocks, YOLOv11 significantly improves the accuracy of multi-scale object detection and the ability to extract their characteristics. These developments further improve the model’s ability to adapt to diverse environments and objects of different sizes, achieving the best balance between computational efficiency and accuracy. Thanks to its sophisticated multitasking capabilities, YOLOv11 extends its capabilities beyond basic detection skills, allowing it to effortlessly perform additional tasks such as object segmentation and pose estimation in addition to standard object detection. These achievements make YOLOv11 a flexible solution for a variety of applications, such as autonomous driving, smart surveillance, and unmanned aerial vehicle vision systems. With its exceptional performance in these areas, YOLOv11 sets a new standard for intelligent and efficient object detection.

The confidence level is an important parameter in object detection models, reflecting the reliability of the predictions obtained. Theoretically, this indicator is determined by the probability of the object’s presence and the exact match with the actual boundaries, described by formula (1):

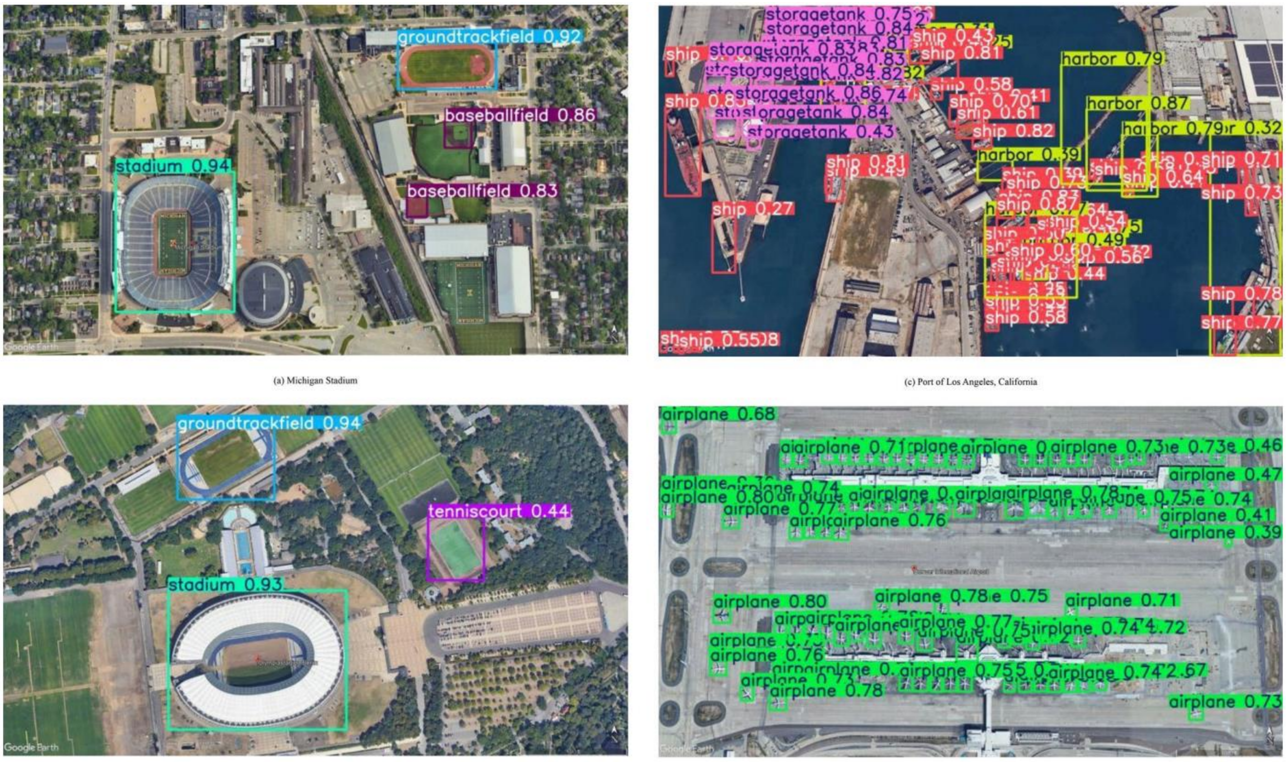
$$Confidence = P(Object) \times IoU_{predicted,true}, \quad (1)$$

where $P(Object)$ – the probability predicted by the model that the region contains a target object$

$IoU_{predicted,true}$ – the intersection over union between the predicted bounding box and the ground-truth bounding box, which quantifies the degree of overlap.

In practice, the model is optimized during training to maximize this metric. During testing, the trained YOLOv11 model demonstrated high performance: the average confidence level for detected objects was over 85% for 80% of the test dataset samples [2].

Four representative sites were chosen from Google Earth in article to thoroughly assess the model's performance in different scenarios: Michigan Stadium, Olympiastadion Berlin, the Port of Los Angeles, and Denver International Airport (Fig. 1).



(a) Michigan Stadium (c) Port of Los Angeles, California

(b) Olympiastadion Berlin (d) Denver International Airport

Figure 1 – Application case test

Defining the boundaries and identifying the characteristic features of the Michigan Stadium and Olympic Stadium in Berlin is a challenging task due to their complex architectural design and numerous details, which are often accompanied by obstacles such as trees and cars [3]. Thus, in this case, the ability of YOLOv11 to work with complex geometry and obstacles was evaluated. As a result, the model successfully identified related objects in complex contexts, accurately determined boundaries, and reproduced the general contours of these structures. These results demonstrate the flexibility and stability of the model in working with complex shapes and in multitasking situations.

Ports are dynamic landscapes with a wide range of objects of varying sizes and densities, including land vehicles and ships that sometimes disappear against the backdrop of water. Therefore, testing focused on the model's ability to manage the recognition of numerous categories and closely spaced objects for the Los Angeles port area.

The model demonstrated high proficiency in detecting multiple objects, accurately identifying densely packed containers at terminals, distinguishing moored vessels from moving vehicles, and performing consistently in these challenging conditions. This demonstrates its high effectiveness in solving tasks related to multi-class classification and localization in complex environments.

Runways, terminals, aprons, and objects such as aircraft and ground vehicles are all part of an airport. These objects, which have static and dynamic aspects, are located in a large area and are characterized by mixed traffic situations. As a result, the main challenge for Denver International Airport was the effectiveness of the model in a spacious, dynamic environment. Despite obstacles such as changing lighting and shading, the model was able to accurately detect both dynamic objects such as aircraft on runways and stationary structures such as terminals. These results demonstrate how flexible the model is and how well it performs in extremely complex and dynamic conditions.

Conclusion. The overall test results highlight the model's ability to generalize and its reliability in complex conditions, demonstrating its high performance in situations with a large number of targets, diverse backgrounds, and a combination of dynamic and static conditions. The model's significant effectiveness in multi-class detection tasks indicates its promise for real-world remote sensing applications. In addition, the tests provide valuable information about areas that require further optimization. For example, to further improve detection accuracy in congested conditions, background suppression methods can be strengthened and resolution for small targets can be improved. More complex applications in traffic monitoring, port management, and urban planning may become possible in dynamic scenarios by improving the model's ability to process sequential data through the integration of temporal information and trajectory prediction approaches.

References

1. Khanam R., Hussain M. YOLOv11: An Overview of the Key Architectural Enhancements. Ithaca, NY: Cornell University, 2024. 9 p. (Preprint. arXiv:2410.17725). DOI: 10.48550/arXiv.2410.17725 (date of access: 11.10.2025).

2. YOLO Evolution: A Comprehensive Benchmark and Architectural Review of YOLOv12, YOLO11, and Their Previous Versions / N. Jegham et al. Ithaca, NY: Cornell University, 2025. 20 p. (Preprint. arXiv:2411.00201). DOI: 10.48550/arXiv.2411.00201 (date of access: 11.10.2025).

3. Research on object detection and recognition in remote sensing images based on YOLOv11 / L.-h. He et al. *Scientific Reports*. 2025. Vol. 15, № 1. DOI: 10.1038/s41598-025-96314-x (date of access: 11.10.2025).