UDC 004.8: 629.7.05: 623.67 **doi: 10.32620/reks.2024.4.09**

Vadym MISHCHUK, Herman FESENKO, Vyacheslav KHARCHENKO

National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine

DEEP LEARNING MODELS FOR DETECTION OF EXPLOSIVE ORDNANCE USING AUTONOMOUS ROBOTIC SYSTEMS: TRADE-OFF BETWEEN ACCURACY AND REAL-TIME PROCESSING SPEED

***The study focuses on** deep learning models for real-time explosive ordnance detection (EO). **This study aimed** to evaluate and compare the performance of YOLOv8 and RT-DETR object detection models in terms of accuracy and speed for EO detection via autonomous robotic systems. **The objectives** are as follows: 1) conduct a comparative analysis of YOLOv8 and RT-DETR image processing models for explosive ordnance (EO) detection, focusing on accuracy and real-time processing speed;2) to explore the impact of different input image resolutions on model performance for identifying the optimal resolution for EO detection task s;3) to analyze how object size (small, medium, large) affects detection efficiency for enhancing EO recognition accuracy; 4) to develop recommendations for EO detection model configurations; 5) to propose methods for enhancing EO detection model performance in complex environments. The following **results** were obtained. 1) The results of a comparative analysis of YOLOv8 and RT-DETR models for EO detection in the context of speed-accuracy trade-offs. 2) Recommendations for EO detection model configurations aimed at improving the efficiency of autonomous demining robotic systems, including optimal camera parameter selection. 3) Methods for improving EO detection model performance to increase its accuracy in complex environments, including synthetic data generation and confidence threshold tuning. **Conclusions**. The main contribution of this study is the results of a detailed evaluation of the YOLOv8 and RT-DETR models for real-time EO detection, helping to find trade-offs between the speed and accuracy of each model and emphasizing the need for special datasets and algorithm optimization to improve the reliability of EO detection in autonomous systems.*

***Keywords:** explosive ordnance; object detection; precision; performance; YOLO; transformers.*

1. Introduction

1.1. Motivation

Real-time object detection is a critical component in various applications, ranging from autonomous driving to security surveillance. In particular, explosive ordnance detection, such as landmines and unexploded ordnance (UXO), requires highly accurate and efficient object detection models to ensure human safety and operational effectiveness.

The use of modern technology in demining can revolutionize the field, offering significant improvements in efficiency and safety. Automated systems, such as those that integrate robotic and biological components [1], enhance the detection and identification capabilities of explosive ordnance. Unmanned aerial vehicles (UAVs) equipped with advanced imaging technologies, such as thermal cameras and high-resolution optical cameras, have shown great promise in explosive ordnance (EO) detection. These UAVs can be part of robotic-biological systems, thus enhancing their capabilities. In addition, Machine learning algorithms can detect EO in various environments by analyzing footage from these cameras. This method is particularly useful for wide-area surveys, where manual detection is time-consuming and hazardous.

While some tasks, like humanitarian demining of agricultural fields or recreational zones, are not time-critical, the speed of object detection is crucial in scenarios where rapid decision-making and fast response are vital for success. For instance, planning evacuation routes for civilians in combat zones requires real-time detection of threats like explosives to ensure that the safest possible path is quickly identified. The dynamic nature of such environments requires continuous monitoring, where delays can endanger lives.

Autonomous robots and UAVs can also be used for real-time object detectors. Because their on-board computing capabilities are limited, they must rely on fast lightweight detection algorithms to navigate hazardous terrains and accomplish missions such as search and rescue. In all these scenarios, striking a balance between speed and accuracy is crucial, with an emphasis on ensuring that detection systems operate quickly enough within specific constraints on detection accuracy.

Ensuring trade-offs between the accuracy and speed of real-time object detectors for EO detection provides many important advantages. First, optimized systems allow sufficient data processing speed, which is critical for rapid response in high-risk situations. This reduces the time delay of decision-making and enables immediate danger detection. Secondly, maintaining an acceptable accuracy level reduces the number of false alarms and minimizes the likelihood of missing explosive ordnance, which increases the overall reliability of the system.

This trade-off, along with the active use of advances in machine learning and edge computing, will allow for the development of solutions that can operate in resource-limited environments on mobile platforms such as UAVs, unmanned ground vehicles (UGVs). This will facilitate the rapid deployment of the EO detection system in the appropriate locations, increase its flexibility and adaptability to changing operational environments, and allow it to be scaled up as needed.

1.2. State of the art

The application of deep learning techniques to UAV-based detection systems has been demonstrated in several studies. For instance, authors of [2] presented a study focusing on the detection of scatterable landmines using UAVs equipped with multispectral and thermal imaging systems. Their methodology, which employs a Faster R-CNN (Region-based Convolutional Neural Network) model, is calibrated for detecting scatterable plastic landmines, such as PFM-1, and has shown promising results in automating landmine detection through supervised learning algorithms.

In another study [3], the authors proposed a real-time detection system for surface landmines that uses optical imaging integrated within a demining robot. This system uses YOLOv8 (You Only Look Once) models to achieve high recall rates in detecting PFM-1 (butterfly) and PMA-2 (starfish with tripwire) landmines. The integration of this system into a demining robot demonstrates the potential of combining optical imaging and artificial intelligence to enhance demining operations. Both “nano” and “small” YOLOv8 models used in this study demonstrated strong capacity for detecting objects similar to those in the training set. However, the relatively high false positive rates suggest that further optimization is required.

Object detection has seen significant advancements over the past few decades, transitioning from classical methods to modern deep learning-based models. The study [4] provides a comprehensive review of the field, covering the evolution of object detection from the 1990s to 2022. This review discusses milestone detectors, key datasets, evaluation metrics, and speed-up techniques, providing a detailed understanding of the advances and current state-of-the-art object detection techniques. Classic object detection methods often relied on handcrafted features and simple classifiers; however, they have largely been superseded by deep learning-based models due to the latter's superior performance and ability to learn features autonomously. The authors have identified lightweight object detection as a promising future direction. They mentioned that despite significant progress in recent years, there is still a considerable speed gap between machine detection and the human eye, particularly when it comes to identifying small objects or processing information from multiple sources.

Several studies have conducted comparative analyses of different object detection models to evaluate their performance across various parameters. For instance, authors of [5] evaluated YOLOv3, YOLOv5, and YOLOX models on edge computing devices such as NVIDIA Jetson Nano and Google Coral Dev Board. In this study, the MS COCO dataset was used for evaluation, providing insights into the trade-offs between accuracy and inference speed across different devices and input sizes.

A previous study [6] analyzed deep learning algorithms in the context of smart cities to identify accurate models for urban object detection, emphasizing the importance of real-time performance. The findings highlight the challenges of achieving high accuracy in urban scenes while meeting real-time requirements, with Dynamic Head emerging as a top-performing model due to its high precision and recall at medium IoU thresholds. Authors stress that high computational demands of advanced algorithms, like Dynamic Head, could also present challenges for real-time deployment in resource-limited tasks. Additionally, established metrics like AP and mAP, while useful, may not fully reflect the peculiarities of the subject area, suggesting the need for more context-specific evaluations.

In [7], the authors compared three major image processing algorithms Single Shot Detection (SSD), Faster R-CNN, and YOLO to determine the fastest and most efficient model. Their results indicated that YOLOv3 outperforms SSD and Faster R-CNN in terms of speed and efficiency, making it a preferred choice for real-time applications.

In study [8], an in-depth analysis of various object detection algorithms. This study evaluates different models on different datasets, focusing on key factors, such as accuracy, speed, resource usage, and robustness. The results also highlight the importance of understanding how different variables, such as backbone architecture and image size influence the performance of these detectors. Notably, the study finds that keypoint-based detectors, like NanoDet, generally perform well across multiple metrics, offering a good balance between accuracy and speed while being resource-efficient. In contrast, anchor-based detectors such as SSD and YOLO are more optimized for edge devices, making them suitable for applications with limited computational resources. Overall, this survey serves as a guideline for selecting suitable object detection models for specific applications, considering both their strengths and potential limitations.

The authors of the study [9] provided a guide for selecting object detection architectures based on specific speed, memory, and accuracy requirements. It evaluates Faster R-CNN and SSD models, highlighting that SSD with MobileNet is the fastest, making it suitable for real-time applications, while Faster R-CNN with Inception Resnet offers the highest accuracy but at a slower speed. This paper also explores the impact of image resolution and proposal numbers on performance and demonstrates how speed can be increased with minimal accuracy loss. Although they are insightful, their findings are somewhat outdated due to advancements in object detection technology since their release.

In the study [10], is presented YOLOv10, an advancement over previous YOLO models, focusing on improved efficiency and performance. The proposed method introduces a dual assignment strategy for NMS-free training to optimize accuracy and inference speed while reducing computational redundancy. Achieving state-of-the-art results, YOLOv10 outperforms its predecessors and contemporary models in terms of speed and parameter reduction. This study highlights the need for optimal confidence thresholds and future improvements in spatial and label quality for high-accuracy applications such as UAVs, static cameras, and orbital sensors.

The performance evaluation of object detection algorithms requires a thorough understanding of various metrics. The study [11] presents an extensive review of the most frequently used metrics for object detection, highlighting their differences, applications, and main concepts. The proposed method proposes a standard implementation for benchmarking different datasets with minimal adaptation to annotation files. The proposed comprehensive evaluation framework is crucial for consistent performance assessment across various models and datasets.

The authors of the study [12] addressed an often overlooked aspect of object detection: selecting the optimal confidence score threshold for model deployment. Typically, models are optimized for benchmark datasets like COCO, which favors low thresholds to maximize detection scores, resulting in a higher number of false positives. However, this approach may be inadequate in scenarios in which high confidence is crucial. The authors proposed a method to identify the optimal performance points of models, thereby enabling fairer comparisons and deeper insights into the trade-offs between true positives, false positives, and false negatives. They highlight the importance of balancing accuracy and efficiency, particularly for edge devices such as UAVs or static cameras, where model selection and confidence thresholds are critical.

The article [13] addresses the critical challenge of landmine detection and removal, focusing on UAV-based Airborne Magnetometry for identifying magnetic anomalies. The proposed method highlights the integration of edge computing for real-time data analysis to enhance the efficiency, security, and decision-making of landmine detection processes. The proposed Magnetometry Imaging-based Classification System (MAGICS) demonstrated high accuracy by leveraging deep learning to achieve a mean average precision of 97.8% for landmine identification.

These studies demonstrate significant advancements in deep learning applications for real-time detection, particularly EO detection. Techniques like Faster R-CNN and YOLOv8 exhibit high detection accuracy and are optimized for speed in field applications; however, they often have trade-offs in precision. Many of these studies emphasized the need to balance accuracy and speed, especially for edge devices, making the compromise between high precision and fast inference a central focus (Table 1).

1.3. Objectives and the methodology

The aim of this study was to evaluate and compare the performance of YOLOv8 and RT-DETR object detection models in terms of accuracy and speed for EO detection via autonomous robotic systems.

The objectives are as follows:

1) to perform a comparative analysis of YOLOv8 and RT-DETR image processing models for EO detection, focusing on accuracy and real-time processing speed;

2) explore the impact of different input image resolutions on model performance to identify the optimal resolution for EO detection tasks;

3) to analyze how the object size (small, medium, large) affects detection efficiency to enhance EO recognition accuracy;

4) to develop recommendations for EO detection model configurations;

5) to propose methods to enhance EO detection model performance in complex environments.

The research methodology is based on the principles of comparative analysis and integration of the YOLOv8 and RT-DETR models to identify EO in the context of speed-accuracy trade-offs.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

Table 1
Approaches to providing speed-accuracy trade-offs in the reviewed studies

Reference	Approach to providing speed-accuracy trade-offs
[2]	Faster R-CNN was optimized for UAV-based landmine detection, achieving high accuracy with manageable processing times for UAV deployment.
[3]	YOLOv8 models in a demining robot achieved high recall but need optimization to reduce false positives, emphasizing the speed-accuracy trade-off in real-time detection.
[5]	Studies on YOLOv3, YOLOv5, and YOLOX show that smaller configurations (e.g., "nano" versions) trade some accuracy for faster inference on edge devices like NVIDIA Jetson Nano.
[9]	Comparing Faster R-CNN and SSD, the authors found SSD with MobileNet to be faster but less accurate, making it more suitable for real-time applications prioritizing speed.
[10]	YOLOv10 enhances both speed and accuracy using a dual assignment strategy and NMS-free training, significantly lowering inference latency while maintaining performance.

The research methodology consists of the following steps:

1. Dataset selection and preparation. The dataset used in the experiment was a subset of the COCO 2014 validation dataset, which is a common benchmark for object detection research. This subset was selected randomly to reflect the diversity of object categories, sizes, and complexity of the environment. Cluttered scenes, multiple objects, and different lighting conditions were included in the data selection, and the variety required to test the robustness of the models.

2. Model selection and configuration. This study focused on two advanced object detection models, YOLOv8 and RT-DETR, due to their efficiency in balancing accuracy and inference speed. Each model was configured and tested using the Ultralytics YOLOv8 and RT-DETR frameworks implemented in the PyTorch environment.

3. Experimental design and input resolutions. The models were examined at 384, 448, 512, 576, and 640 pixels to evaluate the impact of input data resolution on model performance. The models were tested at each resolution to evaluate changes in mean accuracy (mAP) and inference time as the input size varied.

4. Object classification by size. To evaluate model performance across different object scales, the objects in the dataset were categorized into three size groups based on pixel area: small (less than 32^2 pixels), medium (32^2–96^2 pixels), and large (more than 96^2 pixels).

5. Choice of performance evaluation metrics. The primary evaluation metrics used to assess model performance were as follows:

- Mean Average Precision (mAP) to evaluate how accurately each model detects objects across different resolutions and sizes;
- Precision and Recall to provide a holistic view of model performance across the entire dataset;
- Inference Time to measure each model's suitability for real-time applications;
- Confidence Score Analysis to provide insight into how well each model is calibrated to detect EO, focusing on reducing false positives without compromising detection accuracy.

6. Data analysis. Experimental results were presented in

- tables showing mAP, precision, and recall at various resolutions and object sizes;
- charts to compare model performance in terms of detection reliability;
- scatter plots to assess model behavior across varying object scales and to help optimize confidence thresholds for each model based on detection requirements.

The article is structured as follows. Section 2 discusses the trade-off between accuracy and processing speed in real-time object detection models, highlighting challenges and solutions relevant to practical applications. Section 3 presents a comparative performance analysis of YOLOv8 and RT-DETR models, focusing on their behavior across different resolutions and object sizes. Section 4 discusses the implications of the findings, including their limitations and potential applications in real-world scenarios such as explosive ordnance detection. Section 5 concludes the paper by summarizing the key contributions, highlighting practical recommendations, and proposing directions for future research.

2. Accuracy vs. speed trade-off

Evaluating performance metrics across different object scales (small, medium, and large) is crucial in the context of landmine and UXO detection using UAVs or ground robots. UAVs capture images at varying altitudes, and ground robots encounter objects at different distances, leading to significant changes in object size. Assessing AP across scales ensures that the model performs well under diverse conditions, which enhances its robustness and reliability. This metric is critical for detecting small objects like landmines, where missing an object can have severe consequences. Evaluating mAP across scales also helps identify the model's strengths and weaknesses, thus guiding improvements for better overall detection. By ensuring high detection accuracy regardless of object size, AP across scales contributes to operational efficiency and safety in real-world applications.

The main challenge in real-time object detection is to optimize the trade-offs between accuracy and image processing time. Highly accurate models often require significant computational resources, which increases the output time. Understanding how different object detection models perform at different input signal resolutions is essential for selecting and tuning models that meet the requirements of the particular industry.

Real-time object detection models have improved significantly, with modern neural network architectures achieving impressive accuracy and speed. Advanced deep learning models, such as YOLO [14] and RT-DETR (Real-Time Detection Transformer) [15], offer a compelling balance between performance and accuracy:

- Known for their speed and efficiency, YOLO models process images in a single neural network pass.
- RTDETR uses a transformer architecture to improve detection capabilities, particularly in understanding the relationships between objects in an image and increasing detection accuracy. RTDETR models provide high accuracy, although often at the cost of increased computational requirements.

The purpose of this study is to evaluate and compare the performance of two state-of-the-art object detection models, YOLOv8 [16] and RT-DETR, and the degradation of their performance as the input image size changes; to demonstrate how these models balance accuracy and inference time and how object size affects their performance. Although the focus is on detecting objects of a "general nature", the results can be applied to critical areas such as landmine and unexploded ordnance detection, where fast and reliable object identification is crucial.

3. Performance analysis of detection models

The following outputs are produced:

- Comparison of the YOLOv8 and RT-DETR models in terms of accuracy and inference time.
- Assessment of the impact of different input data resolutions on model performance.
- Understanding how the size of objects in an image (small, medium, large) affects detection performance.

The dataset used in this study was a randomly selected subset of the COCO 2014 validation dataset. The original COCO dataset [17] is known for its diverse and large collection of annotated images, which were designed to facilitate object detection research.

The subset was randomly selected and shows an imbalance of classes, which is typical for such datasets. Despite this imbalance, the subset preserved the diversity and complexity of the original dataset, showing cluttered scenes, numerous objects, and different lighting conditions. The number of annotated objects in the original dataset is shown in Fig. 1.

The neural networks used for this study were implemented using the Ultralyics [16] YOLOv8 and RTDETR frameworks, which are known for their performance in object detection tasks. All experiments were performed on an NVIDIA RTX 2060 GPU. The runtime environment was PyTorch [18], with batch processing disabled. The confidence score threshold was left at a default value of 0.2.

The parameters evaluated in this study include the mean Average Precision (mAP) at different Intersection over Union (IoU) thresholds (0.5 and 0.75) and the average processing time for each image. The objects were classified as small (area less than 322 pixels), medium (area from 322 to 962 pixels), and large (more than 962 pixels) to evaluate the model's performance at different scales. We also included the percentage of correct predictions (precision) and the percentage of detected objects (recall) for the entire dataset without dividing it into classes to provide an understanding of the model's overall performance. Custom-made scripts based on [19] were used for performance evaluation. Results are shown in **Ошибка! Источник ссылки не найден.**.

As can be seen, YOLOv8l and YOLOv8x demonstrate high mean average precision with a high percentage of correct predictions and balanced percentage of detected objects. Both RTDETR models had higher mAP rates than the YOLOv8 models; however, the RTDETR model had a significantly higher false positive rate, with a significantly higher percentage of detected objects.

Compared to YOLOv8, the proposed RTDETR model also demonstrated high mAP for large and medium-sized objects, with noticeable improvements in detecting small objects.

Tables 2 to 5 summarize the performance of four object detection models (RTDETR-L, RTDETR-X, YOLOv8l, and YOLOv8x) at different input resolutions (384, 448, 512, 576, and 640 pixels). The evaluated performance includes the average accuracy at IoU thresholds of 0.5 and 0.75, as well as the average accuracy for large, medium, and small objects at IoU 0.5.

Predictably, a decrease in input resolution leads to a decrease in mAP@0.5 and mAP@0.75 for all models. RTDETR-L and RTDETR-X demonstrate high average accuracy at all resolutions, although performance

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

degrades very sharply with decreasing resolution.

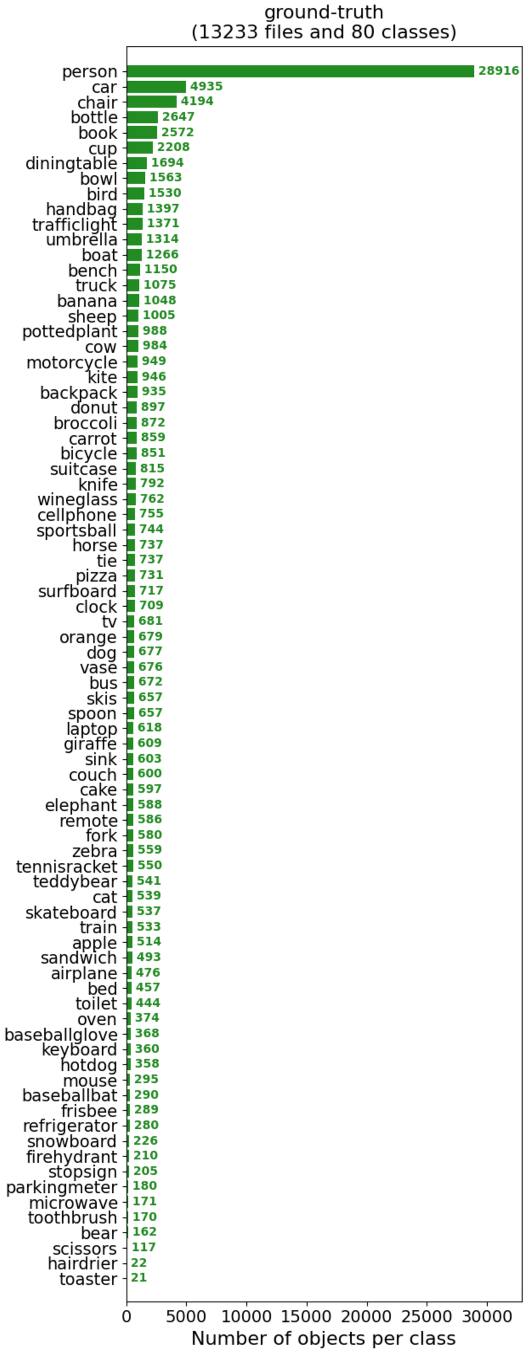

Fig. 1. Number of ground-truth objects per class in selected dataset

In the detection of small objects, RTDETR outperformed YOLO even at lower resolutions. All models performed well on large objects with less variability in mAP with decreasing resolution. The gap in average accuracy between RTDETR and YOLOv8 decreased with increasing localization accuracy requirements.

It is worth noting that the data in the tables represent average precision and should not be taken as absolute values. First, the MS COCO dataset is very diverse and contains a wide range of object categories, from animals to household items; however, when detecting EOs, the diversity of objects is much smaller. Second, the images in the COCO dataset have different backgrounds and environments; however, when it comes to finding EOs, the number of typical contexts is much more limited. The combination of these factors can improve the accuracy of detection methods in demining tasks. The degradation of mAP and the increase in performance associated with the change in resolution are shown in Figs. 2 and 3.

As can be seen, with a decrease in resolution, RTDETR models show a slight increase in performance compared to YOLO, but the mean average precision's degradation rate is much higher. This decrease may have occurred because the models were not retrained when scaling the resolution. As mentioned in [12], the numbers of false positives/negatives (and, consequently, precision and recall) are highly dependent on the selected confidence score threshold. To optimize this threshold, we analyzed the confidence score distributions for true and false positives (Fig. 4.) Table 6 provides a detailed characterization of the confidence scores for different object sizes (all, big, medium, and small) at IoU@0.5 across four models: YOLO8l, YOLO8x, RTDETR-l, and RTDETR-x. This shows the distribution of confidence scores by listing the first, median, and third quartiles for both true and false positives.

From the histogram of confidence scores in combination with the previous table, we can derive some common trends regarding the performance and behavior of YOLO and RTDETR models in terms of the detection confidence for both true and false positives:

1. The confidence scores for true positives in both YOLO models are heavily skewed toward the higher end of the confidence range. RTDETR models also show a similar right-skewed distribution for true positives but with a slightly larger volume of true positives compared to YOLO models. From these results, we conclude that both RTDETR and YOLO models are well-calibrated for true positives, providing high confidence when the model is correct.

2. YOLO had fewer false positives across all confidence values. Most false positives occur at lower confidence scores, with fewer false positives at high confidence. RTDETR models also show similar left-skewed distributions for false positives but with a much larger number of false positives compared to YOLO models.

3. Both models showed a similar tendency: larger detections have higher confidence scores for both false positives and true positives. Smaller detections result in lower confidence thresholds. However, with an increase in the object size difference between the median confidence scores for false positives and true positives, it is easier to filter out the majority of false positives with minimal effect on the true positive count.

Table 2

Object detectors' performance at original resolution

Model	mAP @0.5	mAP @0.75	Precision	Recall	Big object mAP@0.5	Medium object mAP@0.5	Small object mAP@0.5
YOLOv8l	70.92	63.53	77.04	72.86	85.04	68.56	33.59
YOLOv8x	72.55	65.39	76.53	74	85.92	70.44	36.5
RTDETR-L	78.14	66.04	48.15	84.24	87.74	73.19	45.43
RTDETR-X	78.32	66.74	48.67	84.61	87.78	73.75	44.01

Table 3

YOLOv8-X performance at different resolutions

Resolution	mAP@0.5	mAP@0.75	mAP(big)@0.5	mAP(medium) @0.5	mAP(small)@0.5
384	64.23	57.2	84.88	60.88	17.65
448	67.53	60.4	85.85	65.11	22.8
512	68.87	61.88	85.71	66.57	26.77
576	70.13	62.85	85.57	67.63	30.73
640	70.92	63.53	85.04	68.56	33.59

Table 4

RTDETR-L performance at different resolutions

Resolution	mAP@0.5	mAP@0.75	mAP(big)@0.5	mAP(medium) @0.5	mAP(small)@0.5
384	66.26	59.24	86.06	64.09	19.81
448	69.65	62.51	86.71	68.35	26.53
512	70.68	63.77	86.62	69.02	30.42
576	71.85	64.74	86.44	70.26	33.8
640	72.55	65.39	85.92	70.44	36.5

Table 5

RTDETR-X performance at different resolutions

Resolution	mAP@0.5	mAP@0.75	mAP(big)@0.5	mAP(medium)@0.5	mAP(small)@0.5
384	69.65	56.06	84.49	62.71	27.67
448	73.21	60.23	86.57	67.73	33.97
512	75.43	62.98	87.3	70.46	37.43
576	76.88	64.69	87.52	72.15	41.42
640	78.14	66.04	87.74	73.19	45.43

To further refine the confidence score threshold selection, we can examine the Scatter Plot of Confidence vs Detection Size should be examined (Fig. 5), showing the relationship between object size and confidence score. Understanding how confidence varies with object size provides deeper insight into selecting appropriate thresholds for each model.

The scatter plot of the confidence versus detection size reveals the key trends of the YOLO and RTDETR models. In YOLO models, true positives consistently show higher confidence scores, especially for larger objects, whereas false positives generally have lower confidence; thus, it is easier to filter out false positives by setting a threshold of approximately 0.6. In contrast, RTDETR models displayed higher confidence scores for false positives, particularly in medium-sized detections, indicating that thresholding was less effective without considering detection size. Although both models showed an increase in confidence for larger objects, the overlap of confidence scores between true and false positives was more pronounced in RTDETR, indicating that size-based filtering might be necessary to boost precision. Thus, YOLO benefits more from simple confidence thresholding, whereas RTDETR requires a combination of strategies to achieve similar precision improvements.

Based on these plots and tables, the key recommendation for selecting a camera with a fixed field of view, resolution, and flight altitude for a drone searching for EO is to optimize for medium to large object detection. Both YOLO and RTDETR models demonstrate higher confidence scores for larger objects, particularly true positives, which means that setting the drone's altitude and field of view to capture larger detection areas will likely improve detection precision.

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

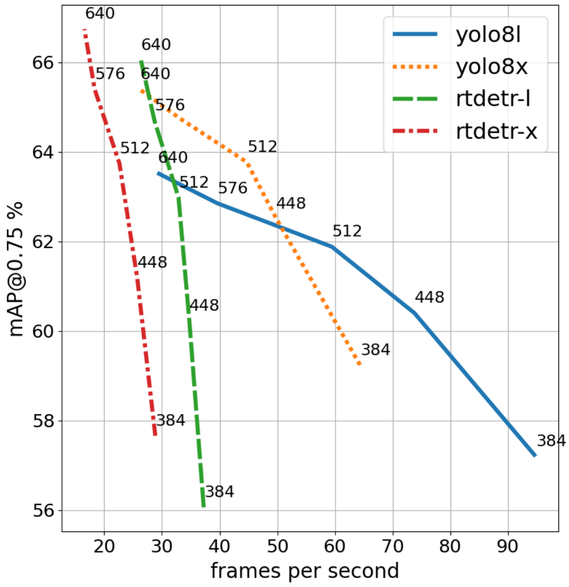

Fig. 1. Precision vs inference speed curves for different resolutions at IoU=0.75

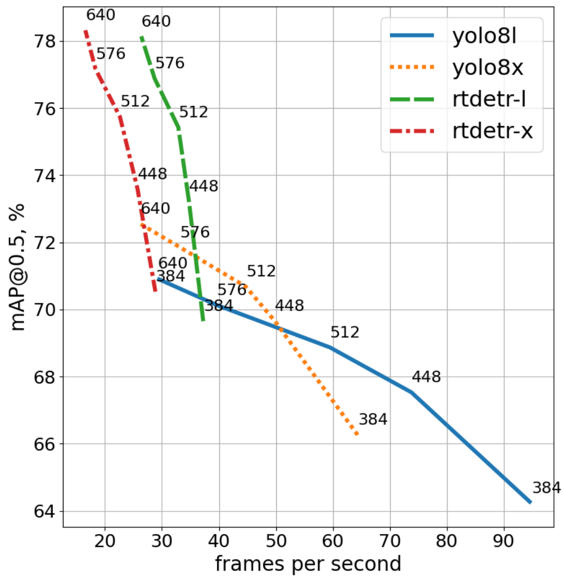

Fig. 2. Precision vs inference speed curves for different resolutions at IoU=0.5

Table 6

Confidence Score Distribution for YOLO and RTDETR Models by Object Size

Metrics		Models			
		YOLO8l	YOLO8x	RTDETR-L	RTDETR-X
All objects IoU@0.5					
True positives	First quartile	0.599	0.615	0.557	0.559
	Median	0.809	0.821	0.813	0.813
	Third quartile	0.907	0.914	0.918	0.917
False positives	First quartile	0.304	0.304	0.281	0.281
	Median	0.388	0.39	0.327	0.327
	Third quartile	0.526	0.537	0.412	0.413
Big objects IoU@0.5					
True positives	First quartile	0.83	0.846	0.854	0.851
	Median	0.911	0.919	0.927	0.926
	Third quartile	0.941	0.946	0.952	0.951
False positives	First quartile	0.328	0.33	0.287	0.287
	Median	0.448	0.454	0.346	0.347
	Third quartile	0.649	0.663	0.46	0.46
Medium objects IoU@0.5					
True positives	First quartile	0.578	0.596	0.56	0.558
	Median	0.767	0.784	0.785	0.785
	Third quartile	0.86	0.868	0.883	0.883
False positives	First quartile	0.315	0.319	0.287	0.285
	Median	0.414	0.423	0.342	0.341
	Third quartile	0.579	0.601	0.444	0.442
Small objects IoU@0.5					
True positives	First quartile	0.407	0.413	0.389	0.390
	Median	0.567	0.581	0.546	0.556
	Third quartile	0.721	0.731	0.732	0.736
False positives	First quartile	0.295	0.295	0.277	0.277
	Median	0.366	0.364	0.317	0.317
	Third quartile	0.487	0.49	0.391	0.392

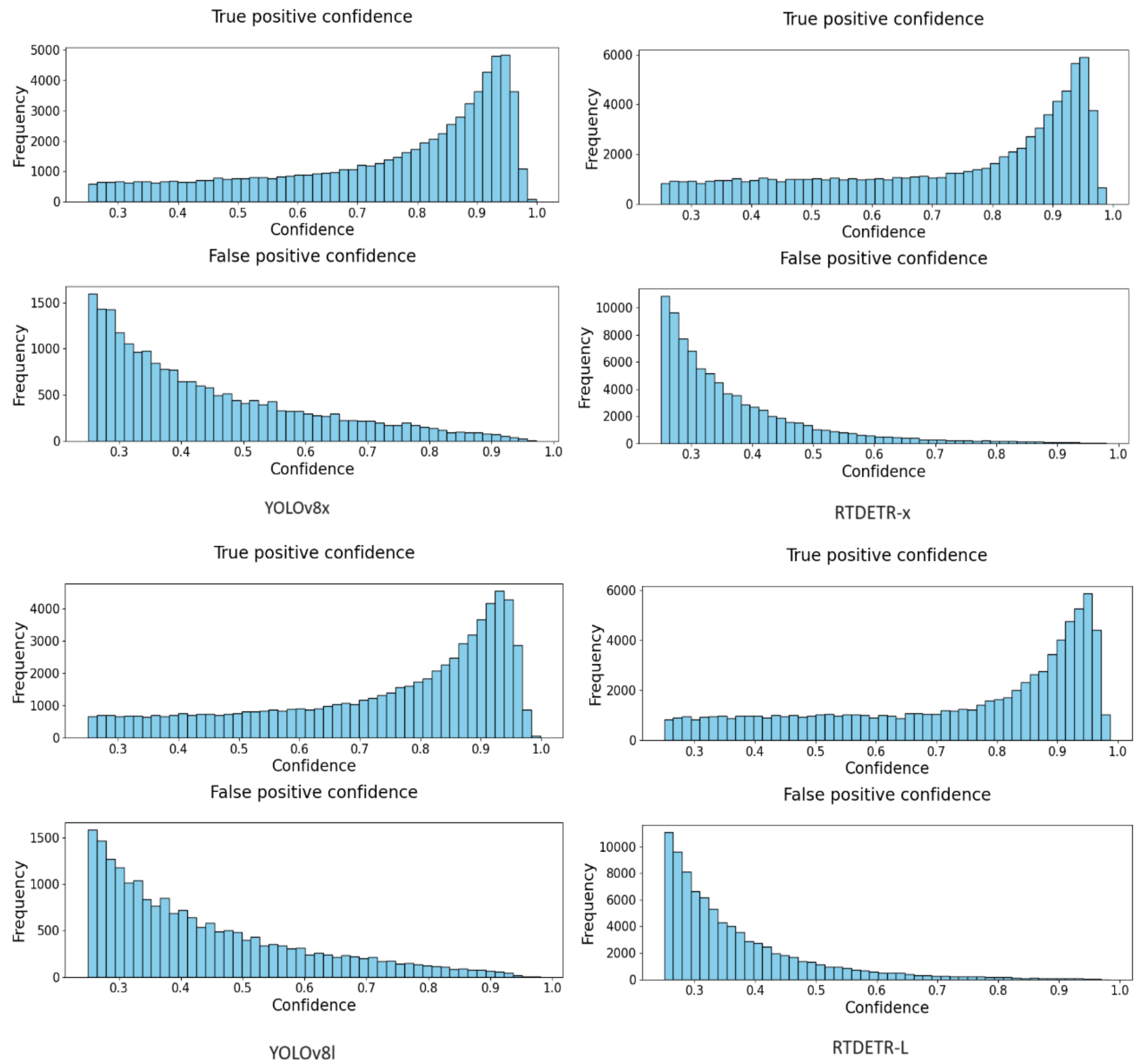

Fig. 4. Distribution of detections confidence scores for different models for objects of all sizes

Given that false positives in RTDETR are more frequent, especially in medium-sized objects, it is important to fine-tune the flight altitude to a level where the EO values are sufficiently large in the image to trigger higher confidence in true positive detections. A higher camera resolution is also essential to resolve small or medium-sized EO better because the detection accuracy for small objects is more challenging for both models, especially in RTDETR.

4. Discussion

This study utilized a subset of the COCO 2014 validation dataset, which is known for its wide range of object categories, environments, and image contexts. However, the unique and repetitive background and object characteristics that often occur in EO detection scenarios may not be fully covered by this generic dataset. The complexity and chaos in COCO images provide a good benchmark for overall performance; however, EO images are limited, meaning that detection accuracy may vary in real-field conditions.

To analyze the performance of the model at different scales, this study used a range of input resolutions from 384 to 640 pixels. Real-world EO detection applications may benefit from even higher resolutions, especially for small or partially enclosed objects. This is even though it is useful for understanding the trade-off between accuracy and processing time. Thus, limiting testing to these resolutions may result in a limited understanding of the full potential of detection models at higher resolutions.

To balance detection sensitivity with false positives, the default confidence threshold for detection was set to 0.2. However, EO detection applications require higher confidence levels to mitigate risks in high-stake environments and reduce false positives. The threshold affects the generalizability of the results

ISSN 1814-4225 (print)
ISSN 2663-2012 (online)

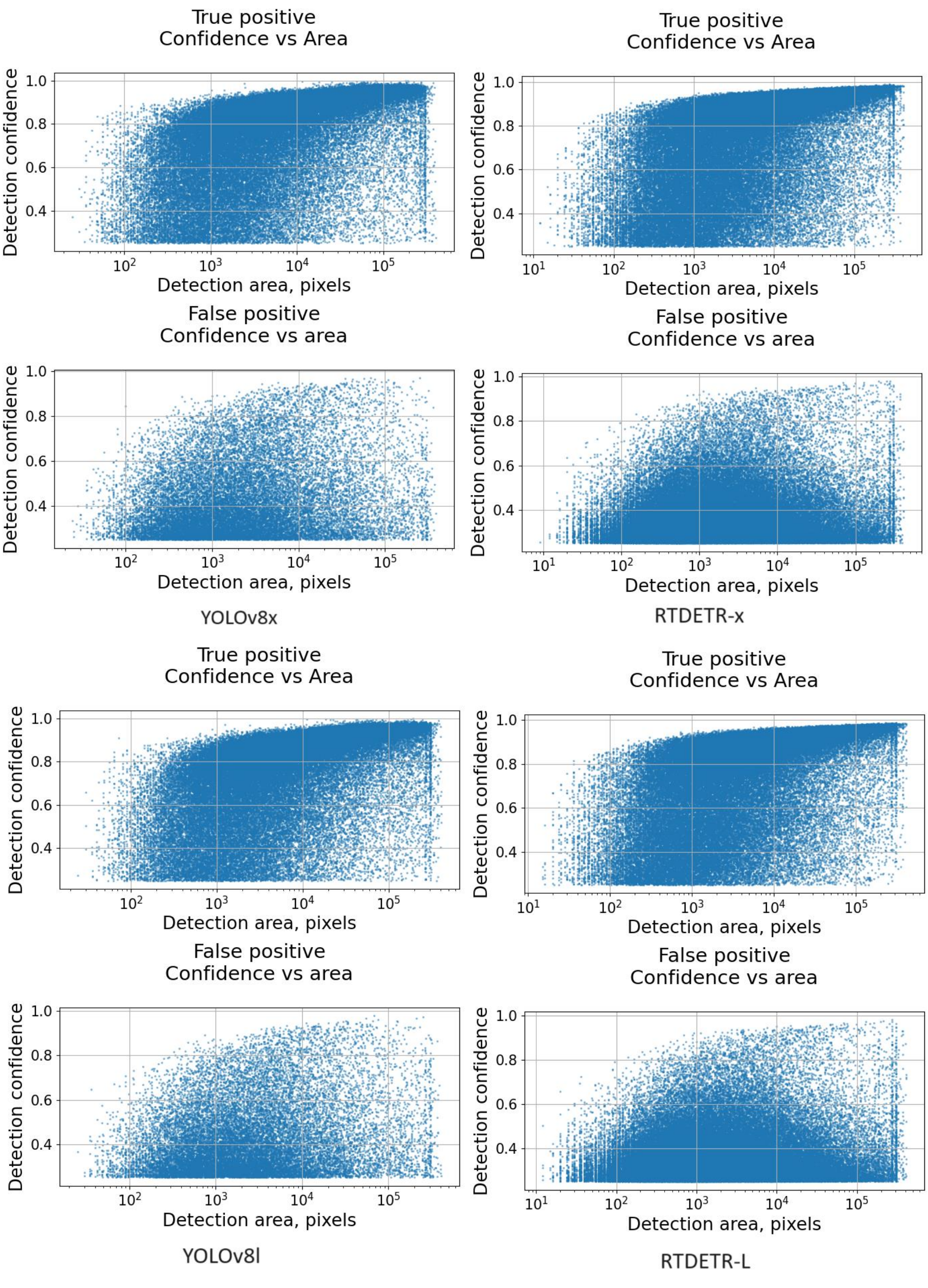

Fig. 3. Scatter plot of confidence vs detection size for different models

becauset he optimal confidence level for fault detection is likely to depend on the facility size, environmental conditions, and model type.

Although YOLO models are designed for speed and efficiency, they do not perform as well on smaller objects as RT-DETR, which uses a transformer-based architecture to capture the relationships between objects. However, the higher computational requirements of RT-DETR may not be feasible to install on every UAV or peripheral used for EO detection, limiting its practical application.

This study assesses small, medium, and large objects, but it assumes a balanced distribution between these categories. Nevertheless, the EO detection may assume a higher detection rate for smaller objects, such as landmines. This may affect the accuracy of these models in real-world situations, where the model may have trouble maintaining extremely high accuracy for much smaller objects.

5. Conclusions

The main contribution of this study is the results of a detailed evaluation of the YOLOv8 and RT-DETR models for real-time EO detection, helping to find trade-offs between the speed and accuracy of each model and emphasizing the need for special datasets and algorithm optimization to improve the reliability of EO detection in autonomous systems.

The real-time object detection models YOLOv8 and RT-DETR demonstrate effective trade-offs between accuracy and inference time. The mean precision of the RT-DETR models was higher than average; RT-DETR-X reached 78.32% at an IoU of 0.5, outperforming YOLOv8x, which had 72.55%. However, the RT-DETR models showed higher false alarm rates, indicating that the performance must be balanced.

The input resolution significantly affects the detection performance. YOLOv8x achieved a mAP@0.5 of 72.55% at the highest test resolution of 640 pixels, whereas RT-DETR-X achieved a mAP@0.5 of 78.32%. When the resolution was reduced to 384 pixels, mAP@0.5 decreased to 66.26% for YOLOv8x and 70.53% for RT-DETR-X, indicating a significant performance degradation when the resolution was reduced.

The RT-DETR model performed well for large objects, with RT-DETR-X achieving 87.78% mAP@0.5. YOLOv8x also detects large objects well, achieving 85.92% mAP@0.5. However, it lags behind RT-DETR in detecting small objects.

YOLO models excel at output speed, processing images in a single pass. This feature makes it more suitable for time-sensitive applications, where the additional computational requirements of RT-DETR may present limitations despite its higher accuracy for certain tasks.

The transformer-based architecture of RT-DETR improves the understanding of object relationships in images, thereby improving detection accuracy but leading to higher computational requirements. For example, RT-DETR-L achieved 78.14% mAP@0.5, although it had a higher false positive rate than YOLO models.

The COCO 2014 subset validation dataset, which preserves diversity with complex backgrounds and varying object scales, is ideal for testing object detection. However, this generic dataset may not be fully contextualized for explosive detection tasks where constant backgrounds and object types are common.

According to the confidence estimate analysis, the YOLO models maintain a high calibration for true positives; for large sites, the average confidence estimate is approximately 0.92. However, RT-DETR models show a larger discrepancy between true and false positive estimates, which requires the use of deeper filtering strategies to achieve a higher level of accuracy.

For UAV-based detection applications, optimizing the camera resolution and altitude improves the detection accuracy of medium- and large objects. The YOLO and RT-DETR models improve the detection reliability of large objects.

Applying the above results to the field of humanitarian demining, it can be concluded that the design of an optoelectronic system and the planning of UAV or ground-robot routes should be carried out, because increasing the size of the EO in an image will positively affect the detection accuracy. This can be achieved using high-resolution sensors, lenses with a narrow field of view, and a camera height above the ground. However, meeting such requirements means that either the number of images that need to be processed to survey the same area or the resolution of these images will increase. This increases the computing power requirements of robotic systems.

Future research can include:

- minimizing false positives in RT-DETR and YOLO models, especially in complex EO detection scenarios, to improve accuracy and reliability;
- training models on specialized datasets tailored to EO environments rather than general datasets like COCO, to increase model adaptability and precision in relevant contexts;
- in real-world conditions, exploring the impact of incorporating multispectral or thermal imaging to enhance detection accuracy, particularly for smaller or camouflaged objects.

Since reliability and autonomy are important characteristics for the systems under consideration, providing the speed-accuracy-reliability-autonomy trade-off can also be considered a promising area of research.

Contributions of authors: conceptualization, methodology, formulation of tasks – **Vadym Mishchuk**, **Herman Fesenko**, **Vyacheslav Kharchenko**; model selection and configuration – **Vadym Mishchuk**; verification, analysis of results, visualization, writing, original draft preparation – **Vadym Mishchuk**; review and editing – **Herman Fesenko**, **Vyacheslav Kharchenko**.

Conflict of Interest

The authors declare that they have no conflict of interest concerning this research, whether financial, personal, authorship or otherwise, that could affect the research and its results presented in this paper.

Financing

This research was conducted without financial support.

Data Availability

Data will be made available upon reasonable request.

Use of Artificial Intelligence

The authors confirm that they did not use artificial intelligence methods while creating the presented work.

All authors have read and agreed to the publication of the finale version of this manuscript.

References

1. Fedorenko, G., Fesenko, H., Kharchenko, V., Kliushnikov, I. & Tolkunov, I. Robotic-biological systems for detection and identification of explosive ordnance: concept, general structure, and models. *Radioelectronic and Computer Systems*, 2023, no. 2, pp. 143 –159. DOI:10.32620/reks.2023.2.12.

2. Baur, J., Steinberg, G., Nikulin, A., Chiu, K., & de Smet, T. S. Applying Deep Learning to Automate UAV-Based Detection of Scatterable Landmines. *Remote Sensing*, 2020, vol. 12, no. 5, article no. 859. DOI:10.3390/rs12050859.

3. Vivoli, E., Bertini, M., & Capineri, L. Deep Learning-Based Real-Time Detection of Surface Landmines Using Optical Imaging. *Remote Sensing*, 2024. vol. 16, no. 4, article no. 677. DOI:10.3390/rs16040677.

4. Zou, Z., Chen, K., Shi, Z., Guo, Y, & Ye, J. Object Detection in 20 Years: A Survey. *Proceedings of the IEEE*, 2023, vol. 111, no. 3, pp. 257–276. DOI: 10.1109/JPROC.2023.3238524.

5. Lema, D. G., Usamentiaga, R., & García, D. F. Quantitative comparison and performance evaluation of deep learning-based object detection models on edge computing devices. *Integration*, 2024, vol. 95, article no 102127. DOI: 10.1016/j.vlsi.2023.102127.

6. Tian, J., Jin, Q., Wang, Y., Yang, J., Zhang, S., & Sun, D. Performance analysis of deep learning-based object detection algorithms on COCO benchmark: a comparative study. *Journal of Engineering and Applied Science*, 2024, vol. 71, article no. 76. DOI: 10.1186/s44147-024-00411-z.

7. Srivastava, S., Divekar, A. V., Anilkumar, C., Naik, I., Kulkarni, V., & Pattabiraman, V. Comparative analysis of deep learning image detection algorithms. *Journal of Big Data*, 2021, vol. 8, article no. 66. DOI: 10.1186/s40537-021-00434-w.

8. Arani, E., Gowda, S., Mukherjee, R., Magdy, O., Kathiresan, S., & Zonooz, B. *A Comprehensive Study of Real-Time Object Detection Networks Across Multiple Domains: A Survey*. Available at: https://arxiv.org/abs/2208.10895 (accessed 20.11.2024).

9. Huang, J., Rathod, V., Sun, C., Zhu, M., Korattikara, A., Fathi, A., Fischer, I., Wojna, Z., Song, Y., Guadarrama, S. & Murphy, K. *Speed/accuracy trade-offs for modern convolutional object detectors*. Available at: https://arxiv.org/abs/1611.10012 (accessed 20.11.2024).

10. Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J. & Ding, G., 2024. *YOLOv10: Real-Time End-to-End Object Detection*. Available at: https://arxiv.org/abs/2405.14458 (accessed 20.11.2024).

11. Padilla, R., Passos, W. L., Dias, T. L. B., Netto, S. L., & da Silva, E. A. B. A Comparative Analysis of Object Detection Metrics with a Companion Open-Source Toolkit. *Electronics*, 2021, vol. 10, no. 3 article no. 279. DOI: 10.3390/electronics10030279.

12. Wenkel, S., Alhazmi, K., Liiv, T., Alrshoud, S. & Simon, M. Confidence Score: The Forgotten Dimension of Object Detection Performance Evaluation. *Sensors*, 2021, vol. 21, no 13, article no 4350. DOI:10.3390/s21134350.

13. Barnawi, A., Kumar, K., Kumar, N., Alzahrani, B., & Almansour, A. A Deep Learning Approach for Landmines Detection Based on Airborne Magnetometry Imaging and Edge Computing. *CMES - Computer Modeling in Engineering and Sciences*, 2024, vol. 139, no. 2, pp. 2117–2137. DOI:10.32604/cmes.2023.044184.

14. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A.. You Only Look Once: Unified, Real-Time Object Detection. *2016 IEEE Conf. Comput. Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788. DOI: 10.1109/CVPR.2016.91.

15. Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., Liu, Y. & Chen, J. *DETRs Beat YOLOs on Real-time Object Detection*. Available at: https://arxiv.org/abs/2304.08069 (accessed 20.11.2024).

16. Jocher, G., Chaurasia, A., & Qiu, J. Ultralytics YOLOv8. Available at: https://github.com/ultralytics/ultralytics (accessed 20.11.2024).

17. Lin, T.-Y., Maire, M., Belongie, S., Bourdev, L., Girshick, R., Hays, J., Perona, P., Ramanan, D., Zitnick, C.L. & Dollar, P. *Microsoft COCO: Common Objects in Context*. Available at: https://arxiv.org/abs/1405.0312 (accessed 20.11.2024).

18. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A.,

Chilamkurthy, S., Steiner, B., Fang, L., Bai, J. & Chintala, S. *PyTorch: An Imperative Style, High-Performance Deep Learning Library*. Available at: https://arxiv.org/abs/1912.01703 (accessed 20.11.2024).

19. Cartucho, J., Ventura, R. & Veloso, M., 2018. Robust Object Recognition Through Symbiotic Deep Learning In Mobile Robots. *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 2336–2341. DOI:10.1109/iros.2018.8594067.

Received 11.09.2024, Accepted 18.11.2024

МОДЕЛІ ГЛИБОКОГО НАВЧАННЯ ДЛЯ ВИЯВЛЕННЯ ВИБУХОНЕБЕЗПЕЧНИХ ПРЕДМЕТІВ ЗА ДОПОМОГОЮ АВТОНОМНИХ РОБОТИЗОВАНИХ СИСТЕМ: КОМПРОМІС МІЖ ТОЧНІСТЮ ТА ШВИДКІСТЮ ОБРОБКИ В РЕАЛЬНОМУ ЧАСІ

В. В. Міщук, Г. В. Фесенко, В. С. Харченко

Предметом дослідження є моделі глибокого навчання для виявлення вибухонебезпечних предметів (ВНП) у реальному часі. **Метою** цього дослідження є оцінка та порівняння продуктивності моделей виявлення об'єктів YOLOv8 і RT-DETR з точки зору точності та швидкості виявлення ВНП за допомогою автономних роботизованих систем. **Завдання**: 1) провести порівняльний аналіз моделей обробки зображень YOLOv8 і RT-DETR для виявлення ВНП, зосереджуючись на точності та швидкості обробки в реальному часі; 2) вивчити вплив роздільної здатності вхідного зображення на продуктивність моделі для визначення оптимальної роздільної здатності для завдань виявлення ВНП; 3) проаналізувати, як розмір об'єкта (малий, середній, великий) впливає на ефективність виявлення для підвищення точності розпізнавання ВНП; 4) розробити рекомендації щодо конфігурації моделі виявлення ВНП; 5) запропонувати методи підвищення ефективності моделі виявлення ВНП в складних середовищах. Були отримані наступні **результати**. 1) Результати порівняльного аналізу моделей YOLOv8 і RT-DETR для виявлення ВНП в контексті компромісів швидкодія-точність. 2) Рекомендації щодо конфігурації моделі виявлення ВНП, спрямовані на підвищення ефективності роботизованих автономних систем розмінування, зокрема, оптимальний вибір параметрів камери. 3) Методи підвищення продуктивності моделі виявлення ВНП для підвищення її точності в складних середовищах, включно генерацію синтетичних даних і налаштування порогу достовірності. **Висновки**. Основним внеском дослідження є детальна оцінка моделей YOLOv8 і RT-DETR для виявлення ВНП в реальному часі, підкреслюючи компроміси між швидкістю та точністю кожної моделі та наголошуючи на необхідності спеціальних наборів даних і оптимізації алгоритмів для покращення виявлення та надійності в автономних системах.

Ключові слова: вибухонебезпечні предмети; виявлення об'єктів; точність; швидкодія; YOLO; трансформери.

Міщук Вадим Валерійович– асп. каф. комп'ютерних систем мереж і кібербезпеки Національний аерокосмічний університет ім. М. Є. Жуковського «Харківський авіаційний інститут», Харків, Україна.

Фесенко Герман Вікторович – д-р техн. наук, проф., проф. каф. комп'ютерних систем, мереж і кібербезпеки, Національний аерокосмічний університет ім. М. Є. Жуковського «Харківський авіаційний інститут», Харків, Україна.

Харченко Вячеслав Сергійович – д-р техн. наук, проф., зав. каф. комп'ютерних систем, мереж і кібербезпеки, Національний аерокосмічний університет ім. М. Є. Жуковського «Харківський авіаційний інститут», Харків, Україна.

Vadym Mishchuk – PhD Student of the Department of Networks and Cybersecurity, National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine,
e-mail: v.v.mishchuk@csn.khai.edu, ORCID: 0009-0001-4423-9078.

Herman Fesenko – Doctor of Technical Science, Professor, Professor at the Department of Computer Systems, Networks and Cybersecurity, National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine,
e-mail: h.fesenko@csn.khai.edu, ORCID: 0000-0002-4084-2101, Scopus Author ID: 57190123735.

Vyacheslav Kharchenko – Doctor of Technical Science, Professor, Head of the Department of Computer Systems, Networks and Cybersecurity, National Aerospace University "Kharkiv Aviation Institute", Kharkiv, Ukraine,
e-mail: v.kharchenko@csn.khai.edu, ORCID: 0000-0001-5352-077X, Scopus Author ID: 22034616000.