Detection of military targets from images using deep learning models*

Dmytro Borovyk[1,*,†], Oleksandr Barmak[1,†] and Volodymyr Lytvynenko[2,†]

[1] *Khmelnytskyi National University, Instytutska 11, 29000, Khmelnytskyi, Ukraine*
[2] *Kherson National Technical University, Berislavske Shosse, 24, Kherson, 73008, Ukraine*

Abstract

This article states an approach to the automated detection and classification of military targets based on images, received from unmanned aerial vehicle (UAV) cameras. The approach is based on a multi-level architecture of deep learning models, which includes YOLOv11 and Faster R-CNN models. Three levels of classification are implemented within the framework of the study: (1) initial detection of objects of the type "human" and "military equipment", (2) detailed classification of military equipment by categories (tanks, infantry fighting vehicles, other), (3) refinement of the category "other" with division into multiple launch rocket systems (MLRS), trucks, etc.

The experimental study of the proposed approach included the analysis of 5000 images and 20 hours of video materials. The results obtained demonstrate the effectiveness of the proposed system: the Precision, Recall and F1-score metrics exceeded 91% for all classification levels. In addition, the speed of image processing is such that it allows applying the approach in real time. To assess the advantages of the proposed approach, a comparison was made with modern solutions which confirmed the competitiveness of the proposed one. The presented approach can be used in automatic monitoring systems, combat analysis and real-time situation assessment, providing a high level of accuracy and speed.

Keywords

Unmanned Aerial Vehicles (UAVs), military targets, target recognition, Faster R-CNN, YOLOv11, classification, deep learning.

1. Introduction

Artificial intelligence (AI) is actively used in modern military technologies. One of the tasks where the use of deep learning models can be a promising area of research is the task of detecting and recognizing military objects (potential targets). The mentioned task becomes most complex in the case of using unmanned aerial vehicles (UAVs) for monitoring, reconnaissance, aiming and carrying out strikes. This is explained by the fact that it is during the detection and recognition of potential targets that it is necessary to process large amounts of visual information in real time, which is a complex computational task. At the same time, object recognition is one of the most important tasks of aerial images using UAVs, as it provides the ability to identify and classify objects of interest (potential targets) in the obtained images, which can play a vital role in the course of military operations [1].

Nowadays, existing systems for searching and recognizing objects in video images are usually evaluated according to the following parameters [2]:

1. recognition accuracy - the system must provide stable detection and classification of objects;

Intelitsis'25: The 6th International Workshop on Intelligent Information Technologies & Systems of Information Security, April 04, 2025, Khmelnytskyi, Ukraine
* Corresponding author.
† These authors contributed equally.
dborovyk86@gmail.com (D. Borovyk); barmako@khmnu.edu.ua (O. Barmak); immun56@gmail.com (V. Lytvynenko);
0009-0001-5337-3519 (D. Borovyk); 0000-0003-0739-9678 (O. Barmak); 000-0002-1536-5542 (V. Lytvynenko)

2. processing speed - the system must perform the task in real time under conditions of limited resources for a particular computing device.

It is worth noting that not all methods, from among the existing ones, are effective according to the given parameters. Therefore, the current task is to develop approaches to improve existing systems or propose new systems for solving the problem of target detection and recognition.

The contribution of the study is the proposed approach to the detection and classification of military targets from aerial images, which basically uses a sequential method of target classification, which is different from existing solutions. This approach allowed to increase the accuracy of target classification under conditions of limited computing resources.

The structure of the article is as follows. The Related works section analyzes the latest publications on the topic under study, namely, the detection of military targets from images using deep learning. The Materials and methods section propose a sequential classification approach for military targets. The Results and discussion section present the results of experiments that confirm the effectiveness of this approach and compares it with existing solutions

2. Related works

Among modern deep learning architectures focused on object classification, a special place is occupied by such as R-CNN (Regions with CNN), Fast R-CNN, Faster R-CNN, YOLO (You Only Look Once), SSD (Single Shot Detector), MobileNet and SqueezeNet [3,4]. All these architectures belong to convolutional neural networks (CNN). CNNs automatically highlight discriminative features from images, learning on large data sets, which allows them to successfully recognize objects even in complex scenes with a rich background. For the successful application of CNN for the analysis of aerial images, it is important to consider the following aspects.

1. Representativeness of the data for training. Annotated aerial images should cover a wide range of object classes, perspectives, and environmental conditions, allowing the network to learn robust and generalizable features.
2. Optimization of the network architecture. The performance of CNNs largely depends on the choice of parameters such as network depth, convolutional kernels size, activation functions, etc. Regularization, data augmentation, and transfer learning techniques help reduce the risk of overfitting and increase generalizability.
3. Computational resources. CNNs require significant computing power, especially for training and real-time operation. The use of graphics processing units (GPUs) and parallel processing methods allows speeding up the computation, which is critical for practical applications.

Despite the significant computational requirements, CNN provides high recognition accuracy by automatically learning complex hierarchical representations. Successful use of CNN requires careful selection of architecture, high-quality training data, and optimization of resources to achieve real-time performance. Although the model provides high accuracy, its speed is very slow, making it unsuitable for real-time operation.

Fast R-CNN is an enhanced version of R-CNN. It generates regions from the internal representation received after processing the entire image, which accelerates operation. However, even this optimization does not provide the required speed for solving relevant real-time tasks [5].

Faster R-CNN is an enhanced version of Fast R-CNN that generates potential object locations and a convolutional network for their further classification and refinement. Due to process optimization, Faster R-CNN achieves high accuracy and efficiency. However, it still does not provide the necessary speed for addressing the studied tasks in real-time [6].

YOLO is a neural network that processes the entire image in a single pass by dividing it into a grid and predicting objects for each region. The YOLO network achieves high-speed object recognition through the simultaneous division of the image into a grid and the prediction of

bounding boxes and object classes. This network significantly enhances operational speed while maintaining accuracy. YOLO has demonstrated its effectiveness as a solution for aerial images, especially in scenarios where real-time object detection and tracking are critically important. The model operates significantly faster than R-CNN and its modifications [7].

Different algorithms can form the basis of YOLO models, which imparts specific characteristics to them.

In particular, the HSP-YOLOv8 model, more detailed described in the article [8], is popular. This model is applied for the detection of small objects. The authors [8] added an additional prediction head for small targets and a Space-to-Depth Convolution (SPD-Conv) module, which served to reduce information loss regarding the features of small targets and enhance detection accuracy. Experiments conducted on the VisDrone2019 dataset demonstrated an 11% increase in accuracy compared to the baseline YOLOv8s model.

Another notable model is the Improved YOLOv7, the description of which is presented in paper [9]. This model represents an enhancement of the YOLOv7 algorithm. The authors of [9] note that images from UAVs exhibit variable scales, uneven object distribution, and a high prevalence of small targets, which complicates their detection. The proposed improvements are aimed at increasing detection accuracy and speed, as well as reducing computational costs, which is critical for deployment on resource-constrained UAVs.

The UN-YOLOv5s model also merits attention. The UN-YOLOv5s model is oriented towards improving the detection of small targets in aerial images from UAVs. In [10], a more precise small object detection (MASD) mechanism and a multi-scale feature fusion (MCF) path are proposed to enhance the model's feature expression capability. Experiments on the VisDrone dataset showed an 8.4% increase in mean Average Precision (mAP) compared to the original YOLOv5s algorithm.

SSD, in its concept, is analogous to YOLO but employs the VGG16 architecture for feature extraction. SSD predicts bounding boxes and object classes at each level, achieving a balance between accuracy and speed, which makes it suitable for real-time applications. However, the use of this model on devices with limited computational resources is restricted [11].

Lightweight deep learning architectures, such as MobileNet and SqueezeNet, are designed for applications in computationally constrained environments, for example, on UAV platforms. They employ efficient techniques, including depthwise separable convolutions, quantization, and network pruning, to reduce model size and computational power requirements, while maintaining sufficient accuracy for real-time operation [12].

The conducted analysis of contemporary deep learning architectures focused on object recognition demonstrates that, among the existing models and architectures for real-time tasks, YOLO models are an optimal choice. Their speed and accuracy enable effective processing of video streams and object recognition even under complex conditions.

It should be noted that in addition to the object recognition algorithms mentioned above, there are other algorithms applicable to solving the problem under investigation.

A number of researchers, in their pursuit of improving deep learning methods aimed at object recognition, have focused their attention on various features, characteristics, approaches, and so forth. An evaluation of some of them will be conducted.

Thus, in the paper [13] the authors considered the use of various UAV detection and classification technologies based on machine and deep learning algorithms, and also investigated optimization models for real-time operation.

In the study [14] a multi-stage method for UAV detection and classification was proposed, which uses differences in communication signals between UAVs and controllers. A compressed sampling technique was used for signal processing, and two neural networks were developed for UAV detection and identification: a deep neural network (DNN) for detectors and a convolutional neural network (CNN) for UAV type and flight mode classification. The method was evaluated using 10-fold cross-validation, which confirmed its effectiveness in UAV detection and classification.

The study [15] focuses on land cover classification using UAV images using deep learning. The authors used convolutional neural networks (CNN) to classify different land cover types with high accuracy, which is important for environmental mapping and monitoring.

The study [16] develops an automated system for detecting flooded areas based on images obtained from UAVs. Using deep learning, the system analyzes aerial images to quickly and accurately detect flooded areas, which is critical when responding to natural disasters.

The paper [17] is devoted to evaluating the performance of deep learning methods for crop classification based on aerial images. It presents various deep learning architectures, including convolutional neural networks (CNN), long short-term memory (LSTM), and transformers, as well as techniques such as data augmentation, transfer learning, and multimodal fusion to improve the accuracy of the models.

The research [18] aimed to classify coastal marsh vegetation using high-resolution images obtained from UAVs. The authors used the U-Net architecture to accurately map vegetation at the local level, demonstrating the effectiveness of combining UAV-RGB data and deep learning in environmental monitoring tasks.

The authors of the research [19] worked out a method for multimodal detection, classification and 3D tracking of UAVs. Their proposed approach uses information from various sensors, including stereo vision, different types of lidars, radars and audio arrays, to accurately detect and classify UAVs. The authors developed a new classification system that includes sequence fusion, region of interest (ROI) cropping and key frame selection. The system integrates modern classification techniques and sophisticated post-processing methods to improve accuracy and reliability. Experiments confirmed the effectiveness and precision of their proposed approach.

The authors of the research [20] analyzed the performance of Vision Transformer (ViT) and Convolutional Neural Network (CNN)-based models for drone detection. The authors created different models based on CNN and ViT and demonstrated that for single drone detection, the basic ViT model can be more effective than CNN-based models. However, ViT requires more training data, computational resources, and more complex designs to fully outperform current CNN detectors.

In the study [21], the problem of UAV detection in images with a complex background and in the presence of rain artifacts was considered. The authors prepared two datasets: one with images of the sky as a background and the other with complex scenes. A test set of images with rain artifacts was also created. An evaluation of modern object detection methods was carried out to determine their effectiveness in complex conditions.

In the study [22], the application of computer vision algorithms based on deep learning for real-time processing of images from UAV onboard cameras is described. Four use cases are considered: detection, classification and localization of targets; segmentation of roads for autonomous navigation; segmentation of the human body; recognition of human actions. The algorithms were developed using modern image processing methods based on deep neural networks.

The analyzed studies demonstrate various aspects of the application of modern deep learning methods, in general, and the variety of approaches to the detection and classification of targets in aerial images and videos from UAVs.

The analysis of the application of deep learning methods, in general, and on the example of solving the problem of object recognition in video images, in particular, allows us to conclude that there are no universal methods that would be applicable to solving a wide class of problems. Therefore, in order to correctly choose a way to solve the problem under study, it is necessary to detail the latter and highlight those aspects in it that could help choose a way to solve the problem through improving existing methods or forming completely new ones.

Therefore, the purpose of this paper is to increase the accuracy of military targets classification under conditions of limited computing resources by developing a system for detecting and classifying military targets using UAVs in real time, which would provide increased efficiency according to the classical parameters of methods and systems for searching and recognizing objects in video images, minimize the shortcomings of each of the currently applicable models for solving similar problems, and use their advantages. To achieve the stated goal, the following research tasks were defined:

1. to propose an architecture for cascading target classification;
2. to develop a system that implements the proposed approach.

3. Materials and methods

The conducted above analysis suggests the feasibility of developing our own model to solve the formulated problem based on the model of the most modern version of YOLO, namely YOLOv11, and the Faster R-CNN model.

Analysis of the features of the problem under study allows us to propose a research hypothesis, which consists in the step-by-step and sequential application of the Faster R-CNN, YOLOv11s, YOLOv11m models trained separately on different datasets. Below the justification for the proposed assumption will be provided.

With multi-class classification using one model, there are often cases when a deep neural network does not distinguish the features of each class. Therefore, it is considered appropriate to abandon multi-class classification and use neural networks for classification with two or three classes. By reducing the number of object classes during training, each model can better learn and distinguish important features of each type. Each of the models should be trained on a dataset that corresponds to the sequential level at which it will be used. In addition, a feature of the sequential approach is that it can be expanded infinitely by adding new layers and applying models to recognize new types of military objects (vehicles).

Figure 1 presents the structure of the proposed sequential classification approach.

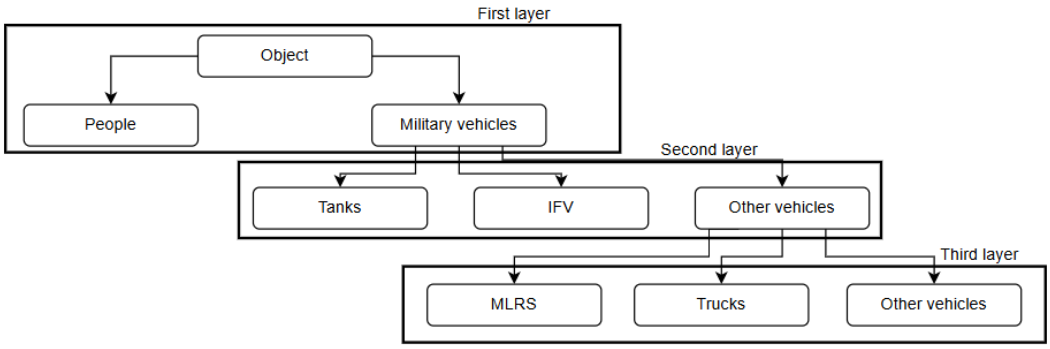

Figure 1: Structure of the proposed sequential classification approach.

The presented sequential classification approach consists of three levels.

Its implementation, as envisioned by the authors, is as follows. For the classification at the first level, a CNN model of the Faster R-CNN neural network was utilized. At this level, the identified objects are categorized into 2 classes: "Person" and "Military Equipment". The primary challenge of the first sequential classification level is the detection of "Person" class objects within images. This is attributed to the variability in human poses. Given the absence of a discernible consistent pattern for this class, the majority of neural networks fail to identify the class features and, consequently, do not recognize it in images. However, the Faster R-CNN model is advantageous due to its accuracy and sensitivity to small objects. Therefore, the application of this specific neural network model is considered appropriate at this classification level.

The Faster R-CNN model was trained on a synthetically generated dataset, which exclusively included the classes: "Person" and "Military Equipment." This dataset was created based on the Roboflow dataset, typically used for multi-class classification.

The architecture of the Faster R-CNN model consists of several key components, each performing a distinct task in the process of image processing and object detection.

1. Backbone (Feature Extractor). The Faster R-CNN architecture commences with the utilization of a convolutional neural network, which serves to extract features from the input

image. This network takes an image as input and generates a spatial feature map, which is a compact representation of textures, contours, and visual patterns. The feature map serves as the foundation for subsequent processing.

2. Region Proposal Network (RPN). The RPN is a central component of Faster R-CNN, which provides efficient generation of Region Proposals (Regions of Interest). This module takes as input the feature map obtained from the feature extractor. In the operation of this component, a convolutional grid is used to generate anchors, assess each anchor for the probability of belonging to an object, and regress anchor boundaries for more precise border determination. The output of this layer is a list of Region Proposals, sorted by probability.
3. ROI Pooling Operation. This operation allows for the standardization of the dimensions of input regions, irrespective of their scale. This layer is designed to consolidate information from regions of varying sizes into a uniform shape that can be processed by subsequent network layers.
4. Classifier and Bounding Box Regressor. Following ROI Pooling, the obtained features are passed to two separate modules: a classifier, which determines the class to which the object in each region belongs, and a regressor, which precisely adjusts the object boundaries (bounding box) to minimize positioning errors.

Based on the conducted analysis, the YOLOv11 neural network was selected for classification at subsequent levels.

YOLOv11 represents an evolution of previous YOLO versions, incorporating new features and enhancements to further improve performance and flexibility while delivering impressive speed and high efficiency. YOLOv11 offers five models of varying sizes: nano, small, medium, large, and extra-large. To balance detection accuracy and processing speed, we utilize the YOLOv11m and YOLOv11s models.

At the second classification level, the general class "Military Equipment" is subdivided into the subclasses "Tank," "Infantry Fighting Vehicle (IFV)," and "Other Equipment." This categorization is driven by the fact that, in modern military operations, these two types of vehicles are the most frequently used. Consequently, datasets contain significantly more images of these vehicle types compared to other military equipment. Since training a neural network on balanced class distributions enhances accuracy and improves feature recognition, the dataset was structured according to these defined categories. Moreover, these classes share common features that distinguish them significantly from other military equipment types. Due to these similarities, objects are often misclassified and assigned to incorrect categories. Therefore, to ensure high classification accuracy, it was necessary to select a neural network model capable of differentiating these shared features between the specified classes.

For classification at this level, we selected the YOLOv11m CNN model. The advantages of this network include its accuracy when working with medium and large objects. Additionally, YOLOv11 models are highly suitable for real-time recognition tasks due to their speed and flexibility in configuration.

The architecture of YOLOv11 is similar to that of previous versions and consists of the following components:

1. Backbone – This layer is responsible for feature extraction from the input image. YOLOv11 utilizes CSPDarknet53 as its primary network, ensuring efficient and deep feature extraction. It employs C2f (Cross Stage Partial Fusion) modules, which enables deeper learning with fewer layers compared to C3.
2. Neck – This component facilitates multi-scale feature fusion using a combination of Feature Pyramid Network (FPN) and Path Aggregation Network (PAN). This allows the preservation of information about objects of varying scales.
3. Head – This layer performs the final predictions, including object localization, classification, and confidence score estimation.

At the third classification stage, the "Other Equipment" class is further refined. At this level, objects are categorized into "Truck," "MLRS" (Multiple Launch Rocket System), and "Other Equipment." All detected objects that were not classified at the current or previous recognition levels are assigned to the "Other Equipment" category. The selected classes for this stage were chosen to ensure a balanced number of images within the datasets.

Since the third-level classes differ more significantly from one another than those at the second level, their distinguishing features are easier to extract and separate. As a result, a less powerful model can be used for classification at this level. However, it is reasonable to compensate for the reduced computational power by increasing processing speed. To achieve this, a smaller YOLO model, specifically YOLOv11s, can be applied for classification.

Figure 2 illustrates the data processing sequence of suggested approach.

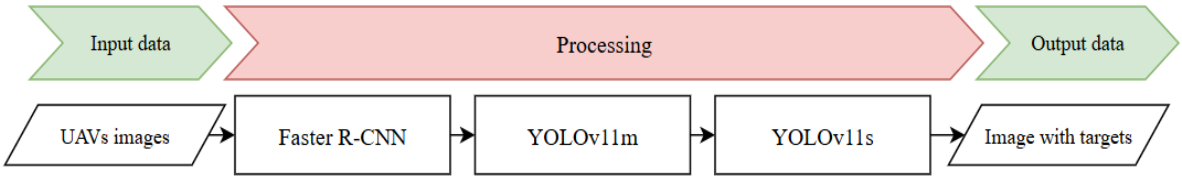

Figure 2: Data processing sequence.

To evaluate the performance of the proposed system, it is considered necessary to calculate overall statistical metrics such as Precision, Recall, and F1-score. The following formula was used to determine the values of these metrics (M) f1-score, recall and precision for the entire developed system:

$$M_{avg} = (M_1 + M_2 + M_3)/3, \quad (1)$$

where

M1 – metric value at the first classification level;

M2 – metric value at the second classification level;

M3 – metric value at the third classification level.

4. Results and discussion

4.1. Dataset

To train each of the sequential classification layers, it was necessary to synthetically generate training datasets at each level by dividing the total dataset into subclasses, with usage of method of creation of custom datasets [23], that are used at a particular recognition level. The FECL dataset [24] was used as a basis.

For training and experiments at each level, the prepared datasets were divided three times in the ratio of 80% for the training sample and 20% for the test sample.

4.2. Experiment results

To assess the efficiency of the proposed method, experiments were conducted and their results were compared. As part of the study, a sequential classification system was implemented for object recognition in images and videos obtained from UAVs. The main goal of the experiments was to assess the quality of the proposed system using the Precision, Recall, F1-score metrics. The results of the developed system were also compared with existing approaches to solving the problem under study.

Initially, each of the sequential classification models was trained on different data sets described above, corresponding to the level at which they are used.

At the first level of classification, the division occurs into 2 classes: “People” and “Military equipment”. In figure 3, the result of the first level of sequential recognition can be seen. In figure 3, the “People” class corresponds to the “Person” class, and military equipment corresponds to the “Vehicle” class.

Figure 3: Results of the first level of classification.

Tables 1 and 2 show the values of statistical metrics for validating the accuracy of the first classification class.

The tables show the results of the metrics for three experiments with different dataset divisions, and the average values and standard deviation are calculated.

Table 1
Calculation of the values of the Precision and Recall metrics for the first level

	Precision					Recall				
	1	2	3	Avg	Std	1	2	3	Avg	Std
Train dataset	94,4%	93,7%	94,6%	94,23%	0,9%	91,2%	91,9%	90,8%	91,3%	1,1%
Test dataset	91,7%	91,0%	90,5%	91,06%	1,2%	90.0%	90,4%	89,7%	90%	0,7%

Table 2
Calculation of F1-score metric values for the first level

	F1-score				
	1	2	3	Avg	Std
Train dataset	92,8%	93,1%	92,6%	92,83%	0,5%
Test dataset	91,1%	91,8%	90,8%	91,06%	1%

The results obtained during the validation of the first classification level show that the best metric values on the training dataset are Precision – 94,6%, Recall – 91,9%, and F1-score – 93,1%. On the test dataset, the values are Precision – 91,7%, Recall – 90,4%, and F1-score – 91,8%. The metrics slightly decreased on the test dataset. This is due to the fact that the neural network works with images on the test dataset that it has not previously processed. The standard deviation for all metrics was less than 5%.

At the second level of classification, the class “Military equipment” is refined using the YOLOv11m neural network model. In figure 4, you can see the results of the second level of recognition. In figure 4, the class “Tank” corresponds to the class “Tank”, and other military equipment corresponds to the class “Vehicle”.

Figure 4: Results of the second level of classification.

Tables 3 and 4 show the values of statistical metrics for validating the accuracy of the first classification class.

Table 3
Calculation of the values of the Precision and Recall metrics for the second level

	Precision					Recall				
	1	2	3	Avg	Std	1	2	3	Avg	Std
Train dataset	90,4%	89,9%	90,7%	90,33%	0,8%	90,7%	88,1%	89,7%	89,5%	2,6%
Test dataset	88,7%	88,3%	88,9%	88,63%	0,6%	89,2%	89,0%	88,7%	88,96%	0,5%

Table 4
Calculation of F1-score metric values for the second level

	F1-score				
	1	2	3	Avg	Std
Train dataset	91,0%	91,8%	90,4%	91,06%	1,4%
Test dataset	89,7%	90,0%	89,8%	89,83%	0,3%

The results obtained during the validation of the second classification level show that the best metric values on the training dataset are Precision – 90,4%, Recall – 90,4%, and F1-score – 90,4%. On the test dataset, the values are Precision – 88,9%, Recall – 89,2%, and F1-score – 90,0%. The metrics slightly decreased on the test dataset. This is due to the fact that the neural network works with images on the test dataset that it has not previously processed. The standard deviation for all metrics was less than 5%.

At the third level of classification, the “Other equipment” class is refined using the YOLOv11s neural network model. In figure 5, you can see the results of the third level of recognition, which is

the final result of the entire classification sequence. In figure 5, the “Truck” class corresponds to the “Car” class, and other military equipment corresponds to the “Vehicle” class.

Figure 4: Results of the third level of classification.

Tables 5 and 6 show the values of statistical metrics for validating the accuracy of the first classification class.

Table 5
Calculation of the values of the Precision and Recall metrics for the third level

	Precision					Recall				
	1	2	3	Avg	Std	1	2	3	Avg	Std
Train dataset	93,7%	93,2%	93,3%	93,4%	0,5%	94,0%	93,6%	93,3%	93,63%	0,7%
Test dataset	90,7%	91,1%	90,4%	90,73%	0,7%	92,8%	92,6%	91,8%	92,4%	1%

Table 6
Calculation of F1-score metric values for the third level

	F1-score				
	1	2	3	Avg	Std
Train dataset	94,1%	94,5%	93,4%	94%	1,1%
Test dataset	92,9%	92,7%	91,8%	92,46%	1,1%

The results obtained during the validation of the third classification level show that the best metric values on the training dataset are Precision – 93,7%, Recall – 94,0%, and F1-score – 94,5%. On the test dataset, the values are Precision – 91,1%, Recall – 92,8%, and F1-score – 92,9%. The metrics slightly decreased on the test dataset. This is due to the fact that the neural network works with images on the test dataset that it has not previously processed. The standard deviation for all metrics was less than 5%.

In addition, ROC curves were built and AUC analysis was performed to validate the learning quality of the models of each classification level. The results can be seen in Figure 6.

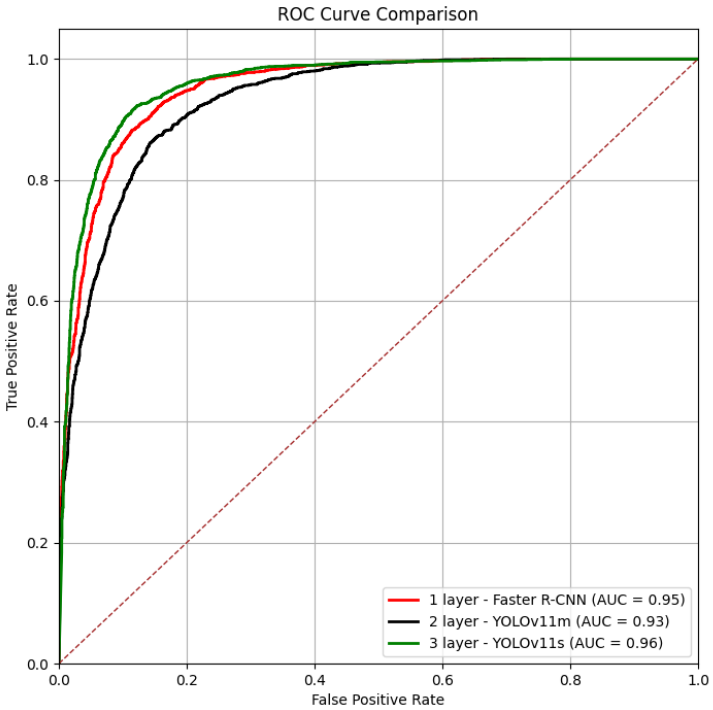

Figure 6: ROC curves and AUC analysis results.

Using formula (1), the overall metric values were calculated and compared with other existing approaches. Table 7 shows a comparison of the results of the proposed method with others.

Table 7
Comparison of the results of the existing method with others

Method	Precision	Recall	F1-score
Suggested method	**91,4%**	**91,29%**	**91,87%**
Existing method 1 [25]	**91,4%**	91,0%	91,4%
Existing method 2 [26]	90,51%	90,9%	90,6%

The suggested sequential classification approach outperformed some existing approaches and showed the best results in terms of Recall and F1-score metrics. Thus, for complex scenarios with a large number of classes, the suggested sequential approach demonstrates better accuracy.

The system implementing the suggested approach was trained on aerial images obtained from a UAV camera and shows its best results when working with them.

The limitations of the proposed approach include its application to image processing with poor resolution, which may lead to incorrect classification or failure to detect objects, as well as a limited number of classification classes at the current level of the sequential structure.

Further research may be aimed at adding new levels to the sequential classification, which will allow recognizing more types of military equipment and improving classification accuracy when working with low-resolution images.

5. Conclusions

This study developed and analyzed a system for automatic detection and recognition of military objects based on aerial and video materials obtained using UAVs. The suggested system is based on a multi-level architecture that includes modern deep learning algorithms, in particular YOLOv11 and Faster R-CNN, which allows achieving high classification accuracy. Three levels of classification were implemented: initial object recognition (person, military equipment), detailed classification of

equipment by category (tanks, infantry fighting vehicles, and other) and additional analysis of the "other" category to clarify the types of objects (MLRS, trucks, etc.).

The conducted experimental study included the analysis of 5000 images and 20 hours of video materials. The analysis of these materials allowed us to assess the effectiveness of the proposed system in conditions close to real use. The obtained results demonstrate a high level of accuracy (Precision, Recall and F1-score exceed 91% at all classification levels), which indicates the effectiveness of the approach. In addition to high accuracy, the system provides fast processing of input data, which allows its use in real time.

Additional comparative analysis with modern methods confirmed the competitiveness of the developed solution, especially in terms of recognition accuracy and data processing speed. This indicates the prospects for further implementation of the system for solving applied military tasks.

The limitations of the proposed approach include its application to image processing with poor resolution, which may lead to incorrect classification or failure to detect objects, as well as a limited number of classification classes at the current level of the sequential structure.

Further research may be aimed at expanding the capabilities of the system, including integration with other deep learning technologies, increasing resilience to changing shooting conditions, and adaptation to new types of military equipment.

Declaration on Generative AI

During the preparation of this work, the authors used GPT-4o and Grammarly in order to: Grammar and spelling check. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] H. Liu, Y. Yu, S. Liu, W. Wang, A Military Object Detection Model of UAV Reconnaissance Image and Feature Visualization, Appl. Sci. 12.23 (2022) 12236. doi:10.3390/app122312236.

[2] G. Tang, J. Ni, Y. Zhao, Y. Gu, W. Cao, A Survey of Object Detection for UAVs Based on Deep Learning, Remote Sens. 16.1 (2023) 149. doi:10.3390/rs16010149.

[3] S. Liu, H. He, Z. Zhang, Y. Zhou, LI-YOLO: An Object Detection Algorithm for UAV Aerial Images in Low-Illumination Scenes, Drones 8.11 (2024) 653. doi:10.3390/drones8110653.

[4] X. Zhao, Y. Chen, YOLO-DroneMS: Multi-Scale Object Detection Network for Unmanned Aerial Vehicle (UAV) Images, Drones 8.11 (2024) 609. doi:10.3390/drones8110609.

[5] D. Yan, G. Li, X. Li, H. Zhang, H. Lei, K. Lu, M. Cheng, F. Zhu, An Improved Faster R-CNN Method to Detect Tailings Ponds from High-Resolution Remote Sensing Images, Remote Sens. 13.11 (2021) 2052. doi:10.3390/rs13112052.

[6] HR-YOLOv8: A Crop Growth Status Object Detection Method Based on YOLOv8 / J. Zhang et al. Electronics. 2024. Vol. 13, no. 9. P. 1620.

[7] Implementation of a Modified Faster R-CNN for Target Detection Technology of Coastal Defense Radar / H. Yan et al. Remote Sensing. 2021. Vol. 13, no. 9. P. 1703. URL: https://doi.org/10.3390/rs13091703.

[8] H. Zhang, W. Sun, C. Sun, R. He, Y. Zhang, HSP-YOLOv8: UAV aerial photography small target detection algorithm, Drones 8.9 (2024) 453. doi:10.3390/drones8090453.

[9] Z. Bai, X. Pei, Z. Qiao, G. Wu, Y. Bai, Improved yolov7 target detection algorithm based on UAV aerial photography, Drones 8.3 (2024) 104. doi:10.3390/drones8030104.

[10] J. Guo, X. Liu, L. Bi, H. Liu, H. Lou, UN-YOLOv5s: A uav-based aerial photography detection algorithm, Sensors 23.13 (2023) 5907. doi:10.3390/s23135907.

[11] X. Wei, L. Yin, L. Zhang, F. Wu, DV-DETR: improved UAV aerial small target detection algorithm based on RT-DETR, Sensors 24.22 (2024) 7376. doi:10.3390/s24227376.

[12] X. Luo, Y. Wu, F. Wang, Target detection method of UAV aerial imagery based on improved yolov5, Remote Sens. 14.19 (2022) 5063. doi:10.3390/rs14195063.

[13] HR-YOLOv8: A Crop Growth Status Object Detection Method Based on YOLOv8 / J. Zhang et al. Electronics. 2024. Vol. 13, no. 9. P. 1620. URL: https://doi.org/10.3390/electronics13091620.
[14] Implementation of a Modified Faster R-CNN for Target Detection Technology of Coastal Defense Radar / H. Yan et al. Remote Sensing. 2021. Vol. 13, no. 9. P. 1703.
[15] M. H. Rahman, M. A. S. Sejan, M. A. Aziz, R. Tabassum, J.-I. Baik, H.-K. Song, A comprehensive survey of unmanned aerial vehicles detection and classification using machine learning approach: challenges, solutions, and future directions, Remote Sens. 16.5 (2024) 879. doi:10.3390/rs16050879.
[16] Y. Mo, J. Huang, G. Qian, Deep learning approach to UAV detection and classification by using compressively sensed RF signal, Sensors 22.8 (2022) 3072. doi:10.3390/s22083072.
[17] T. Lu, L. Wan, S. Qi, M. Gao, Land cover classification of UAV remote sensing based on transformer–cnn hybrid architecture, Sensors 23.11 (2023) 5288. doi:10.3390/s23115288.
[18] H. S. Munawar, F. Ullah, S. Qayyum, A. Heravi, Application of deep learning on uav-based aerial images for flood detection, Smart Cities 4.3 (2021) 1220–1243. doi:10.3390/smartcities4030065.
[19] I. Teixeira, R. Morais, J. J. Sousa, A. Cunha, Deep learning models for the classification of crops in aerial imagery: A review, Agriculture 13.5 (2023) 965. doi:10.3390/agriculture13050965.
[20] K. Kim, D. Lee, Y. Jang, J. Lee, C.-H. Kim, H.-T. Jou, J.-H. Ryu, Deep learning of high-resolution unmanned aerial vehicle imagery for classifying halophyte species: A comparative study for small patches and mixed vegetation, Remote Sens. 15.11 (2023) 2723. doi:10.3390/rs15112723.
[21] A. Munir, A. J. Siddiqui, S. Anwar, A. El-Maleh, A. H. Khan, A. Rehman, Impact of adverse weather and image distortions on vision-based UAV detection: A performance evaluation of deep learning models, Drones 8.11 (2024) 638. doi:10.3390/drones8110638.
[22] Z. Liu, P. An, Y. Yang, S. Qiu, Q. Liu, X. Xu, Vision-Based drone detection in complex environments: A survey, Drones 8.11 (2024) 643. doi:10.3390/drones8110643.
[23] T. ISAIEV, T. KYSIL, METHOD OF CREATING CUSTOM DATASET TO TRAIN CONVOLUTIONAL NEURAL NETWORK, Comput. Syst. Inf. Technol. No. 4 (2024) 37–44. doi:10.31891/csit-2024-4-5.
[24] Military Object Detection Dataset by militarypy. URL: https://universe.roboflow.com/militarypy/military-epgmt.
[25] X. Zhao, W. Zhang, Y. Xia, H. Zhang, C. Zheng, J. Ma, Z. Zhang, G-YOLO: A lightweight infrared aerial remote sensing target detection model for uavs based on yolov8, Drones 8.9 (2024) 495. doi:10.3390/drones8090495.
[26] X. Du, L. Song, Y. Lv, S. Qiu, A lightweight military target detection algorithm based on improved yolov5, Electronics 11.20 (2022) 3263. doi:10.3390/electronics11203263.