UDC 681.327.12(045)
DOI:10.18372/1990-5548.86.20560

A. T. Kot

COMPUTER VISION FOR UAV-BASED RECONNAISSANCE UNDER CONDITIONS OF MODERN WARFARE

Educational and Research Institute for Applied System Analysis, National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute" Kyiv, Ukraine
E-mail: anatoly.kot@gmail.com, ORCID 0000-0002-7490-8834

Abstract*—This paper considers the application of computer vision and deep learning methods for automated aerial reconnaissance using unmanned aerial vehicles under the conditions of modern warfare. The main classes of reconnaissance objects are analyzed, including military vehicles, fortifications, artillery positions, and groups of personnel. An approach to building an object detection system based on deep neural networks is proposed, in particular using YOLO-type detectors and U-Net segmentation models. The process of data preparation and augmentation with consideration of combat factors (smoke, explosions, low illumination, image shift, and noise) is described. An experimental evaluation of object detection quality under different scenarios is performed. It is shown that the use of specially adapted augmentation significantly increases the robustness of the models to interference. The limitations of the proposed approach and directions for further research are discussed.*

Keywords—Unmanned aerial vehicle; computer vision; deep learning; object detection; military reconnaissance; neural networks.

I. INTRODUCTION

Modern warfare is characterized by high dynamics, intensive use of technical means, and a significant volume of information received in real time. One of the key sources of operational information has become unmanned aerial vehicles (UAVs), which provide surveillance, reconnaissance, fire adjustment, and assessment of strike results. The volume of video data transmitted from UAVs is continuously increasing, which leads to operator overload and increases the risk of missing important targets.

The traditional working model, in which visual observation by a human remains the main means of analysis, has a number of significant limitations. These include operator fatigue, subjectivity of assessment, limited ability to simultaneously analyze multiple video streams, and deterioration of decision-making quality under stressful conditions.

The development of convolutional neural networks and deep learning methods has created the prerequisites for the wide implementation of automated image analysis systems in real time [12]. Single-stage detectors of the YOLO family have demonstrated the ability to combine high accuracy and high processing speed, which is critically important for application on UAVs [1], [2].

The aim of this work is to study the possibilities of applying computer vision methods for automated reconnaissance using UAVs under the conditions of modern warfare, as well as to develop and experimentally evaluate an approach to the detection of military objects in aerial images.

The following research questions are addressed in this study:

- Is it possible to ensure robust object detection under complex combat imaging conditions?
- How do flight altitude and illumination affect detection accuracy?

What trade-off between accuracy and processing speed is acceptable for practical use on UAVs?

II. REVIEW OF EXISTING SOLUTIONS

The first approaches to automated analysis of aerial images were based on classical image processing methods. These included threshold filtering, contour analysis, feature extraction methods such as HOG and SIFT, and subsequent classification using machine learning techniques, in particular SVM and *k*-NN. Such methods demonstrated limited effectiveness under complex conditions and with high variability of images.

With the emergence of deep neural networks, the situation has changed significantly. Modern object detectors are divided into two main groups: two-stage detectors (Faster R-CNN) and single-stage detectors (YOLO, SSD) [1], [2], [4]. Two-stage methods provide higher accuracy but have significantly lower processing speed. Single-stage detectors are oriented toward operation in near real-time mode, which makes them more suitable for use on UAVs.

A separate class of tasks is image segmentation, which makes it possible not only to detect an object but also to determine its exact shape and position in the frame. The most widespread architectures are U-Net, DeepLab, and their modifications [3]. Segmentation is especially important for the detection of engineering structures, trenches, and fortifications that have complex geometric shapes.

Most existing studies in the field of aerial reconnaissance are based on open civilian datasets such as DOTA and xView [6], [7]. However, these datasets usually do not take into account specific combat factors (smoke, explosions, camouflage, night imaging, and camera instability). This creates a gap between laboratory results and the real effectiveness of systems under combat conditions.

III. PROBLEM STATEMENT

The task of automated reconnaissance is reduced to the problem of object detection and, when necessary, object segmentation in a sequence of frames obtained from a UAV camera.

An image or a video frame of size $H \times W \times C$ is provided as the input to the system.

At the output, a set of objects is formed, each of which is characterized by the coordinates of a bounding box, the object class, and a confidence score.

Mathematically, the problem is formulated as the search for the parameters of a model f_{\theta} that approximates the conditional distribution $P(y|x)$.

Within this work, the following basic object classes were considered: armored vehicles (tanks, IFVs, APCs), artillery positions, cargo and passenger vehicles, buildings and fortifications, trenches and engineering structures, and groups of personnel.

For training and testing the models, a combined dataset was used, which consisted of anonymized real aerial videos, synthetic images generated in the AirSim simulation environment [8], and open civilian datasets such as DOTA and xView [6], [7]. The distribution of objects by classes in the training dataset is presented in Table I.

TABLE I. DISTRIBUTION OF OBJECTS BY CLASSES IN THE TRAINING DATASET

Object Class	Number of Objects	Share, %
Tanks	840	13.4
IFVs / APCs	620	9.9
Trucks	910	14.5
Passenger Cars	760	12.1
Artillery Positions	430	6.8
Buildings and Fortifications	1200	19.1
Trenches	980	15.6
Groups of Personnel	550	8.7
Total	6290	100

All images were converted to a unified format and divided into training, validation, and test sets. The annotation was performed in the YOLO format.

To increase the robustness of the models, specialized data augmentation was applied, including the addition of smoke and fog, simulation of explosion flashes, motion blur, changes in brightness and contrast, addition of sensor noise, and minor geometric distortions [11].

Visual examples of the application of specialized combat-oriented augmentation are shown in Fig. 1.

IV. APPROACH

The core of the system is a single-stage object detector adapted for processing aerial images based on the YOLO family [1], [2]. The architecture is configured for the detection of small objects typical for high-altitude flights. The input image size was modified, anchor boxes were optimized, and multi-scale training was applied. The overall architecture of the automated UAV-based reconnaissance system is shown in Fig. 2.

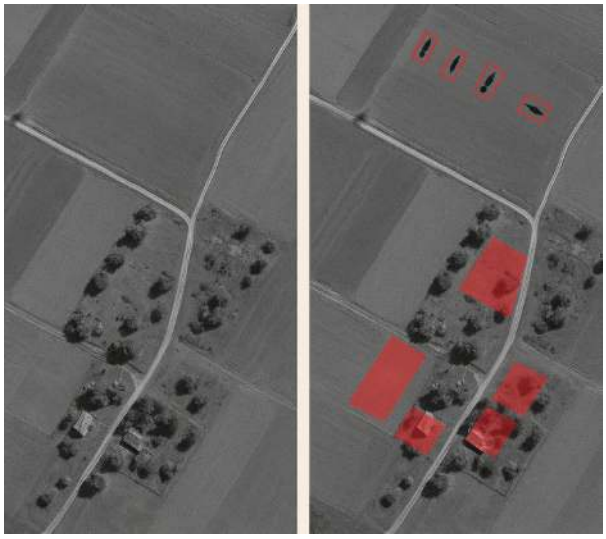

Fig. 1. Example of Combat-Oriented Image Augmentation

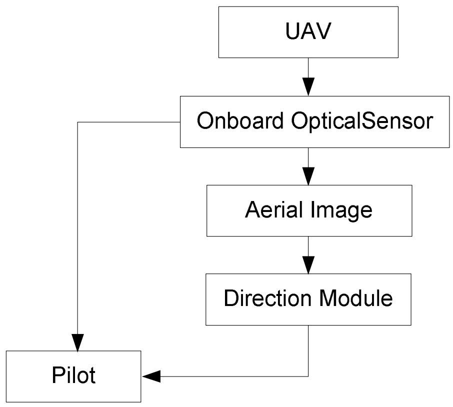

Fig. 2. Overall Architecture of the UAV-based Reconnaissance System

For the task of segmentation of engineering structures, a separate neural network of the U-Net type was used [3]. This made it possible to obtain masks of trenches and fortifications that have complex shapes and weak visual expression.

The processing pipeline includes the following stages:

- acquisition of the video stream from the UAV;
- selection of key frames;
- image pre-processing;
- object detection;
- confidence-based filtering;
- aggregation of results within a temporal window;
- transmission of the results to the operator.

The implementation was carried out using the PyTorch framework. Embedded systems optimized for neural network execution in the edge-inference mode were considered as the hardware platform, using deep model optimization and computational compression techniques [9], [10].

V. RESULTS OF MODELING

All experimental results in this work were obtained using synthetic data generated in a simulation environment and are used to demonstrate the feasibility of the proposed approach.

The experiments were conducted under four scenarios: daytime imaging at low altitude, daytime imaging at high altitude, imaging under low-light conditions, and imaging in the presence of smoke and explosions in the frame.

The evaluation was carried out using the precision, recall, F1-score, and mAP@0.5 metrics. The FPS metric was used to assess the processing speed.

Model optimization for execution on embedded platforms was performed taking into account modern approaches to edge inference and neural network acceleration [9], [10].

A quantitative comparison of detection accuracy under different imaging conditions is given in Table II.

TABLE II. COMPARISON OF DETECTION ACCURACY (MAP@0.5) UNDER DIFFERENT IMAGING CONDITIONS

Imaging Conditions	Baseline Model	Model with Combat Augmentation
Daytime, low altitude	0.68	0.78
Daytime, high altitude	0.59	0.7
Low illumination	0.52	0.63
Smoke / heavy smoke	0.47	0.59

As shown in Fig. 3, the use of combat-oriented augmentation provides a significant increase in detection accuracy under all imaging conditions.

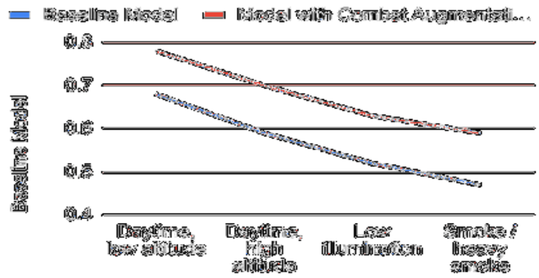

Fig. 3. Detection Accuracy (mAP@0.5) under Different Imaging Conditions

The comparison of accuracy and processing speed for models of different sizes is presented in Table III.

TABLE III. ACCURACY (MAP@0.5) AND FPS FOR MODELS OF DIFFERENT SIZES

Model Size	mAP@0.5	FPS
Small	0.71	42
Medium	0.76	27
Large	0.79	14

The results showed that the use of specialized combat-oriented augmentation increases the mAP under difficult conditions by 10–18% compared to the baseline model without special preparation [11].

The largest decrease in accuracy was observed during imaging at high altitude and under heavy smoke conditions. A typical example of the results of military object detection in an aerial image is shown in Fig. 4.

Fig. 4. Example of Military Object Detection Results in an Aerial Image

The trade-off between detection accuracy and processing speed for models of different sizes is illustrated in Fig. 5.

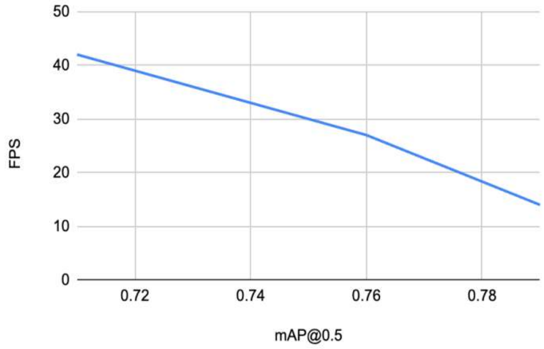

Fig. 5. Trade-off between Accuracy (mAP@0.5) and Processing Speed (FPS) for Models of Different Sizes

Typical classification errors of objects from different classes are presented in the form of a confusion matrix in Fig. 6.

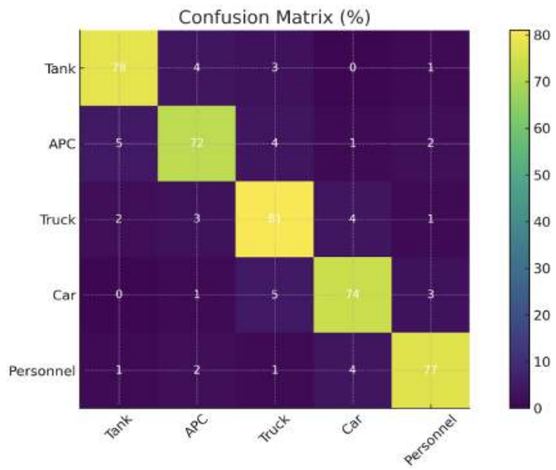

Fig. 6. Confusion Matrix of Military Object Classification

VI. Conclusions

This study investigates the possibilities of applying computer vision methods for automated reconnaissance using UAVs under the conditions of modern warfare. An approach to the development of a system for the detection and segmentation of military objects with consideration of real combat factors is proposed. The experimental results demonstrate the potential of deep neural networks to reduce operator workload and increase the effectiveness of reconnaissance. The obtained results can be used in the development of practical decision support systems for UAVs.

The obtained results indicate that modern computer vision methods are capable of effectively supporting military reconnaissance tasks. Large objects with clear geometric features are detected most reliably. The most challenging tasks remain the detection of small groups of personnel and objects that are masked against the background of the terrain.

Smoke, explosions, and low illumination have a strong impact on detection quality. Therefore, adaptation of the models to real combat conditions is critically important. The balance between accuracy and processing speed is also essential, since excessively complex models are not suitable for real-time operation.

The main limitations of the proposed approach include dependence on the quality of training data, computational resources, and the complexity of scaling the system to a large number of UAVs. Issues of safety, ethics, and responsibility in the use of artificial intelligence in military systems also require separate consideration.

Further research should be focused on:

- combining detection with multi-frame tracking;
- the use of thermal imaging cameras;
- integration of the results into decision support systems;
- application of multi-agent analysis using data from multiple UAVs.

References

[1] J. Redmon and A. Farhadi, “YOLOv3: An Incremental Improvement,” arXiv preprint arXiv:1804.02767, 2018. https://arxiv.org/abs/1804.02767

[2] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, “YOLOv4: Optimal Speed and Accuracy of Object Detection,” arXiv preprint arXiv:2004.10934, 2020. https://arxiv.org/abs/2004.10934

[3] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” Proc. MICCAI, 2015. https://arxiv.org/abs/1505.04597 https://doi.org/10.1007/978-3-319-24574-4_28

[4] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal Loss for Dense Object Detection,” *IEEE International Conference on Computer Vision (ICCV)*, 2017. https://arxiv.org/abs/1708.02002 https://doi.org/10.1109/ICCV.2017.324

[5] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, “The PASCAL Visual Object Classes (VOC) Challenge,” *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, 2010. https://link.springer.com/article/10.1007/s11263-009-0275-4

[6] G.-S. Xia, J. Hu, F. Hu, X. Shi, X. Bai, Y. Zhong, L. Zhang, and X. Lu, “DOTA: A Large-Scale Dataset for Object Detection in Aerial Images,” *Proc. IEEE Conference on Computer Vision and Pattern*

Recognition (CVPR), 2018. https://arxiv.org/abs/1711.10398 https://doi.org/10.1109/CVPR.2018.00418

[7] D. Lam, et al., "xView: Objects in Context in Overhead Imagery," *Proc. IEEE CVPR*, 2018. https://arxiv.org/abs/1802.07856

[8] S. Shah, D. Dey, C. Lovett, and A. Kapoor, "AirSim: High-Fidelity Visual and Physical Simulation for Autonomous Vehicles," *Proc. IEEE CVPR*, 2017. https://arxiv.org/abs/1705.05065 https://doi.org/10.1007/978-3-319-67361-5_40

[9] A. G. Howard, et al., "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications," arXiv preprint arXiv:1704.04861, 2017. https://arxiv.org/abs/1704.04861

[10] A. Garcia, A. Gordon, J. Hurst, S. Konecny, S. Kim, and K. An, "Glider: A Deep Learning Framework for Edge Inference," arXiv preprint arXiv:2003.03119, 2020. https://arxiv.org/abs/2003.03119

[11] C. Shorten and T. M. Khoshgoftaar, "A Survey on Image Data Augmentation for Deep Learning," *Journal of Big Data*, vol. 6, no. https://journalofbigdata.springeropen.com/articles/10.1186/s40537-019-0197-0 60, 2019.

[12] I. Goodfellow, Y. Bengio, and A. Courville, "Deep Learning," *MIT Press, Cambridge*, 2016. https://www.deeplearningbook.org

Received Oktober 02, 2025

Kot Anatoly. ORCID 0000-0002-7490-8834. PhD Student.
Artificial Intelligence Department, Educational and Research Institute for Applied System Analysis, National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute,", Kyiv, Ukraine.
Education: National Technical University of Ukraine "Ihor Sikorsky Kyiv Polytechnic Institute", (2025).
Research area: artificial Intelligence.
Publications: 11.
E-mail: anatoly.kot@gmail.com

А. Т. Кот. Комп'ютерний зір для розвідки з БПЛА в умовах сучасної війни
У статті розглянуто застосування методів комп'ютерного зору та глибокого навчання для автоматизованої повітряної розвідки з використанням безпілотних літальних апаратів в умовах сучасної війни. Проаналізовано основні класи об'єктів розвідки, включаючи військову техніку, укріплення, артилерійські позиції та групи живої сили. Запропоновано підхід до побудови системи виявлення об'єктів на основі глибоких нейронних мереж, зокрема детекторів типу YOLO та сегментаційних моделей U-Net. Описано процес підготовки та аугментації даних з урахуванням бойових факторів (дим, вибухи, низька освітленість, зсув зображення та шум). Проведено експериментальну оцінку якості виявлення об'єктів у різних сценаріях. Показано, що використання спеціально адаптованої аугментації значно підвищує стійкість моделей до завад. Обговорено обмеження запропонованого підходу та напрями подальших досліджень.
Ключові слова: безпілотні літальні апарати; комп'ютерний зір; глибоке навчання; детекція об'єктів; військова розвідка; нейронні мережі.

Кот Анатолій Тарасович. ORCID 0000-0002-7490-8834. Аспірант.
Кафедра штучного інтелекту, Навчально-науковий інститут прикладного системного аналізу, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна.
Освіта: Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, Україна, (2025).
Напрям наукової діяльності: штучні нейронні мережі, програмування.
Кількість публікацій: 11.
E-mail: anatoly.kot@gmail.com