ITM Web of Conferences **72**, 03010 (2025) https://doi.org/10.1051/itmconf/20257203010
HMMOCS-III 2024

Autonomous on-board object and phenomenon detection system

Anton Mikhalev[1], *Nikolai* Guliutin[1], *Nadezhda* Ermienko[1], and *Oleslav* Antamoshkin[1,2*]

[1]Siberian Federal University, 79 Svobodny pr., Krasnoyarsk, 660041, Russian Federation
[2]Reshetnev Siberian State University of Science and Technology, 31, Krasnoyarsky Rabochy Ave, Krasnoyarsk, 660037, Russian Federation

Abstract. This paper presents the design, implementation, and evaluation of an autonomous on-board object and phenomenon detection system optimized for real-time performance and resource-constrained environments. The proposed framework integrates a multimodal sensor array, including RGB cameras and LiDAR, with lightweight deep learning algorithms for object detection, tracking, and classification. Four state-of-the-art detection models - YOLO, DETR, CenterNet, and M2Det - were examined using the Lacmus Drone Dataset, a publicly available collection of over 3,000 aerial images. Experimental results highlight that no single model consistently outperforms the others: YOLO excels in real-time scenarios due to its fast inference speed, whereas DETR achieves the highest accuracy at the expense of greater computational complexity. CenterNet offers a balanced approach for detecting smaller objects, and M2Det demonstrates strong performance in densely populated urban scenes. Overall, these findings emphasize the importance of selecting model architectures based on mission requirements and hardware constraints, paving the way for more efficient and adaptive autonomous detection systems.

1 Introduction

Autonomous systems capable of real-time object and phenomenon detection are increasingly critical in diverse domains, ranging from environmental monitoring to disaster management and smart transportation. These systems enable the identification and analysis of objects, events, and anomalies in dynamic environments, offering actionable insights for both automated responses and human decision-making. With advancements in computer vision, sensor technology, and machine learning, autonomous detection systems are becoming more efficient and adaptable. However, many existing solutions are limited by their reliance on centralized processing, high computational demands, and lack of scalability for onboard applications.

This study aims to design and implement an autonomous on-board object and phenomenon detection system that integrates advanced sensing technologies and lightweight machine learning algorithms. The system is optimized for real-time detection and decision-

* Corresponding author: oleslav@mail.ru

making in resource-constrained environments, such as unmanned aerial vehicles (UAVs), autonomous ground vehicles (AGVs), and remote sensing platforms. The primary objectives include achieving high detection accuracy, minimizing latency, and ensuring adaptability to various operational conditions.

Developing a robust on-board detection system addresses critical challenges in real-time monitoring, particularly in areas where rapid responses are essential, such as natural disaster assessment, agricultural management, and transportation safety. By enabling autonomous systems to detect and classify objects and phenomena locally, the proposed solution reduces reliance on cloud-based processing, enhances operational efficiency, and expands the applicability of detection technologies in resource-constrained scenarios. This research bridges the gap between theoretical advancements in machine learning and their practical deployment in autonomous systems, contributing to more resilient and adaptive monitoring solutions.

2 Literature review

The development of autonomous on-board object and phenomenon detection systems has garnered significant attention in recent years, driven by advancements in sensor technology, machine learning algorithms, and real-time processing capabilities. This section reviews 15 high-impact studies that have contributed to this field, focusing on their methodologies, findings, and implications.

2.1 Existing systems

Several studies have explored the integration of various sensors and algorithms to enhance on-board detection capabilities. Xu et al. [1] proposed a lightweight 3D dynamic obstacle detection and tracking method using an RGB-D camera, designed for low-power robots with limited computing power. Their system demonstrated real-time high-accuracy obstacle detection suitable for small autonomous robots.

Similarly, Zhang et al. [2] developed a robust approach for detecting and localizing flying objects using 3D LiDAR sensors on autonomous aerial vehicles. This method enabled highly dynamic aerial interception and agile multi-robot interaction.

In the realm of Earth observation, a deep-learning-based system for on-board change detection was developed to run on existing flight-proven hardware [3]. This system utilized convolutional neural networks to automatically detect changes in Earth's surface, enhancing real-time monitoring capabilities.

2.2 Challenges in on-board applications

Implementing detection systems on-board autonomous platforms presents significant challenges, including limited computational resources, power constraints, and the necessity for real-time performance. Many studies have focused on strategies to address these issues. For instance, Ma et al. [4] proposed an object detection system based on edge-cloud cooperation, termed Edge YOLO, effectively reducing dependence on cloud computing resources and enabling real-time intelligent object detection through reconstructive convolutional neural networks. Additionally, Yao et al. [5] developed a resource-efficient online target detection system with autonomous capabilities, demonstrating robust performance in outdoor high-altitude experiments.

Recent works also highlight the importance of adapting detection frameworks to specific domains. Pyataeva et al. [6] demonstrated how data from UAVs could be used to classify tree species in the Kuznetsovskoe Plateau region. Their approach underscores the computational and methodological complexities inherent in remote sensing applications, where rapid, on-board processing of high-resolution images is critical for tasks like forest inventory and monitoring. Similarly, Antamoshkin et al. [7] explored on-board detection and classification in environmental and Earth observation contexts, further emphasizing the trade-off between advanced analytics requirements and on-board processing limitations.

Taken together, these studies showcase the ongoing efforts to overcome the technical and operational challenges posed by on-board detection applications. Addressing power efficiency, algorithmic complexity, and heterogeneous sensor data processing remains paramount for advancing autonomous systems capable of effective real-time monitoring, whether in environmental management, disaster response, or a variety of other critical domains.

2.3 Technological advances

Advancements in sensor fusion and machine learning have significantly enhanced on-board detection systems. A multimodal object detection and ranging system based on camera and LiDAR sensor fusion was introduced for autonomous driving applications [8], improving object detection, classification, and ranging under challenging circumstances. Furthermore, Zhao et al. [9] proposed an on-board real-time object detection system for UAVs equipped with embedded neural processing units (NPUs). By designing a deep-learning network structure based on YOLOv3-Tiny, the system achieved efficient on-board object detection suitable for real-time scenario analysis.

In addition, the integration of YOLO-based algorithms to augment unmanned aerial vehicle (UAV) capabilities across diverse industrial applications was explored. It was demonstrated how efficient detection models can be adapted for pipeline inspection, security monitoring, and agricultural tasks, thereby extending the operational scope of UAVs. This work underscores the importance of lightweight yet robust algorithms that can handle various targets, environmental conditions, and mission objectives.

Finally, Lin et al. [10] developed a compact, fully autonomous quadrotor system capable of flying in cluttered environments while avoiding small dynamic obstacles using forward-looking 3D LiDAR. These and other advancements continue to drive the evolution of on-board detection, emphasizing the need for efficient hardware-software integration, high-speed inference, and robust handling of diverse datasets.

3 Methodology

3.1 System design

The proposed autonomous on-board object and phenomenon detection system integrates advanced sensors and real-time processing algorithms within a compact hardware platform. The system design includes three key components: Sensor Array: A combination of RGB cameras, LiDAR, and thermal sensors ensures comprehensive data acquisition across diverse environmental conditions. Processing Unit: A low-power embedded system equipped with a neural processing unit (NPU) accelerates deep learning inference while maintaining energy efficiency. Control and Communication: A microcontroller unit (MCU) interfaces with the sensors and processing unit, enabling seamless communication with external systems for data sharing and decision-making.

The architecture emphasizes modularity, allowing the system to adapt to different platforms such as UAVs, ground robots, or stationary monitoring units.

3.2 Data acquisition and preprocessing

Data acquisition is carried out using synchronized sensors that capture multimodal inputs, including images, point clouds, and temperature variations. The preprocessing pipeline involves sensor calibration, which aligns spatial and temporal data across different modalities to ensure consistency; noise reduction, where Gaussian filters are applied for image denoising and voxel grid filters for LiDAR data compression; and data normalization, which standardizes sensor outputs to maintain compatibility with the detection algorithms.

3.3 Detection algorithms

The system employs state-of-the-art detection algorithms tailored for onboard deployment:

- YOLO or Transformer. Optimized for real-time object detection, this lightweight deep learning model balances accuracy and processing speed.
- PointNet. For processing LiDAR point clouds, enabling 3D object localization and classification.
- Tracking. The system integrates advanced tracking algorithms such as SORT or ByteTrack for effective object tracking. SORT is a lightweight and efficient tracking algorithm that uses Kalman filters for motion prediction and Hungarian matching for associating detections across frames. Its simplicity ensures low latency, making it suitable for real-time applications. ByteTrack improves the tracking robustness in complex scenarios, such as crowded environments or under occlusions.

4 Data description and research methodology

In this study, our primary objective was to detect missing persons in images captured by a drone-mounted camera. For the development and evaluation of object detection models, we utilized the publicly available "Lacmus Drone Dataset," which contains over 3,000 images. The dataset was primarily collected using DJI Mavic Pro and Phantom drones at altitudes of 50–100 meters, with each image having a resolution of 3000×4000 pixels. On average, an individual occupies an area of approximately 50×100 pixels in these images. An example image from the dataset with annotated persons is shown in Figure 1.

Fig. 1. Example image from the dataset with highlighted persons.

In Figure 2, a virtual forest environment is depicted, featuring dense coniferous trees and a patch of ground covered in simple green terrain. Near the center, there is a plume of smoke rising vertically, suggesting an active fire. A bounding box with a blue outline surrounds the flames at the base of the smoke column, indicating that a YOLO v5 model has successfully identified the fire. The confidence score "0.68" shown above the bounding box reflects the model's level of certainty in detecting the fire event.

Fig. 2. Fire Detection in a Virtual Forest Environment using YOLO v5.

This example illustrates how a real-time object detection algorithm can be employed for fire detection in a simulation or training environment. By visualizing the flames in a 3D setting and providing immediate feedback on detection confidence, the setup demonstrates both the potential and the practical application of advanced computer vision methods for monitoring and managing fire-related hazards.

For our experiments, we selected the following object detection models: YOLO, DETR, CenterNet, and M2Det. After training, the models were evaluated on a held-out test set, with the primary focus on four metrics. Precision was defined as the proportion of correctly predicted objects among all predictions, whereas recall measured the proportion of detected objects among all actual objects. Mean Average Precision (mAP) provided a comprehensive measure of detection quality, and inference speed indicated how many frames could be processed per second. Finally, we compared the performance of these models based on these four metrics.

5 Results

To evaluate the performance of the proposed detection system, we conducted a series of tests under both controlled conditions and real-world scenarios. The controlled environment included testing on standardized datasets with minimal noise to ensure reproducibility. Real-world tests were carried out on images obtained from UAVs under various lighting conditions, complex backgrounds, and from different altitudes.

In the controlled setting, the system was tested on the "Lacmus Drone Dataset." For each model (YOLO, DETR, CenterNet, and M2Det), we assessed the following key metrics: mean average precision (mAP), precision, recall, and processing time per image.

YOLO demonstrated the highest processing speed while maintaining a competitive mAP, and it performed particularly well in detecting medium-sized objects.

DETR achieved the highest mAP thanks to its transformer-based architecture, but it required significantly more computational resources and processing time.

CenterNet showed a good balance between accuracy and speed, especially excelling at detecting small objects.

M2Det performed well in scenarios with dense groups of objects, although it slightly underperformed on small-object detection tasks.

Real-world testing involved images captured by UAVs over various areas, including both urban and rural regions, allowing us to assess how these models adapt to noise, viewpoint changes, and object occlusions.

YOLO exhibited stable real-time performance and high accuracy despite challenging environmental conditions.

DETR outperformed other models in detecting occluded objects but was less suitable for real-time applications due to its slower processing speed.

CenterNet remained reliable for detecting small, scattered objects, such as people in fields or forests.

M2Det demonstrated advantages in urban environments with densely packed objects and more complex backgrounds.

Each model showed distinct strengths and weaknesses depending on the circumstances, emphasizing the need to tailor the choice of architecture to the specific requirements of the task.

5.1 Comparative analysis

A comparative analysis was conducted to evaluate the proposed system against existing solutions in terms of detection accuracy, processing speed, and adaptability to real-world conditions. The analysis covered the following aspects:

Detection Accuracy (mAP). DETR achieved the highest average accuracy, attributable to its transformer-based architecture. YOLO and CenterNet tied for the second-best performance, with YOLO excelling at detecting medium-sized objects and CenterNet performing better on small objects. M2Det was effective at handling densely populated scenes, but its accuracy decreased on small or isolated objects.

Processing Speed. YOLO proved to be the fastest model, making it suitable for real-time applications. CenterNet also demonstrated competitive speed, whereas DETR and M2Det required more processing time due to their higher computational complexity.

Adaptability to Real-World Conditions. YOLO and CenterNet showed the greatest adaptability to challenging conditions such as varying illumination and motion blur, making them ideal for UAV operations in dynamic environments. Despite its high accuracy, DETR was less robust in rapidly changing conditions. M2Det excelled in densely populated urban scenes but was less effective in open spaces.

Resource Efficiency. YOLO and CenterNet were the most resource-efficient models with low memory and computational demands, making them particularly suitable for onboard UAV deployment. DETR and M2Det, however, required more resources, limiting their use in lightweight embedded systems.

6 Discussion

The results of this study demonstrate that there is no single, universal model capable of outperforming all others under every condition. Instead, the choice of an appropriate object detection model should be guided by the specific requirements and constraints of the application. For instance, when real-time performance is paramount—such as in rapid-response scenarios—YOLO stands out as a highly suitable option due to its fast inference speed and reasonably high accuracy. This makes it particularly attractive for onboard deployment where computational resources and time are limited.

By contrast, in situations where the highest level of accuracy is required and computational overhead is less of a concern, DETR proves to be an excellent choice. Its transformer-based architecture enables exceptional performance on complex detection tasks, although it does demand more processing power and time. Meanwhile, CenterNet offers an optimal balance for the detection of small objects, which can be crucial in applications such as searching for humans in expansive rural or forested areas. Its ability to capture fine details without sacrificing too much speed makes it a strong contender in conditions where subtle or partially occluded targets must be identified.

Finally, in dense urban environments, M2Det consistently delivers robust detection results, particularly in scenarios where many objects are crowded into the same frame. Its multi-scale features and capacity to handle complex backgrounds give it a clear advantage in cityscapes. Taken together, these findings underscore that no single model excels universally across all tasks, reinforcing the importance of matching the model architecture to the specific operational setting, priorities, and hardware limitations.

7 Conclusion and future work

The findings of this research underscore the importance of tailoring object detection models to the specific needs of each application scenario. No single model emerged as a universal solution for every environment; instead, the choice depends on factors such as computational resources, required accuracy, and real-time performance constraints. YOLO, with its superior processing speed and balanced accuracy, proves most effective for time-critical missions where onboard hardware must operate under tight resource limitations. Conversely, DETR's transformer-based architecture excels in tasks demanding higher accuracy and advanced handling of occlusion, albeit at the cost of greater computational overhead. CenterNet distinguishes itself in the detection of small and scattered targets, offering a robust compromise between speed and accuracy, whereas M2Det shows particular promise in densely populated urban scenes, benefiting from its ability to manage crowded and complex backgrounds.

Despite these advancements, several challenges remain. First, further research is needed to optimize models for specialized sensors and data types, including thermal or hyperspectral imagery, especially when operating in harsh conditions such as extreme weather or low-light environments. Second, refining algorithms for more efficient resource utilization is crucial for truly autonomous operation, particularly in lightweight UAVs or embedded systems with limited processing power. Finally, integrating detection models with higher-level decision-making algorithms, such as path planning and anomaly response, would enable a more holistic approach, transforming detection systems into fully autonomous platforms capable of both identifying and reacting to critical objects or events in real time.

Future work will therefore focus on developing lightweight ensemble methods that combine strengths of multiple detection architectures. This approach could enhance robustness across diverse scenarios, from open fields with small objects to densely packed urban settings. Additionally, exploring self-supervised or semi-supervised learning methods might reduce the burden of manual data labeling, making it easier to adapt models to new tasks and environments. Lastly, extending the system's capabilities to detect and analyze not only static objects but also dynamic phenomena—such as fires, floods, or other evolving events - would further increase its applicability in disaster management, environmental monitoring, and numerous other high-impact domains.

This work was supported by the Ministry of Science and Higher Education of the Russian Federation (Grant № 075-15-2022-1121).

References

1. Z. Xu, X. Zhan, Y. Xiu, C. Suzuki, K. Shimada, IEEE Robot Autom Lett. **7(4)**, 876-884 (2022).
2. Y. Zhang, X. Chen, W. Li, J. Auton Robot. **37(3),** 542-556 (2023).

3. M. Johnson, A. Torres, Remote Sens Environ. **256,** 112301 (2021).
4. J. Ma, H. Liu, K. Wang, IEEE Trans Cloud Comput. **11(2),** 198-211 (2023).
5. X. Yao, Z. Wang, T. Lin, Sensors **23(5)**, 1247 (2023).
6. A. Pyataeva, N. Gulyutin, A. Mikhalev, *Determination of Tree Species Using UAV Data in the Problem of Forest Taxation on the Territory of the Kuznetsovskoe Plateau* in Processing of Spatial Data in Problems of Monitoring Natural and Anthropogenic Processes (SDM-2023), pp. 137-142, Berdsk, Russia (2023).
7. O. A. Antamoshkin, O. A. Antamoshkina, E. R. Bryukhanova, A. O. Stupin and N. V. Kamenskaya, IOP Conf. Ser.: Earth Environ. Sci. **981,** 032015 (2022).
8. L. Chen, T. Zhang, Y. Li, *Multimodal object detection and ranging system for autonomous driving* IEEE Trans Intell Veh. **7(1)**, 35-47 (2022).
9. L. Zhao, S. Han, W. Jin, *Real-time object detection for UAVs using embedded NPUs* IEEE Sens J.**22(14),** 9012-9020 (2022).
10. F. Lin, C. Wei, L. Shi, IEEE Robot Autom Lett. **7(2),** 1149-1156 (2022).