

Object Detection Models in Counter-Unmanned Aircraft Systems Based on Devices with Limited Computing Capabilities

Oleh Zaritskyi^{1,*}

¹ Taras Shevchenko National University of Kyiv, 64/13, Volodymyrska Street, Kyiv, 01601, Ukraine

Abstract

The study provides a detailed analysis and identifies the most common and powerful computer vision methods for detecting and identifying unmanned aerial vehicles (UAV) depending on environmental conditions from the perspective of using these models on low-power computing devices on board an interceptor UAV. The author investigated the performance of some popular machine learning (ML) models, such as MobileNetV2 SSD FPN-Lite, FOMO (Faster Objects, More Objects). The simulation was carried out using a training dataset obtained by storyboarding a video of a real UAV use and pre-processing the images by scaling them and labeling them with appropriate labels. The obtained result of the accuracy of even the model converted with the help of Tensor Flow light with a reduced bit depth of feature weights indicates the possibility of using such models in computing modules of the Raspberry PI, Arduino etc. types on board the interceptor UAV. Specific imaging factors that have the greatest impact on detection and identification accuracy were also identified. A combination of various machine-learning methods have been proposed to build an effective Counter-Unmanned Aircraft Systems at any time of day and environmental conditions.

Keywords

C-UAV, UAV, computing vision, object detection, salient objects

1. Introduction

The widespread use of unmanned aerial vehicles (UAVs) of different classes and types in recent armed conflicts has resulted in significant material and human losses for both military units and civilians. The Shahed-type barrage munitions, also known as tactical UAVs of the second class according to NATO classification (weighing 150-600 kg, with a practical ceiling of up to 5000 m), pose a particular threat to the country's critical infrastructure [1] and civilian people.

Figure 1 shows the dynamics of Shahed UAV launches against critical infrastructure and civilian targets during the full-scale war in Ukraine. The analysis is based on the "Massive Missile Attacks on Ukraine" database, which contains information on launched and shot down missiles and drones during massive strikes on infrastructure since October 2022 as part of the invasion of Ukraine.

The database was created manually using official reports from the Air Force Command of the UA Armed Forces and the General Staff of the Armed Forces of Ukraine, published on social media [2].

Since the start of the full-scale invasion of Ukraine around 8000 UAVs of this type UAV have been launched till the January of this year, as shown in Figure 1. The cursory analysis of the data and the trends in Fig. 1 suggest that UAV attacks will continue to increase by the end of 2024. It's worth noting that UAVs are used in combined attacks with different types of missiles to overload air defense systems.

Given the growing production and use of UAVs, as illustrated in Figure 1, and the significant cost of neutralizing them using current air defense systems like MANPADS (Man-portable air-defense systems) and operational/tactical air defense systems, a critical scientific and technical goal is to

ICST-2025: Information Control Systems & Technologies, September 24-26, 2025, Odesa, Ukraine.

* Corresponding author.

† These authors contributed equally.

✉ olegzaritskyi@gmail.com (O. Zaritskyi)

ORCID ID 0000-0002-6116-4426 (O. Zaritskyi)



© 2025 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

develop cost-effective, unmanned counter-UAV systems. Research in this area has been ongoing for several years yielding specific advancements as detailed in the cited reference [3].

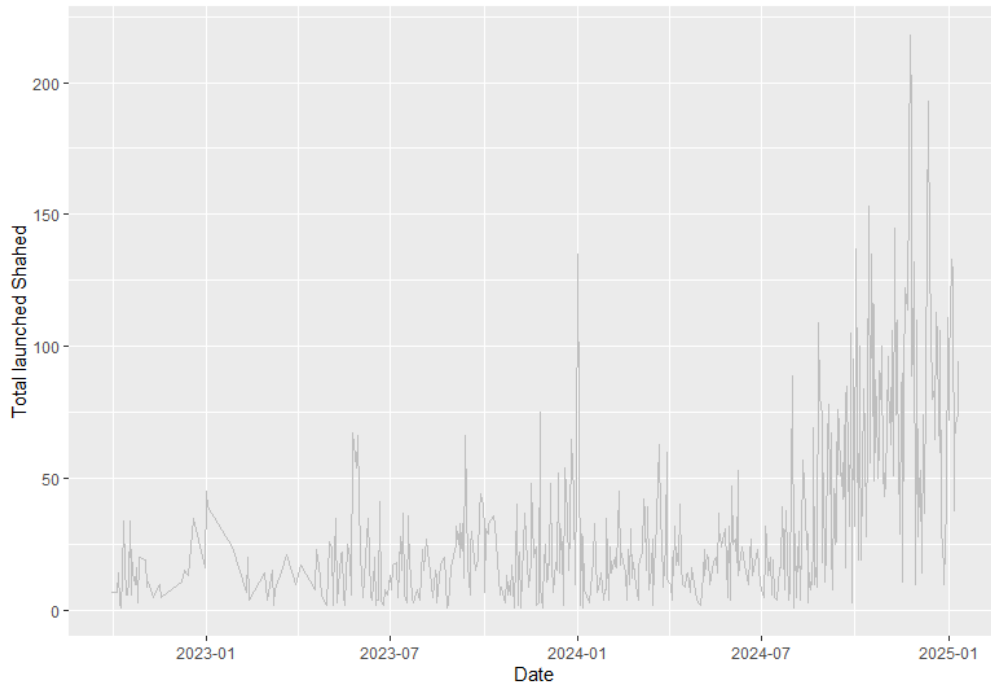


Figure 1: Dynamics of UAV Shahed use

The purpose of this paper is formalizing the most effective methods of UAV detection by analyzing existing approaches with regard to the real conditions of particular UAV type use from the perspective of using these models on low-power computing devices on board an interceptor UAV.

The main tasks of the research:

1. Analysis of existing methods and approaches for detecting and tracking targets at any time of the day, taking into account the tactics of using a specific type of UAV.
2. Modeling and assessing the accuracy of detecting Shahed-type UAVs using the most powerful models.
3. Evaluation of the possibility of using UAV detection models using low-power computing resources such as Arduino etc.
4. Identification of factors that most affect the accuracy of detecting and identifying a specific type of UAV.
5. Development of a general concept of a comprehensive UAV countermeasure system at any time of the day based on an interceptor UAV.

The emphasis is placed on the Shahed UAV due to its widespread use.

2. Analysis of existing researches

Research on UAV detection and identification is mainly focused on improving existing machine learning methods in computer vision tasks in order to increase the efficiency of detecting different types of UAVs. Let us consider the existing approaches in more detail.

It should be noted at once that there is a large number of works devoted to acoustic object detection. For example, in [4], a neural network was built to recognize the sound pattern of the Shahed-136 engine due to its characteristic spectrum of frequencies and tones.

The review paper [5] is interesting in terms of providing a thorough analysis of the methods used to detect and track UAVs or drones. Common methods are described that allow measuring the

position, speed, and image of UAVs, and then using them for detection and tracking. Hybrid detection methods are also presented. The article is a quick reference for a wide range of methods used in the UAV detection process.

In [6], the authors tested the applicability of an inexpensive long-wave infrared sensor for detecting various UAVs in flight. A large number of works have been devoted to the study of detection methods in the infrared range. For example, article [7] describes a visible and thermal monitoring system that combines detection and tracking modules based on deep learning. The authors present an integrated detection and tracking system that outperforms the performance of each individual module containing only detection or tracking. Experiments have shown that even based on synthetic data the proposed system performs well on real UAV images with complex backgrounds.

Similar to infrared cameras, optoelectronic cameras have been widely researched for use in UAV detection and countermeasures. In recent years, deep convolutional neural networks (DCNN) have become the main tool for the development of visual UAV detection and classification systems. Paper [8] investigates the issues and various approaches to building convolutional networks in object detection tasks.

Object tracking is one of the most important tasks in computer vision, which has many practical applications such as motion monitoring, robotics, autonomous vehicle tracking, etc. Various studies have been conducted in recent years, but due to various problems such as occlusion, lighting variations, fast movement, etc., research in this area continues. Article [9] explores various methods of object tracking and presents a comprehensive classification that categorizes tracking methods into four main categories: feature-based, segmentation-based, evaluation-based, and learning-based methods.

The development of machine learning methods for UAV detection and classification requires training datasets that allow training and adjusting model parameters to obtain an acceptable level of detection accuracy, which in turn has created a powerful impetus for the development of research in the field of creating these datasets. In [10], the authors present a new dataset created entirely for training computer vision-based machine learning algorithms for object detection. The dataset extends the existing multi-class image classification and object detection datasets (ImageNet, MS-COCO, PASCAL VOC, anti-UAV) with a diverse set of UAV images.

As a rule, in most studies, object detection is described as the problem of determining the location of an object in an image (localization task) and which category (class) this object belongs to (classification task) and involves several stages: selecting the target area in the image, obtaining semantic features, and classification itself [11].

Faster R-CNN achieves state-of-the-art performance on generic object detection. However, a simple application of this method to a large vehicle dataset performs unimpressively. In paper [12], authors take a closer look at this approach as it applies to vehicle detection. They show that through suitable parameter tuning and algorithmic modification, we can significantly improve the performance of Faster R-CNN on vehicle detection.

Paper [13] presents YOLO, a new approach to object detection. Prior work on object detection repurposes classifiers to perform detection. Instead, authors frame object detection as a regression problem to spatially separated bounding boxes and associated class probabilities. A single neural network predicts bounding boxes and class probabilities directly from full images in one evaluation.

Paper [14] introduces G-CNN, an object detection technique based on CNNs which works without proposal algorithms. G-CNN starts with a multi-scale grid of fixed bounding boxes. Authors train a regressor to move and scale elements of the grid towards objects iteratively. G-CNN models the problem of object detection as finding a path from a fixed grid to boxes tightly surrounding the objects. G-CNN with around 180 boxes in a multi-scale grid performs comparably to Fast R-CNN which uses around 2K bounding boxes generated with a proposal technique. This strategy makes detection faster by removing the object proposal stage as well as reducing the number of boxes to be processed.

Paper [15] presents a method for detecting objects in images using a single deep neural network. The approach, named SSD, discretizes the output space of bounding boxes into a set of default boxes over different aspect ratios and scales per feature map location. Compared to other single stage methods, SSD has much better accuracy even with a smaller input image size.

The analysis of works in the field of C-UAS systems development allows us to draw the following conclusions:

1. Existing approaches involve the widespread use of optoelectronic cameras in the visible and infrared ranges of the radiation spectrum in UAV detection tasks.
2. Researchers are mainly improving existing machine learning methods and widely applying transfer learning to object detection and tracking tasks.
3. None of the systems under consideration fully satisfies the requirements for the implementation of highly effective C-UAS systems in terms of ensuring the detection of UAVs within the framework of possible scenarios of their use.
4. All the papers are devoted to theoretical research and, in some cases, description of elements of practical implementation of UAV detection and classification systems. None of the papers considers the implementation of the interception concept as a logical extension of their research.

That is why the author's idea of building a combined anti-drone system using methods of recognizing the intruder UAV, implemented on low-power computing boards on board the interceptor UAV, day and night, is an urgent practical task. The article presents the intermediate results of the research and the general architecture of building such a system.

3. Machine learning methods in objects detection tasks

Existing object detection methods are conditionally divided into several large groups, such as detection of ordinary objects (detection of faces, human figures, contours of machinery, etc.) and detection of salient objects.

3.1. Detection of ordinary objects

UAV detection through machine vision during daylight employs two primary approaches:

1. Object area detection with subsequent classification (R-CNN models: R-CNN, Fast R-CNN, Faster R-CNN) [12].
2. Regression and classification methods (MultiBox, AttentionNet, YOLO, G-CNN, SSD, YOLOv2) [13-15].

Both techniques aim to efficiently recognize and identify objects in images. It is worth noting the large size of the Shahed-136 UAV, which is a classic wing measuring about 3x4 meters, which allows it to be visually detected at low altitudes with the naked eye and simplifies the setup of the recognition model by its characteristic silhouette.

In order to compare several algorithms, a training dataset was formed, which included up to 1000 images of the Shahed-136 flight in 1280x720px. size, obtained by storyboarding the corresponding videos. The image set was divided into training and testing in a ratio of 78 to 22 percent, respectively. All images were pre-processed with appropriate labels to create the training set and implement supervised learning. Due to the small amount of labeling and the large number of errors when trying to implement AI labeling, the labeling was done manually by applying a bounding box and a corresponding mark.

It's very hard to build a good working computer vision model from scratch, as you need a wide variety of input data to make the model generalize well, and training such models can take days on a GPU. To make this easier and faster for researching tasks transfer learning is usually used. Such

approach lets to use a well-trained model, only retraining the upper layers of a neural network, that lead to much more reliable models that train in a fraction of the time and work with substantially smaller datasets.

The simulation input data was the same for all models to be able to compare the training results. 20 epochs and a learning rate of 0.05 were used for training. The validation set was 20 percent of the training set; batch size was set from 2 to 4. Hardware acceleration was not used, nor was the GPU used during training. To use the trained model, the architecture was simplified by Quantization (from float32 to int8) to reduce size and speed up. The model was then exported to .tflite format (TensorFlow Lite) for deployment on the target device (Raspberry PI 3, Cortex-A53 (4 cores) – 1,5GHz).

3.1.1. MobileNetV2 SSD FPN-Lite.

MobileNetV2 SSD FPN-Lite – a pre-trained object detection model designed to locate up to 10 objects within an image, outputting a bounding box for each object detected. The model is around 3.7MB in size. It supports an RGB input at 320x320px. The architecture is specifically optimized to work on devices with limited computing capabilities, which is achieved thanks to the lightweight MobileNetV2 base model. Due to the model using input images of size 320x320, their reduction was carried out using several methods. 'Squash' will resize via interpolation, using the entire image. The longer axis will look squashed. 'Fit shortest axis' will first crop the outsides of the longer axis to the desired aspect ratio, then interpolate to desired size. 'Fit longest' will 'letterbox' the short axis to the desired aspect ratio, then interpolate.

The model achieved an 0.88 mean average precision (mAP) during training (validation dataset) and 0.77 on the test dataset. Increasing the number of epochs to 40 and setting learning rate to 0.02 increase accuracy to 86% on the test dataset (fig.2).



Figure 2: a) Video frame with Shahed-136 image, and b) probability of object detection with MobileNetV2 SSD FPN-Lite model

MobileNetV2 SSD FPN-Lite is an efficient architecture for object detection on resource-constrained devices, providing a good balance between speed and accuracy. However, running it on standard Arduino boards is practically impossible due to significant hardware limitations. To implement such computer vision tasks on microcontrollers, it is more appropriate to use much simpler models or specialized hardware solutions, such as the Arduino Portenta H7 (or Arduino Nano 33 BLE Sense) in combination with TensorFlow Lite for microcontrollers and additional quantization and optimization of the model. To increase the accuracy of the model and fine-tune the task of recognizing specific objects of a characteristic shape, freezing the layers of the MobileNetV2 block (18 layers) and training SSD Head and FPN-Lite on 1000 samples was used.

3.1.2. FOMO (Faster Objects, More Objects) MobileNetV2 0.35.

An object detection model based on MobileNetV2 (alpha 0.35) designed to coarsely segment an image into a grid of background vs objects of interest. This architecture represents an innovative approach to solving the problem of fast and accurate detection on edge devices, including microcontrollers and single-board computers. These models are designed to be <100KB in size and support a grayscale or RGB input at any resolution. FOMO is a brand-new approach to run object detection models on constrained devices. FOMO is a ground-breaking algorithm that brings real-time object detection, tracking and counting to microcontrollers for the first time. FOMO is 30 times faster than MobileNet SSD and can run under 200K of RAM. Simulation results for the test set (proportion 78/22 of the data set) using images from the previous experiment are presented by Confusion matrix (table1) and fig.3.

Table 1

Confusion matrix (validation set)

	Background	Shahed-136
Background	100%	0%
Shahed-136	5.3%	94,7%

The model achieved an accuracy of 0.97 (F1 Score) during training (validation dataset) and 0.88 on the test dataset.

The model was also converted using Tensor Flow light (reduced the number of feature weights to 8) in order to deploy the model on an Arduino device and obtained the following indicators: inferencing time – 833,98 ms, peak RAM usage - 2,8 MB, flash usage - 11,3 KB. At the same time, it should be noted that the detection accuracy decreased to 0,84.



Figure 3: a) Video frame with Shahed-136 image, and b) probability of object detection with FOMO model

FOMO radically changes the standard approach to object detection. Unlike traditional models that assume bounding boxes and perform classification within these boxes, FOMO uses semantic segmentation as the basis for detection. This engineering change allows the model to be significantly lighter and faster. Due to its small size, the model can be deployed on microcontrollers such as Arduino Portenta, ESP32-CAM, or STM32 for basic object detection.

The main challenges in presented training models:

1. Poor detection of distant (small) objects that blend into the background (ground, buildings, trees). The performance of FOMO is highly dependent on the resolution of the input image. Very small objects in low-resolution images may be missed.

2. Poor detection of large objects (foreground) when the UAV is almost the same size as the house against which it was shot (the geometry of the house, the triangular roof (similar to the shape of the UAV, etc.) affect the detection). This problem was typical for MobileNetV2 SSD, and the FOMO model solved this problem completely (fig.4).
3. Sensitivity to input quality. Compact models are often more sensitive to noise, poor lighting, or unusual angles of objects.

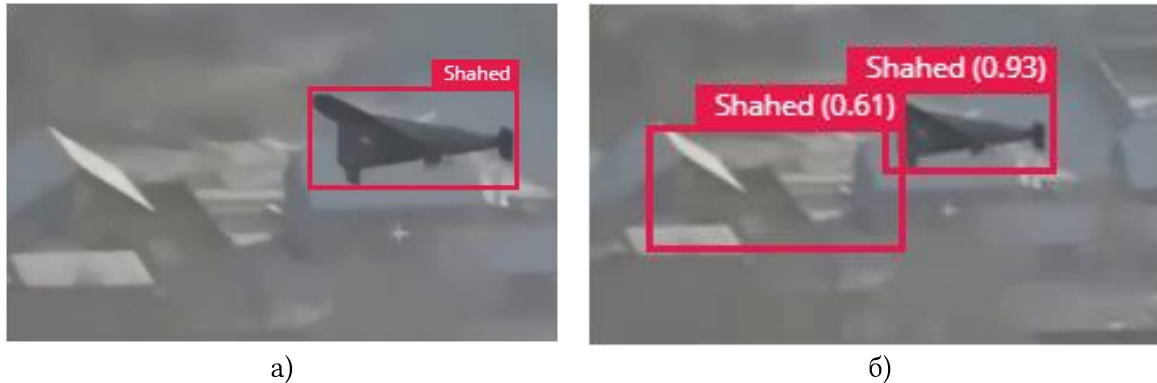


Figure 4: a) Video frame with Shahed-136 above the triangular roof, and б) the probability of detecting a roof instead of a UAV

4. The need for fine-tuning hyperparameters to achieve the optimal balance between speed and accuracy in a specific problem.

3.2. Detection of salient objects

One of the most challenging tasks in the field of computer vision is the detection of salient objects, which involves highlighting the most dominant parts or the object as a whole in an image. There are several approaches to solving the problem of detecting salient objects, the most common of which are bottom-up (BU) and top-down (TD).

The first approach involves studying the contrast of local and global features of image pixels, regardless of the overall context of the image, which allows obtaining a low-contrast map of a salient object instead of the image of the object itself. The second approach requires prior knowledge of the object, i.e., contextual detection, which is the basis for building (recognizing) a contrast map. In the case of a combination of the two approaches, the TD approach detects the object of attention (its contours), which subsequently allows rejecting pixels that stand out strongly in contrast but do not belong to this object.

These methods of detecting salient objects, as in case of detecting ordinary objects, use convolutional neural networks as a basis, which was improved by the authors of the studies [16, 17].

A review and comparison of the effectiveness of the main models for detecting salient objects, such as: CHM (Contextual hypergraph modeling), RC (Regional Contrast based salient object detection), DRFI (discriminative regional feature integration), MC (Multi-Context deep learning), MDF (Multiscale Deep CNN Features), LEGS (Local Estimation and Global Search), DSR (Deeply-Supervised Recurrent CNN), MTDNN (Multi-Task Deep Neural Network), CRPSD (Combining Region-level and Pixel-level Salient Detection), DCL (Deep Contrast Learning), ELD (Encoded Low level Distance), NLDF (Non-Local Deep Features), and DSSC (Deeply Supervised Salient object detection with short Connections) are made in [8].

It should be noted that there are no specialized training datasets that would cover a wide range of UAVs, especially dual-purpose ones. The following datasets can be used as training sets for training models for detecting salient objects: ECSSD, HKUIS, PASCALS, and SOD datasets, which are usually represented by 300 to 4000 images with salient objects.

Two standard metrics are usually used to evaluate the quality of a visibility mapping model: F-measure and mean absolute error (MAE). Taking into account the values of Precision and Recall, for

a pre-built binary mask B of the saliency map and a reference image, the F-measure is calculated as ($\beta^2 = 0,3$) (1):

$$F_\beta = \frac{(1 + \beta^2)Precision \times Recall}{\beta^2 Precision + Recall}. \quad (1)$$

The mean absolute error is calculated (2):

$$MAE = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W |\hat{S}(i,j) - \hat{Z}(i,j)|, \quad (2)$$

where

H, W - the height and width of the visible area, respectively;

\hat{S}, \hat{Z} - represent the ground truth and the continuous saliency map, respectively.

CHM, RC, DRFI recognition methods, which are classified as classical classification methods, achieve the highest model quality ($F_\beta = 0.631 - 0.712, MAE = 0.143 - 0.222$) on these datasets, while other mentioned methods based on convolutional neural networks perform much better ($F_\beta = 0.721 - 0.913, MAE = 0.037 - 0.12$) [8].

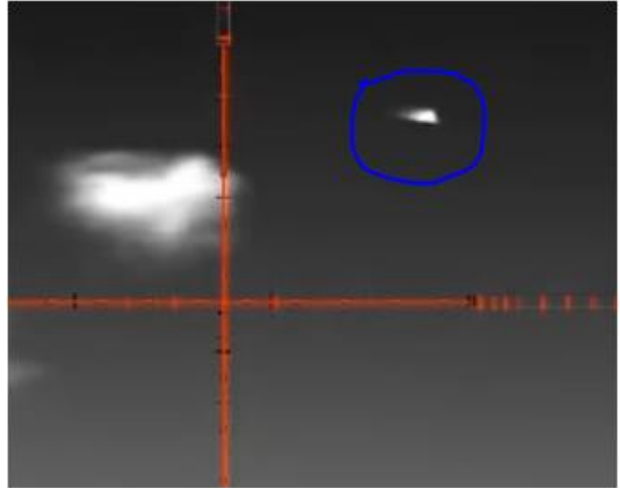
Studies of UAVs with piston engines of the Shahed-136 type in terms of obtaining IR signatures have not been carried out, which significantly complicates the development of appropriate detection models.

The UAV type Shahed 136 uses the L550E (MD-550) engine, a four-cylinder horizontally opposed two-stroke air-cooled gasoline engine that develops a power of 37 kW (Fig. 5a) [18].

The cooling system is air. Working volume - 548 cm³, length - 300 mm, width - 410 mm, height - 301 mm, dry weight - 16 kg, power-to-weight ratio: 2.3 kW/kg.



a)



б)

Figure 5: a) Engine L550E, the downed Shahed fell into the field and b) its IR signatures

The operating temperature of the cylinder walls of an air-cooled engine is usually 150-180 degrees. Air masses passing through the cooling fins at speeds of 150-180 km/h, taking into account the ambient temperature, remove part of the heat from the engine and reduce its temperature to 100-130 degrees, thereby reducing infrared radiation.

Consequently, the infrared emissions of the Shahed-136 engine operate within the long-wave spectrum of 8–12 microns. This property facilitates compatibility with a broad range of infrared cameras and sensors, such as the MLX90640-D55 Thermal Camera, HT U-01 MINI 256 CVBS thermal camera designed for drones, and the FPV camera RunCam Night Cam Prototype, among several others that are compatible with computing modules like Raspberry Pi, NVIDIA Jetson Nano, and Arduino. These devices can detect objects at distances of up to 500–1000 meters, depending on external conditions (Fig. 5b).

4. Concept of the combined system of detection, recognition and interception of UAV

Since the use of Shahed-136 UAVs, it has been established that since the beginning of the armed aggression, these UAVs have been used most often and are most effective at night, between 23:00 and 06:00. This is when it is difficult to visually detect and determine their number.

The flight trajectory is straightforward during the marching phase. The average speed of the Shahed-136 kamikaze drones observed during combat use is 140-150 km/h (although movement at speeds of 80 to 180 km/h was observed), and the flight altitude on the marching section is from 700 m to 2000 m, and in the target area it drops to 200 m.

Thus, for the tasks of detecting UAVs of the Shahed-136 ("Geranium-2") type, taking into account the real data of practical detection of such objects by the Defense Forces of Ukraine, the following detection distances can be formed (Table 2). The most effective methods of detection and identification are combined into one logical scheme of their application (fig.6).

Table 2
Detection distances Shahed-136

Characteristics	Means of detection	Detection distance, km.
Material object (geometric dimensions)	Radar (active radiation)	up to 50
	Visual detection	0,5 - 1
Material object (visibility)	Optical and electronic cameras	0,5 - 2
Thermal radiation	Infrared camera	0,5 - 2
Radio frequency signal	Radio frequency receiver	-
Acoustic signal	High-sensitivity microphone (array)	1 - 5

The visual detection stage begins when the optical-electronic interceptor (OE/IR) camera installed on the UAV clearly identifies the target. At this stage, in addition to target detection, additional target identification can be performed, i.e., confirmation of the correctness of the identification at the previous stage (acoustic, radar, etc. detection and identification). As discussed in the previous section, it is possible to use an infrared camera that will allow the UAV interceptor to be pointed at the hot engine of a UAV such as the Shahed-136. It is obvious that it is necessary to use a combined interception system that will combine various methods to ensure the greatest efficiency and all-weather capability (fig.6). Thus, in normal visibility conditions, optoelectronic cameras can be used, and in night conditions, interception and guidance can be carried out using an infrared camera. The next stage is to track and destroy the target using the segmentation tracking method. Tracking is carried out by tracking a rectangular frame that was superimposed on the detected UAV at the previous stage, and targeting is performed by aligning the Z-axis of the interceptor aircraft with the center of the tracking frame.

Countering drones could be done using one or a combination of several approaches, namely: physically destroying the drone, neutralizing the drone, taking control of the drone and was reviewed by author in article [3] in detail.

The main part of the interceptor UAV can be made in a conventional configuration, i.e., of hard metals, which involves a dynamic impact with the target at high speed and destruction of the propellers and engine, or it can be equipped with a small charge that can be detonated at the moment of impact with the target. The explosion is more relevant for daytime tracking and interception using an optoelectronic camera that allows detecting the target as a whole, rather than its individual elements such as an engine, and imposing a tracking frame on the entire object, which can limit the effects of dynamic damage due to the weight and size ratio of the target and interceptor.

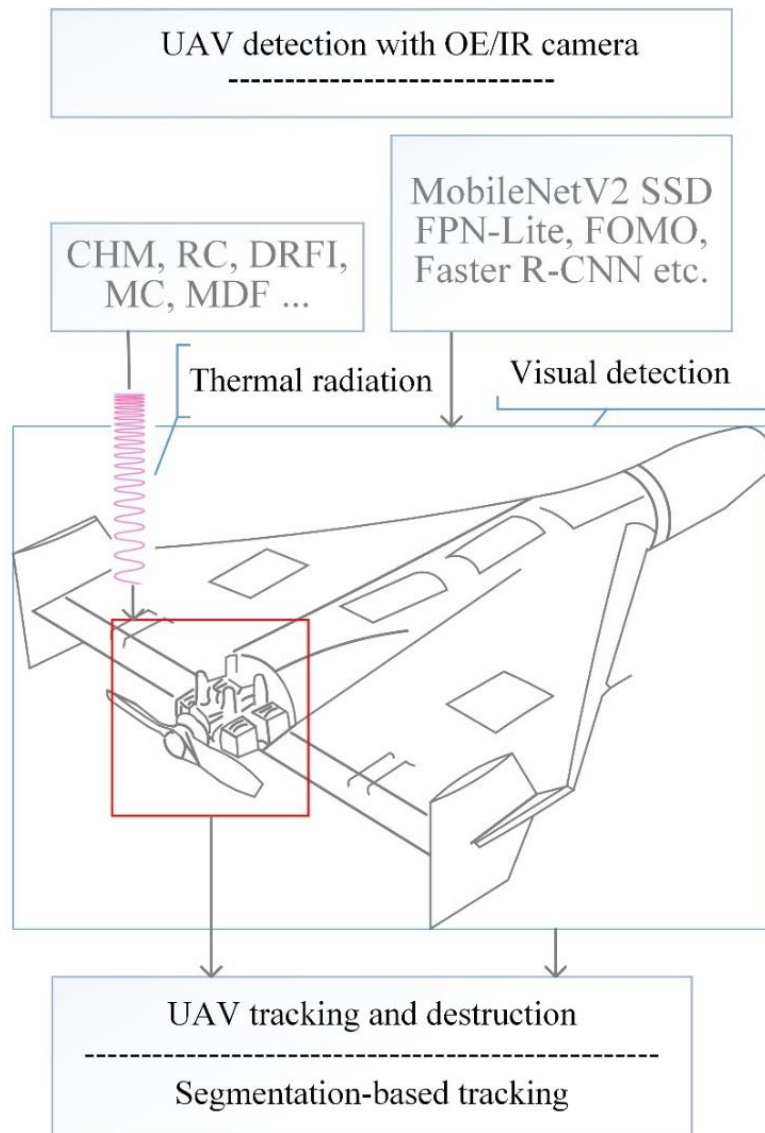


Figure 6: Methods of Shahed-136 detection

5. Conclusion

The results of the research and calculations carried out within the framework of the goals and objectives allow us to draw the following conclusions:

1. The most problematic issue in area of building C-UAS system, as shown by the analysis, is the lack of training datasets that would describe different scenarios of UAV use, not to mention the variety of their types. That is why all further research will depend on filling the training sets with high-quality video and photo materials.
2. The proposed UAV detection by the C-UAS system involves the use of an interceptor UAV as a means of detecting and defeating a target, using a combination of machine vision methods under different interception conditions, which will provide a synergistic effect and significantly expand the capabilities of interceptor UAVs. The relevance of this approach is due to the characteristics of certain types of UAVs (size, engine infrared spectrum, etc.) and the tactics of their use mainly at night, which makes the combination of detection in two ranges (visible and infrared) quite an urgent task.
3. Solving the target detection problem and imposing a rectangular bounding box will allow to proceed to UAV tracking using one of the segmentation-based tracking methods, which

involves solving a regression problem and keeping the longitudinal axis of the interceptor UAV in the center of the bounding box.

4. The results obtained regarding the accuracy of detection and identification at the level of 90% and small size of models allow to speak about the possibility of using the considered models on devices with limited computing power such as Raspberry Pi, NVIDIA Jetson Nano and Arduino computing modules.
5. The Raspberry Pi is a much more powerful platform than Arduino for deploying machine learning models, including the MobileNetV2 SSD FPN-Lite and FOMO. Given its specifications, the Raspberry Pi can provide acceptable performance for real-time object detection tasks, especially with optimized models and hardware accelerators. However, it is important to understand that performance will still be limited compared to more powerful systems. For complex tasks or high-resolution video processing, additional optimizations or even a move to more powerful platforms such as the NVIDIA Jetson Nano or Xavier NX may be required.
6. FOMO does not provide precise bounding boxes, only centroids and segmentation masks. If precise object coordinates are critical for the application, this can be a problem.

Thus, based on the results of the study and the presented conclusions, it is possible to recommend the use of the two presented models, but with certain limitations, namely: model MobileNetV2 SSD FPN-Lite requires a more powerful board, for example Arduino Nano 33 BLE Sense, Raspberry Pi is better and will be more effective in target tracking tasks; model FOMO is more powerful and faster, can use board Arduino Nano 33 BLE Sense, but has certain limitations regarding the implementation of target tracking, since not provide precise bounding boxes.

Declaration on Generative AI

The author has not employed any Generative AI tools.

References

- [1] A Comprehensive Approach to Countering Unmanned Aircraft Systems. Joint Air Power Competence Centr. - 2024. - 644 p.
- [2] Petro Ivaniuk, "Massive Missile Attacks on Ukraine," *Kaggle.com*, 2022. <https://www.kaggle.com/datasets/piterfm/massive-missile-attacks-on-ukraine/data>
- [3] O. Zaritskyi and A. Miroshnyk, "Machine learning in Counter-Unmanned Aircraft Systems," in *XI INTERNATIONAL CONFERENCE. Information Technology and Implementation (IT&Is-2024)*, Kyiv: Taras Shevchenko National University of Kyiv, Nov. 2024, pp. 100–103.
- [4] Oleh Zaritskyi, Andrii Miroshnyk. A combined method for recognizing and intercepting unmanned aerial vehicles using machine learning methods. XII International Scientific and Practical Conference "Modern problems and achievements of radio engineering, telecommunications and information technologies", Zaporizhzhia: December 10-12, 2024. - pp. 127-131.
- [5] Raed Abu Zitar, Mohammad Al-Betar, Mohamad Ryalat, Sofian Kassaymeh. A review of UAV Visual Detection and Tracking Methods. 9th Annual Conf. on Computational Science & Computational Intelligence (CSCI'22), Dec 2022, Las Vegas, United States. - 11 p.
- [6] P. Andrašić, T. Radišić, M. Muštra, and J. Ivošević, "Night-time Detection of UAVs using Thermal Infrared Camera," *Transportation Research Procedia*, vol. 28, pp. 183–190, Jan. 2017, doi: <https://doi.org/10.1016/j.trpro.2017.12.184>.
- [7] Y. Wang, Y. Chen, J. Choi, and C.-C. J. Kuo, "Towards Visible and Thermal Drone Monitoring with Convolutional Neural Networks," *APSIPA Transactions on Signal and Information Processing*, vol. 8, 2019, doi: <https://doi.org/10.1017/atsip.2018.30>.

- [8] Z.-Q. Zhao, P. Zheng, S.-T. Xu, and X. Wu, "Object Detection With Deep Learning: A Review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3212–3232, Nov. 2019, doi: <https://doi.org/10.1109/tnnls.2018.2876865>.
- [9] Z. Soleimanitaleb, M. A. Keyvanrad, and A. Jafari, "Object Tracking Methods: A Review," in *IEEE Xplore*, Mashhad, Iran: IEEE, Oct. 2019, pp. 282–288. doi: <https://doi.org/10.1109/ICCCKE48569.2019.8964761>.
- [10] M. L. Pawelczyk and M. Wojtyra, "Real World Object Detection Dataset for Quadcopter Unmanned Aerial Vehicle Detection," *IEEE Access*, vol. 8, pp. 174394–174409, 2020, doi: <https://doi.org/10.1109/access.2020.3026192>.
- [11] "Object Detection vs. Classification in Computer Vision: Explained," *www.augmentedstartups.com*. <https://www.augmentedstartups.com/blog/object-detection-vs-classification-in-computer-vision-explained>.
- [12] Q. Fan, L. Brown, and J. Smith, "A closer look at Faster R-CNN for vehicle detection," in *IEEE Xplore. 2016 IEEE Intelligent Vehicles Symposium (IV)*, IEEE, Jun. 2016, pp. 124–129. doi: <https://doi.org/10.1109/IVS.2016.7535375>.
- [13] R. Joseph, D. Santosh, G. Ross, and F. Ali, "You Only Look Once: Unified, Real-Time Object Detection," *Cornell University*, Jan. 2016, doi: <https://doi.org/10.48550/arxiv.1506.02640>.
- [14] Mahyar Najibi, M. Rastegari, and L. S. Davis, "G-CNN: An Iterative Grid Based Object Detector," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA: IEEE, Jun. 2016. doi: <https://doi.org/10.1109/cvpr.2016.260>.
- [15] W. Liu *et al.*, "SSD: Single Shot MultiBox Detector," *Computer Vision – ECCV 2016*, vol. 9905, no. 5, pp. 21–37, 2016, doi: https://doi.org/10.1007/978-3-319-46448-0_2.
- [16] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Boston, MA, USA: IEEE, Jun. 2015. doi: <https://doi.org/10.1109/cvpr.2015.7298965>.
- [17] Dashan Gao, Sunhyoung Han, and N. Vasconcelos, "Discriminant Saliency, the Detection of Suspicious Coincidences, and Applications to Visual Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 6, pp. 989–1005, Jun. 2009, doi: <https://doi.org/10.1109/tpami.2009.27>.
- [18] "Aircraft engines from 15 kW to 40 kW. Small and efficient," *Limflug.de*, 2025. <https://limflug.de/en/products/engines-15kw-40kw.php>.