Advances in Science and Technology Research Journal

Advances in Science and Technology Research Journal, 18(7), 364–378
https://doi.org/10.12913/22998624/193368
ISSN 2299-8624, License CC-BY 4.0

Received: 2024.07.15
Accepted: 2024.09.18
Published: 2024.10.06

Artificial Intelligence You Only Look Once – Based Unmanned Aerial System for Remote Sensing in Security Surveillance

Jakub Gutt[1], Michał Kaczor[2], Grzegorz Paleta[1], Krzysztof Polec[1], Kacper Stiborski[1], Filip Strzępek[1], Roman Czyba[1*], Piotr Czekalski[1], Jarosław Domin[3], Nebiyat Gebeyehu[1]

[1] Faculty of Automatic Control, Electronics and Computer Science, Silesian University of Technology, ul. Akademicka 16, 44-100 Gliwice, Poland

[2] Faculty of Transport and Aviation Engineering, Silesian University of Technology, ul. Zygmunta Krasińskiego 8, 40-019 Katowice, Poland

[3] Faculty of Electrical Engineering, Silesian University of Technology, ul. Akademicka 10a, 44-100 Gliwice, Poland

* Corresponding author's e-mail: roman.czyba@polsl.pl

ABSTRACT

Unmanned aerial vehicles are a synergistic technology that complements other new technologies and is in constant development. The paper focuses on using artificial intelligence (AI) in security surveillance. This article aims to develop an unmanned aerial system to monitor border areas and detect human silhouettes in challenging environmental conditions. For this purpose, thermal imaging technology was used for remote sensing in combination with artificial intelligence, particularly Yolo algorithms. After testing various Yolo versions, the target algorithm was implemented on an NVIDIA Jetson Xavier NX edge device. Prototyping of the AI-based thermal detection system was carried out on the DJI S900 multi-rotor aircraft. The final solution was implemented on a vertical take-off and landing aircraft. A summary containing observations and conclusions, as well as perspectives for the development of future work, are included at the end of the paper.

Keywords: drone, UAV, artificial intelligence, artificial neural network, machine learning, control system, thermal imaging, edge computing.

INTRODUCTION

Developing new technologies mainly involves computers, processors, networks, robotisation and automation. Expenditures on devices using IoT (internet of things) technology are constantly increasing, but this is especially visible in the vehicle industry, both manned and unmanned. The transport and logistics sector eagerly uses the option of effective collection, processing, and active exchange of data, where the most crucial thing is streamlining processes and ensuring security and fast service. Over the last two or three years, the influence of AI has become increasingly visible in these activities. Its potential lies in supporting business development in almost every sector of the economy, and not at high costs. In some industries, deep learning, as the most advanced type of machine learning, can contribute to an increase in enterprise revenues by up to 9% [1]. In turn, the benefit to the economy from implementing AI solutions may contribute USD 15.7 trillion to the global economy by 2030, which is more than the current production of China and India combined [2].

The development of the unmanned aerial vehicle (UAV) market, i.e. vehicles using low-altitude spaces (up to 120 m), has also been intensive for several years. The unmanned aircraft sector is becoming the most dynamic sector of the global aviation industry [3]. Currently, drones operating in the airspace at so-called low altitudes are controlled mainly by enthusiasts of modern technologies. However, increasingly bold actions taken by many countries to integrate aviation, telecommunications, and satellite technologies in the near future will ensure the dynamic development of the commercialisation of UAVs that perform autonomous flights. This is intended to make the U-space concept more realistic - a space in which remotely controlled, automatic and, in the future, autonomous flying platforms will safely perform operations thanks to precise air traffic management. Only the dynamic development of airspace management technologies and the parallel development of research on the applications of 5G technology, blockchain, and easy availability of satellite data open up new opportunities for drones and a chance to transition from recreational to entirely professional commercial applications. The automation of services provided using drones, which is already the subject of legislative work, will result in rapid progress, and UAVs will become one of the key elements of IoT, especially if we take into account autonomous flights. With improved flight endurance, payload, and sensor miniaturisation, while their technical capabilities and measurement range increase dramatically, the volume of collected data also increases dramatically. The developed AI technology allows you to organise terabytes of data and provides access to them for all interested parties - GCS operator, pilot, data analysis specialist. It also introduces mechanisms to replace time-consuming, error-prone manual analysis of monitored areas. Instead of uploading high-resolution images to a server or processing them on the drone, artificial intelligence algorithms are used to sample the images and capture only relevant features such as edges, etc. These features can then be sent to a server, requiring much less bandwidth and consuming much less power. This is precisely what the use of AI in UAVs is intended for - developing an automated and intelligent method of acquiring, processing and analysing data. Thanks to this, drones can operate autonomously in a complex and dynamically changing environment. As a result, the development of autonomy will significantly increase the effectiveness of UAV operations and safety in the airspace and on the ground.

The most popular industries using UAVs and AI include agriculture, construction, monitoring and control services (e.g., verification of critical infrastructure conditions). What they have in common is the ability to solve complex problems quickly. Today's UAVs are engaged in much more serious activities than taking photos from a height. A wide range of options for installing components on UAVs, such as thermal cameras, night vision devices, cameras, daylight cameras, direct image transmission, and many others, create virtually unlimited conditions for uniform services to use. Today, we know that they work not only in civilian applications but also in military operations. The conflict in Ukraine showed how much drones have brought changes on the battlefield. The Russian-Ukrainian war is the first drone war on such a scale in which both the so-called FPV (first person view) drones are used, surveying machines and classic military UAVs, including the so-called loitering munitions. Military operations occur on land, air, radio, information and cyberspace. Conducting effective military operations is becoming increasingly complicated and depends on considering more and more elements simultaneously.

The paper is organised into the following sections. The first section introduces the broad subject of unmanned aerial vehicles. The second section reflects on an extensive survey of artificial intelligence techniques used to address many challenges regarding UAV applications. The third section presents the motivation for undertaking this topic and the contribution to its development. The following section introduces the system concept. The fifth section describes the two compared algorithms: Yolo v5 and Yolo v8. The following section presents the network training method, and the results and analyses of the experiments conducted are presented in the seventh section. Section eight provides conclusions, and the last section recommends future work for further research and development.

RELATED WORK

The importance of real-time AI applications in UAVs has been investigated since the growth of the computing power of edge-class devices. Many research articles and applications in the industry have been reviewed, e.g. in [4], presenting the latest developments periodically. This process has even been sped up along with the ongoing Russia-Ukraine war, where the importance of navigation methods based on video and AI has been considered a crucial tool in the navigation of military UAVs during GNSS signal jamming by either party of the conflict.

Several frameworks have been developed for both indoor and outdoor autonomous applications.

In papers [5–7] system for people and object detection in cluttered indoor environments was proposed. In the paper [8] the conceptual framework for moving object detection based on a vision system was presented. In paper [9] a real-time system for small object detection from remote sensing images taken by UAVs was presented.

The systems that implement different artificial intelligence algorithms dedicated to detecting people, various objects, and obstacles or dedicated to monitoring strategy areas (e.g. country borders) are usually based on multispectral video streams, commonly on visual and infrared ranges. An example of a vision system for monitoring a strategic area (airport) for detection of dangerous objects on the runway is presented in the paper [10], another system based on the vision system was presented in the paper [11], this system was used to identify water bodies from satellite images. The vision systems are mainly used to detect objects or people in open space; in the event that other objects obscure the wanted object, the thermal imaging systems are used. An example of such a system was presented in the papers [12–17]. The papers [12–14] presented systems dedicated to human detection, while papers [15–17] presented a fire detection system.

The most common AI model, frequently compared in different papers with other Neural Network models, is Yolo (You Only Look Once [18]). Yolo was initially introduced and developed on the PC platform [18]. Thanks to the model performance optimisation in the following Yolo versions [19], the growth of computing power, and the lowering of the energy demand for computing, off-board object detection and recognition were switched to onboard ones. Nowadays, real-time object recognition is crucial for seamless operations (e.g. without waiting/hanging of UAVs in mid-air) and delivers higher fight efficiency. Thus, current applications prefer on-board object detection and recognition with the use of AI models, which bring real-time results, e.g. presented in papers [20–25]. In particular, the Yolo v5 model (a modified Yolo v5s) and its application in object detection and classification in UAV have been investigated in [26]. A comprehensive review of the performance of object detection using different CNN architectures has been provided in [27]. Hybrid models with preliminary object detection onboard and off-board object classification are also considered applicable in the case of environments with UAV communication links that are stable, solid and of high bandwidth and low latency [28].

A quick review of the Yolo models, their versions, and crucial features is presented below.Yolo v1 - an initial version of the model that treats object detection as a regression problem, predicting spatially separated bounding boxes and class probabilities with a single neural network. YOLO v1 is nowadays considered outdated. It also captures comprehensive object representations and generalises better than DPM and R-CNN on datasets like Picasso and People-Art. Yolo v2, also known as Yolo9000, implemented a 19-layered "Darknet-19" architecture and enhanced the accuracy and performance of the v1 version. It incorporates anchor boxes (single size) in its internal structure and a new loss function. The basis of this loss function involves calculating the sum of squared errors between the predicted bounding boxes and class probabilities and the corresponding values in the ground truth. Yolo v3 introduced improved detection and recognition accuracy with 3 distinct bounding box sizes and a "Darknet-53" internal architecture. Additional "feature pyramid networks" (FPN) increased Yolo v3's performance in detecting and recognising small objects. Yolo v4 is an unofficial branch of Yolo v3 with improved FPN. It also incorporates a CSPNet (Cross Stage Partial Network), a variant of the ResNet architecture, bringing better performance. Yolo v4 uses the CIoU loss function, which is suitable for imbalanced data sets used during model training. Yolo v5, launched in 2020, is an open-source version of Yolo with various cross-platform implementations. It employs EfficientNet and EfficientDet architectures and introduces dynamic anchor boxes and a new pooling layer, SPP (Spatial Pyramid Pooling). Yolo v5 has much better generalisation capabilities than the former versions. Yolo v6 was introduced in 2022 and incorporated EfficientNet-L2 architecture and dense anchor boxes. Yolo v7 employs diverse aspect ratio anchor boxes, which reduce the occurrence of incorrect positive detections. Yolo v8, the most recent version, uses a robust fusion of anchor-free object detection and integrates multiple algorithms. During the project's development, not many facts about this model were known.

MOTIVATION AND CONTRIBUTION

The motivation for performing this project was the problematic situation observed on the Polish-Belarusian border. Attempts to illegally cross it, especially at night, were common and complex to detect due to the dense forests in northeastern Poland. In such a situation, conducting effective operations by border services becomes increasingly complicated and depends on considering an increasing number of factors simultaneously. The growing level of complexity means that the basis for effective action is to have the appropriate so-called situational awareness. ISR (intelligence, surveillance and reconnaissance) capabilities and subsequent target identification are the basis for conducting effective operations in complex and unknown terrain. Information's quantity, quality and timeliness determine the ability to prevent and combat upcoming threats. Only the correct perception of reality and events in relation to time and space, and the significance of these events and their effects, enables an appropriate and effective response.

The ability to install various sensors on the drone perfectly predisposes it to the role of a tool improving situational awareness. A properly equipped drone can provide support as an observation and reconnaissance element. In this situation, the ideal solution to monitor the indicated area is to use an unmanned aircraft equipped with a thermal imaging camera and a software module supported by AI algorithms, which would allow detection and reporting even in poor visibility conditions. However, the detection of objects in images depends on the algorithm's effectiveness, the quality of the collected data, and the computing power of the data processing unit.

The challenges in AI methods for image processing in UAV applications, particularly the constraints of limited resources such as computing power and energy, are significant. These challenges highlight the need for innovative solutions and classify the problem as typical, AI-enabled edge computing.

The current development of the Yolo applications focuses on classical, full-colour images and their processing. There are barely any available applications that use aerial thermal imaging, so the work is innovative in this context and presents a niche, still a new area of research, particularly in the context of edge computing and obtained certainties. Thermal imaging represents different image quality when compared to regular images and has different colour spaces and dynamics; thus, the application of the existing examples, models, and knowledge is not straightforward.

CONCEPT OF THE SYSTEM

The project aimed to develop an unmanned aerial system intended to monitor border areas in terms of uncontrolled crossing of state borders by unauthorised persons in prohibited locations. The assumption of the project was to develop and construct a UAV system capable of flying above the tree line and detecting human silhouettes in challenging environmental conditions, i.e. with limited visibility resulting from operations in forest areas. For this reason, image processing techniques, particularly machine learning technology, were assumed. The image is obtained from a FLIR thermal imaging camera mounted on an unmanned aerial platform. Data and image processing solutions have been implemented to recognise people in a given area and determine their location. The unmanned aerial platform was designed in the configuration of a VTOL (vertical take-off and landing) aircraft with a reconfigurable mechanical and functional structure (Fig. 1). This effect was achieved by a hybrid combination of a conventional airframe and a

Figure 1. A tilt-rotor VTOL aircraft

four-rotor aircraft, with rotors mounted on nacelles. The transition mechanism from vertical to horizontal flight and vice versa involves changing the angular position of individual gondolas (Fig. 2).

A tilt-rotor VTOL aircraft has specific properties - thanks to the reconfigurable mechanical structure, it is able to take off and land vertically without a runway, unlike a classic airframe. This is a vital feature because the plane retains the aerodynamic properties of a classic airframe and thus can perform flights over long distances at high speeds and, at the same time, doesn't need a runway for take-off and landing [29, 30]. The specificity of the system's operation, which involves fast movement along the state border, as well as the need to hover in order to detect potential intruders and perform vertical take-off and landing in a small area, make the VTOL aircraft an ideal carrier of the thermal detection module. The VTOL aircraft is designed to perform long-distance flights and can perform autonomous missions and transport cargo located in a specially designed cargo bay. An example of a practical application may be the delivery of medical supplies to injured people in hard-to-reach areas. In our case, the cargo compartment was adapted to accommodate the installation of a thermal imaging human detection module (Fig. 3), consisting of a FLIR thermal camera, a gimbal, a Herelink V2 long-range transmission system and a dedicated NVIDIA Jetson Xavier NX single-board computer. The technical parameters of the VTOL aircraft are presented in Table 1.

EDGE COMPUTING

The current technical parameters of basic computing units used on unmanned vehicles and

Figure 2. Transition mode: a) front-rear rotors configuration (where: CoG – centre of gravity, δ_t – deflection of the rotor nacelle, f_i – *i-th* rotor force (i = 1,…,4), α – angle of attack, θ – pitch angle, x_b – coordinate x in the body reference frame, x_w – coordinate x in the wind reference frame, x_v – coordinate x in the inertial reference frame with the origin in the CoG), b) tilt mechanism

Figure 3. Cargo compartment with thermal imaging module

Table 1. Technical parameters of the VTOL aircraft

Parameter	Value (unit)
Mass	5 [kg]
Maximum load weight	1.5 [kg]
Minimum horizontal speed	10 [m/s]
Cruising speed	19–21 [m/s]
Maximum speed	30 [m/s]
Flight time	105 [min]
Wing area	64.8 [dm^2]
Wing span	2880 [mm]
Length	1600 [mm]
Propeller diameter and pitch	16 × 8 [inch]

the possibilities of using a data link limit the efficiency of real-time analysis of high-resolution data on board an unmanned aircraft. Lower-resolution sensors or high-resolution data streams are used and filtered to reduce processing time so that the algorithm provides the result in real-time.

Nowadays, a solution with remote power computing is considered outdated and limited as it introduces high latency, low range in video transmission, and low reliability in unstable wireless transmission. Thus, the development of in-UAV AI data processing is becoming more attractive. Low-powered, GPU-enabled hardware solutions and effective object recognition models like YOLO step in and help design and implement AI applications in the edge layer. Those devices are referenced as edge AI processors (coprocessors).

Time is a critical parameter for UAV applications in missions requiring quick response, which is why we are discussing the use of edge computing. Edge computing is a data processing method in which computation, analysis, and data processing occur on end devices such as compact, single-board computers rather than sending data to the cloud or data centre. In the case of artificial intelligence, edge computing allows for data processing (in practice in lower quality) in real-time and outside the server room, directly on end devices. This is particularly useful in applications that require quick response, such as image recognition, detection of objects in camera images, or detection of objects based on acoustic signals or a laser beam.The use of edge computing means faster processing of current information, intelligent, distributed data storage, continuous operation even in conditions of loss of connection with the base station, greater data security, and lower data processing delays.

The NVIDIA Jetson platform is the most known and popular, merging good flexibility (software-based models that are easy to upgrade, compared, e.g., to FPGA) and performance, e.g. Jetson Xavier NX delivers up to 21 TOPS with up to 20 W power consumption. Still, this computing power, even very high, is unsuitable for model training. It is performed in the supervised learning model and later imported from the PC/Mainframe to the edge AI processor. The weight of such hardware, including heatsinks and cabling, is up to 0.5 kg, so it is unsuitable for integration with miniature UAVs. Still, it can be successfully integrated with larger versions, e.g., the DJI Matrice 300 or the VTOL aircraft designed and presented in this paper. Note that the edge AI processor is usually separate from the central flight controller, so it is an additional payload to the UAV.

The concept of our system is based on a FLIR thermal camera, VTOL aircraft and neural networks deployed on the NVIDIA Jetson Xavier NX platform to achieve real-time human silhouette recognition in hard-to-reach environments. In our case, software based on Python scripts and the OpenCV library was developed and implemented for data acquisition and processing. The OpenCV library was employed as a fallback option following the initial implementation of the YOLO object detection framework. In particular, OpenCV's blob detection algorithms were utilized to identify and analyse objects within the image data, providing a robust alternative for feature extraction in scenarios where YOLO's performance was suboptimal. The programs that are created enable effective downloading of images from the FLIR thermal imaging camera and automatically save them at intervals during aircraft flights, which are run on the Jetson Xavier NX onboard computer. The real-time preview function, via written scripts and the Herelink system, allows pilots to precisely monitor and adjust the flight trajectory, directly impacting the quality of the collected data (Fig. 4).

Thermal imaging allows us to carry out missions in low-visibility conditions such as darkness, smoke, and, of course, trees and bushes. The FLIR Vue is compact size and low weight which make it a suitable choice for our applications. The sensor resolution 640 × 512 and image frequency 8.3 Hz in PAL standard was used. The camera feed was directly streamed in real-time through analogue-digital signal video converter to USB port in NVIDIA Jetson.

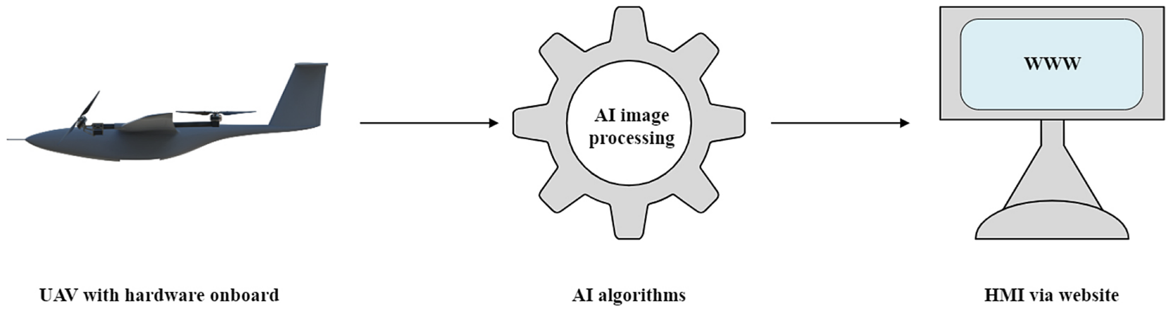

Figure 4. Overall system block diagram

The NVIDIA Jetson platform, known for its powerful AI computing capabilities, processes frames from thermal cameras on the fly, performing parallel calculations on the GPU. Unlike other microcomputers, such as the Raspberry Pi, the NVIDIA Jetson Xavier includes a GPU as a standard feature and offers more RAM. Both the GPU and RAM are crucial for image processing and the use of AI algorithms. All gathered data about detections were transmitted to a web server using an LTE module connected to NVIDIA Jetson Xavier NX. The ground operator could monitor the mission in real-time and check information about detections on an interactive map on the website (Fig. 5).

The system prototype was first tested on the DJI S900 multi-rotor, which was used to collect a dataset for training the network. In general, multirotors are stable and precise, but they are characterised by short flight times (~ 20 min), which are unsuitable for the intended application. Studies show that fixed-wing UAVs seem to be a much better choice regarding flight time and application for planned missions.

ALGORITHMS

This project focused on employing advanced machine learning techniques to recognise human silhouettes in thermal images obtained from a thermal imaging camera. A central aspect of the research was the application of an algorithm suitable for real-time image processing needs. YOLO (You Only Look Once) was selected for this purpose and recognised as a leading tool in dynamic image analysis (Fig. 6).

The specific nature of thermal images, where object contrast and characteristics differ

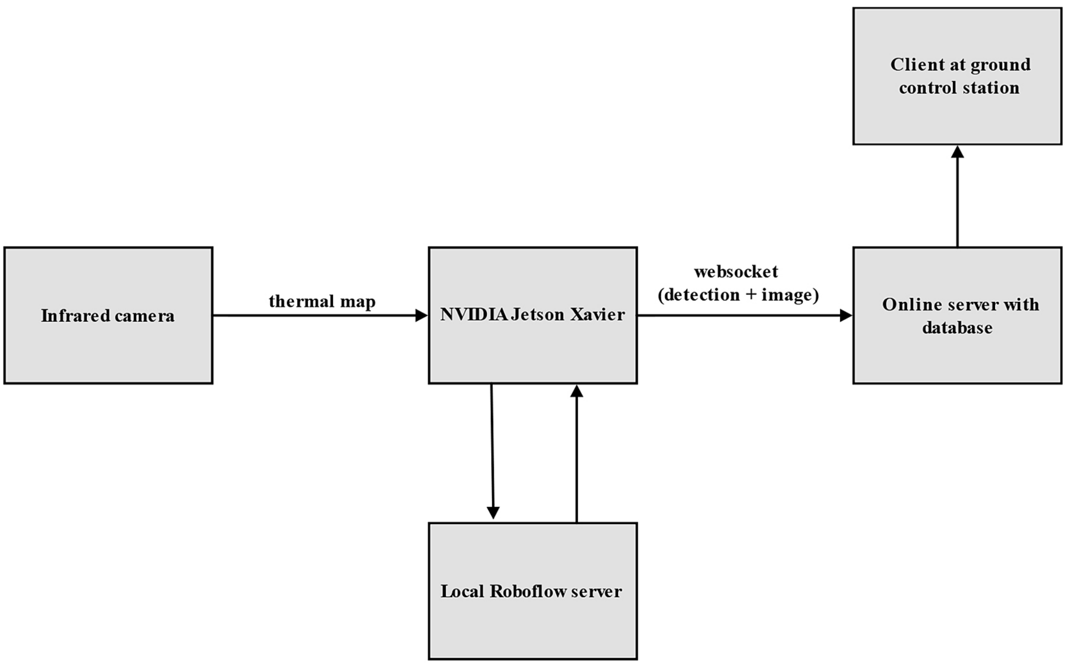

Figure 5. Block diagram of image processing system

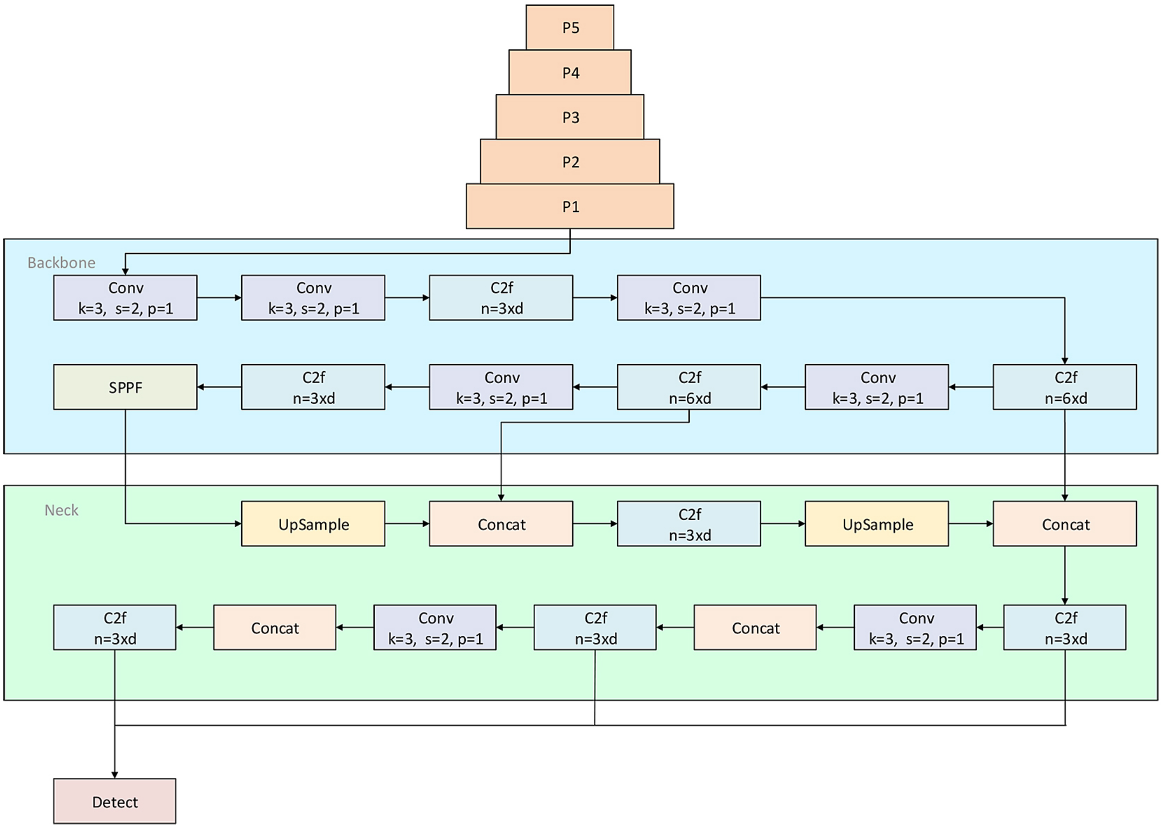

Figure 6. The network architecture of YOLOv5

significantly from images captured by standard RGB cameras, posed a substantial challenge. Therefore, selecting an algorithm capable of effectively operating under these unique conditions was crucial. YOLO was chosen due to its ability to analyse dynamic images in real-time, which is essential for applications requiring immediate system responses, such as public safety monitoring. Its unique architecture allows simultaneous prediction of multiple object attributes (such as location, class, and occurrence probability) in a single network pass, significantly accelerating the image analysis process.

This section covers a detailed comparison of two versions of the YOLO algorithm: YOLOv5 and YOLOv8. Each version has a different structure, directly affecting its performance and effectiveness in various applications. The analysis involves training both algorithm versions on the same input data set and comparing the results.

Network training

This section describes the process of acquiring datasets, the process of training the network itself and comparing their performance.

Details of UAV mission

During collecting images for dataset, some assumptions associated with parameters of UAV mission must have been conducted. From the operational side of using the UAV system, the main aim was to fly as high altitude as possible, to cover as possible the biggest recorded area in a time unit, keeping high detectability of human on the ground. Flights on higher altitudes on this quality type of thermal camera provide problems with human detection, due to the fact that human looks like little point in a circle shape, which can be anything hotter place. Performing flights at lower altitudes than 30–40 meters can be conducted, but the operational risk increases. That risk is associated with possible disturbing of flight by the humans on the ground or might be restricted due to the obstacles like trees. The aim was to conducting flights in a specific area for strategic borders (over fields and forest), so that condition must have been achieved.

In these flight conditions, the images for dataset were collecting. Figure 7 shows the mock area from satellite view, where flights were conducted. It is an area with some trees (red arrow)

Figure 7. Satellite view of the mock area, where test flights were conducted

and grassland (blue arrow). Also the path of flight is marked by the orange line. The green shadows indicates the usable areas covered by the camera. Flights weren't conducted in real conditions on a border, only in a simulated area, which was enough for our purposes. Performing flights in the areas near borders which must be guarded, and for which the system was invented, is mainly restricted for military and borderline guard use only, and flights can be only conducted by them approval. Also the failsafe and damage-tolerant operational aspects of flight weren't for this moment the subject of consideration due to the complexity and military issues. UAVs are used nowadays in such conditions in a manual piloting modes and aspects of safety operation are invented, used and guarded from the open-access by the military [31].

Datasets

The data annotation process was conducted using tools available on the Roboflow platform. Roboflow is an online platform that facilitates working with large image datasets, allowing for easy annotation and conversion into formats compatible with and required by YOLO algorithms. Additionally, Roboflow streamlines the workflow by providing features for dataset management, versioning, and augmentations, making it an essential tool for developing and optimizing computer vision models.

As part of developing an artificial intelligence model to detect human thermal signatures using a thermal imaging camera, a decision was made to gather a dataset independently. This dataset, intended to facilitate effective neural network training, consists of ten thousand video frames obtained from a thermal imaging camera. Data collection occurred in a forest area, using a DJI S900 drone equipped with a thermal imaging camera and an NVIDIA Jetson onboard computer.

During a series of several-minute flights at altitudes of 30–40 meters above ground level, video recordings were made emphasising the diversity of shots. This height has been designated empirically during test flights with specific camera settings, including camera resolution. The manoeuvres performed by the drone aimed to capture the presence of actors in various configurations and camera tilt angles, which contributed to increasing the accuracy and versatility of the future neural network model. These diverse configurations ensured that the dataset encompassed a wide range of scenarios, enhancing the model's robustness and its ability to generalise to different environments.

Upon completion of the data collection phase, the recorded video material was segmented into individual frames. This prepared dataset served as the foundation for further work on developing and training the neural network designed to detect human thermal signatures in thermal images. The segmentation process involved meticulous extraction of frames to ensure that each contained relevant information for training purposes. This comprehensive dataset enabled the neural network to learn intricate details and variations in thermal signatures, thereby improving its detection capabilities.

Moreover, using advanced equipment such as the DJI S900 drone and the NVIDIA Jetson computer facilitated high-quality data acquisition, providing a reliable basis for the subsequent training process. Integrating these technologies was crucial for achieving the precision required for effective thermal imaging and human detection in varied and challenging conditions (Fig. 8).

Training process

This study meticulously compares the training and validation outcomes of two advanced object detection models: YOLOv5s (Fig. 10) and YOLOv8s (Fig. 9). Both models were rigorously evaluated on an identical dataset under consistent settings to ensure a robust comparison of their performance metrics. YOLOv5s and YOLOv8s models were trained on the same images with a resolution of 640 × 640 pixels. The training sessions were conducted on an NVIDIA GeForce RTX 4070 Ti GPU, ensuring optimal performance. The YOLOv8s model was trained for 400 epochs with a batch size of 16. The early stopping mechanism was configured with a patience value of 100, and training was stopped after 87 epochs due to a lack of improvement

a)

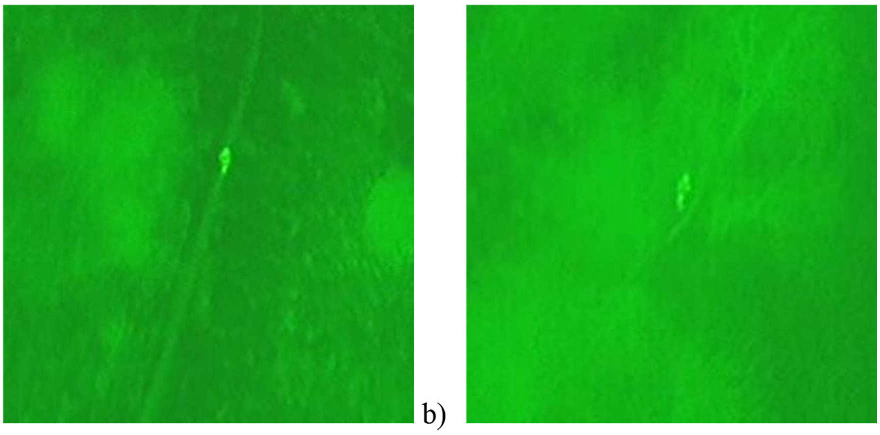

Figure 8. An example of collected video footage with a) a person on the path b) a person hidden among the trees

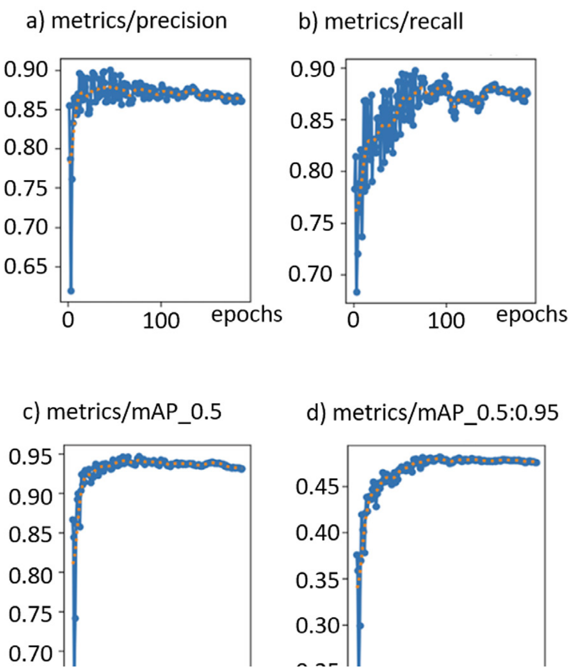

Figure 9. YOLOv8 metrics: a) precision b) recall c) mAP_0.5 d) mAP_0.5:0.95

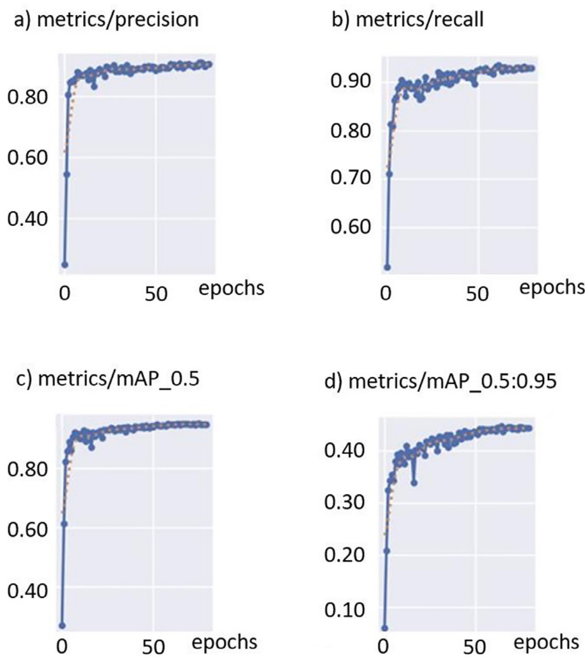

Figure 10. YOLOv5 metrics: a) precision b) recall c) mAP_0.5 d) mAP_0.5:0.95

over the last 100 epochs. Similarly, the YOLOv5s model underwent 400 training epochs, employing the default batch size. The EarlyStopping mechanism, set with a patience value of 15, prompted the termination of training after 64 epochs due to no observed improvement over the last 15 epochs.

Neural networks comparisons

During YOLO training, the following three key metrics were observed, which allow us to assess the quality of the model's object detection: mAP (mean Average Precision), precision, and recall. The meaning of the individual indicators is as follows:

- mAP (mean Average Precision): This metric considers both precision and recall across all object classes in your dataset. It essentially summarises the model's overall detection accuracy. A higher mAP indicates better performance.
- Precision: This metric focuses on the proportion of correctly identified objects. It tells you how many of the model's detections were true positives (objects it identified correctly) and not false positives (objects it incorrectly identified).
- Recall: This metric focuses on the completeness of the detections. It tells you what percentage of actual objects in the image the model actually detected (true positives) and didn't miss (false negatives).

After investigating the YOLOv8 and YOLOv5 algorithms, the comparison of the performance metrics is as follows. Both models achieved real-time inference speeds, but YOLOv5 was faster, YOLOv8 reached mAP_50-95 of 0.483, and YOLOv5 reached mAP_50-95 of 0.447. The comparison of YOLOv5 and YOLOv8 performance metrics is shown in Table 2.

Table 2. Comparison of YOLOv5 and YOLOv8 performance metrics

YOLOv5	YOLOv8
Optimal Result: Epoch 64	Optimal Results: Epoch: 87
Number of Layers: 157	Number of Layers: 168
Precision (P): 0.901	Precision (P): 0.864
Recall (R): 0.931	Recall (R): 0.885
mAP50: 0.949	mAP50: 0.939
mAP50-95: 0.447	mAP50-95: 0.483

Based on Table 2, several conclusions regarding the performance comparison between YOLOv5 and YOLOv8 models can be drawn. Firstly, YOLOv8 features a more complex architecture, reflected in its higher number of layers (168) compared to YOLOv5 (157). Despite this, YOLOv5 achieves higher precision (0.901) compared to YOLOv8 (0.864), suggesting that YOLOv5 generates fewer false positives and is more effective at precise object detection.

It is also noteworthy that YOLOv5 surpasses YOLOv8 in terms of recall, achieving a value of 0.931 compared to 0.885 for YOLOv8. Higher recall indicates a better ability to detect actual objects, which is crucial in object detection tasks.

When analysing the mAP50 metric, YOLOv5 again achieves a slightly higher score (0.949) compared to YOLOv8 (0.939). The mAP50 metric refers to the mean Average Precision at an Intersection over a Union (IoU) threshold of 50%, which is significant for assessing model performance under moderately stringent matching criteria.

However, YOLOv8 shows an advantage in the mAP50-95 metric, achieving a value of 0.483 compared to 0.447 for YOLOv5. The mAP50-95 metric evaluates the Average Precision across various IoU thresholds, suggesting better overall model fitting and effectiveness in diverse detection scenarios for YOLOv8.

The comparison of YOLOv5 and YOLOv8 model performance, presented in Table 2, is also reflected in the graphs in Figures 8 and 9. A characteristic feature of these graphs is the rapid convergence of metrics in the first few dozen epochs, indicating the effectiveness of the training for both models at an early stage. Notably, YOLOv5 demonstrates its strengths in certain aspects of object detection, while YOLOv8 shows advantages in different areas, highlighting its adaptability across varying detection scenarios. These visual trends emphasize the distinct advantages each model offers, as outlined in the corresponding table and detailed analysis.

In summary, YOLOv5 is characterised by higher precision and recall, making it more effective in precise object detection. On the other hand, YOLOv8, despite slightly lower precision and recall, exhibits better results in the mAP50-95 metric, which may indicate its better overall fitting and adaptability in various detection situations. Examples of detection results using the YOLOv8 algorithm with various human identification certainties are shown in Figure 11.

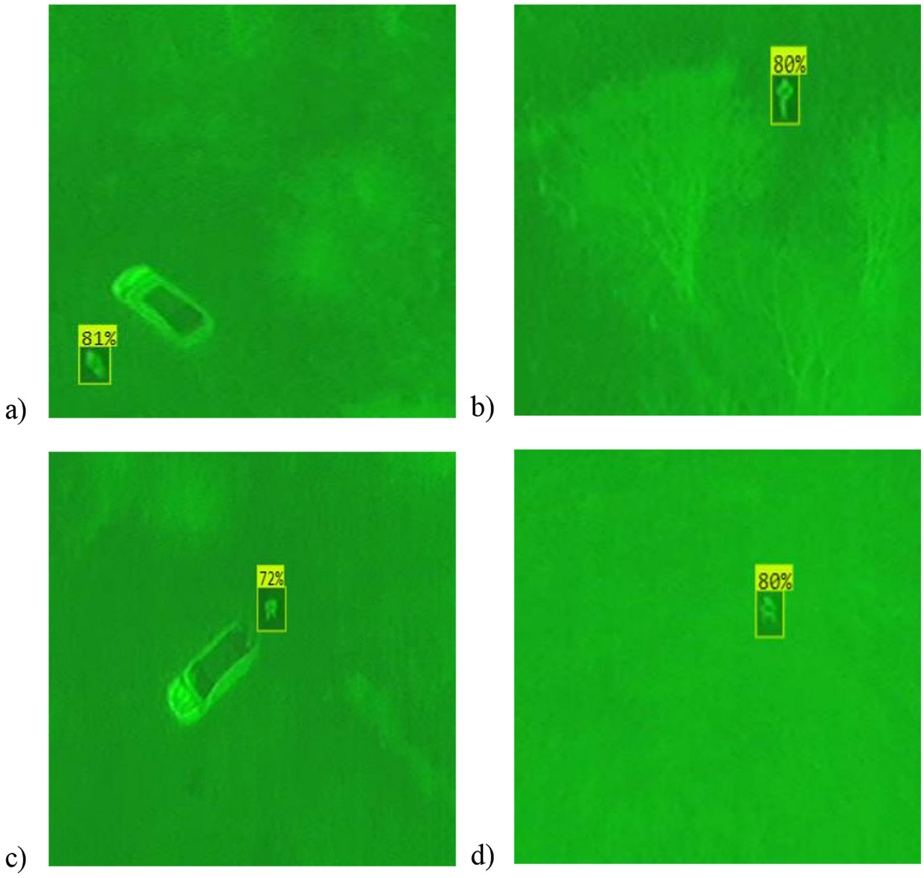

Figure 11. Thermal imaging with YOLOv8 algorithm detection, identifying a human at: a) 81% b) 80% c) 72% d) 80% confidence

CONCLUSIONS

UAVs are a technology in constant development. Platform parameters are being improved, and new components are being integrated into the platforms, which requires the cooperation of various technology providers to create the value chain. They are also a synergistic and complementary technology with other technologies, taking advantage of their capabilities, such as telecommunications and network capacity (5G), communications and satellite imaging. Thanks to them, drone technologies can improve their capabilities and efficiency, stimulate further development, and offer new, previously unknown solutions, contributing to the growth of the data-driven economy.

In the near future, technological development will undoubtedly be able to equip UAVs with an energy source, allowing for longer flights and operations in difficult weather conditions. Still, it will undoubtedly not replace manned flights. At the same time, it can provide significant support, especially considering elements such as the speed of inclusion in activities, services, aviation infrastructure, or the economic factor. The possibilities of using UAVs are basically limited only by technological conditions, both in terms of the devices themselves and the control systems. Developing an effective UAV control system - for example, based on head movement, eye movement, focusing on a selected point of goggles, and monocular - seems to be a challenge that is already within reach of current technological possibilities.

The article presents the possibilities of using artificial intelligence to construct a mobile security surveillance system. For this purpose, thermal imaging technology was used for remote sensing in combination with artificial intelligence. In this research work, the thermal YOLO object detection system was proposed as a smart human silhouette sensing system that should remain effective in all weather and harsh environmental conditions using an end-to-end YOLO deep learning framework. The system has been trained on large-scale thermal newly gathered novel datasets comprising more than 10,000 distinct thermal frames. The study further included deploying deep learning architecture on the edge and mobile devices, which can be interpreted as optimising a small network variant. In conclusion, selecting the most appropriate model depends on the application's specific requirements. The YOLOv8s model is preferable in scenarios demanding higher adaptability in various detection situations. In contrast, the YOLOv5s model may be more suitable for applications where detection accuracy and lower computational demands are paramount. After extensive testing and analysis, YOLOv5 was chosen as the optimal tool for the project. This decision was based on several key factors. YOLOv5 demonstrated better adaptability to the specific nature of thermal images, resulting in higher accuracy in detecting human silhouettes. Additionally, lower computational requirements and greater ease of implementation and customisation to the project's specific needs were significant factors influencing the final choice.

Due to the constantly growing demand for fast transmission of encrypted data (5G), the dynamically expanding world of sensors (IoT), the demand for the highest level of security on the Internet (Block-chain), the need for rapid data processing to improve the quality of transport and the development of new technologies, AI mechanisms in the processing of Big Data sets seem to be the future of unmanned systems development.

We should expect the use of drones to be increasingly common in civilian applications but also on the modern battlefield, especially the so-called drone swarms and the growing use of AI. This article shows the advantages and one of the practical applications of a VTOL aircraft. Another application for which the VTOL aircraft is being prepared is the configuration of a master drone transferring smaller, quieter drones over much longer distances and releasing them only above the targets. Targets may be precisely defined places where light points are dropped, the so-called beacons, marking an evacuation path for rescue services undertaking activities in demanding environmental conditions. The amount of data collected (especially image data) by a swarm of drones during a single mission will be so huge that it will exceed the operator's current analytical capabilities and the capacity of communication systems. Therefore, the implementation of technologies enabling their autonomous operation and the extensive use of decision support systems using AI will be crucial. Despite these promising results, it will still be interesting to experiment with other machine learning and training algorithms and different types of edge computing devices, including using the following versions of Yolo beyond v8 and their structure modifications to obtain better certainties.

Acknowledgements

The project was financed from state budget funds allocated by the Minister of Education and Science under the "Student scientific groups create innovations" program - grant no. SKN/SP/570644/2023. This work was also partially financed by a grant from SUT – a subsidy for maintaining and developing the research potential in 2024.

REFERENCES

1. Chui M., Manyika J., Miremadi M., Henke N., Chung R., Nel P., Malhotra S. Notes from the AI frontier insights from hundreds of use cases. Discussion paper, 2, McKinsey Global Institute, McKinsey&Company, April 2018.
2. Wieczorek S., Ludwig S., Büttner L., Bauer M.R., Markovič P. Forecast of changes in business administration driven by digitalization. open journal of business and management, July 2021; 9(4), https://doi.org/10.4236/ojbm.2021.94092
3. Conejero J.M., Brito I.S., Moreira A., Cunha J., Araújo J. Modeling the Impact of UAVs in Sustainability. In: Proceedings of the 24th International Requirements Engineering Conference Workshops (REW), Beijing, China, 12–16 September 2016. https://doi.org/10.1109/REW.2016.044
4. Cao Z., Kooistra L., Wang W., Guo L., Valente, J. Real-time object detection based on UAV remote sensing: A Systematic Literature Review. Drones 2023; 7(10), 620. https://doi.org/10.3390/drones7100620
5. Sandino J., Vanegas F., Maire F., Caccetta P. Sanderson C.; Gonzalez, F. UAV framework for autonomous onboard navigation and people/object detection in cluttered indoor environments. Remote Sens. 2020; 12, 3386. https://doi.org/10.3390/rs12203386
6. Cheng J., Zhang S. TBFNT3D: Two-Branch Fusion Network With Transformer for Multimodal Indoor 3D Object Detection. IEEE Robotics and Automation Letters 2023; 8(10), 6523–6530. https://doi.org/10.1109/LRA.2023.3309133
7. Samani E.U., Yang X. Banerjee A.G. Visual Object Recognition in Indoor Environments Using Topologically Persistent Features. IEEE Robotics and Automation Letters 2021; 6(4), 2377–3766. https://doi.org/10.1109/LRA.2021.3099460

8. Saif A.F.M.S., Prabuwono A.S., Mahayuddin Z.R. Real time vision based object detection from uav aerial images: a conceptual framework. In: Omar, K., et al. Intelligent Robotics Systems: Inspiring the NEXT. Communications in Computer and Information Science, 376. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-40409-2_23
9. Zhang Z., Liu Y., Liu T., Lin Z., Wang S. DAGN: a real-time uav remote sensing image vehicle detection framework. IEEE Geoscience and Remote Sensing Letters, 2019; 17(11), 1884–1888. https://doi.org/10.1109/LGRS.2019.2956513
10. Parker A., Gonzales F., Trotter P. Live detection of foreign object debris on runways detection using drones and AI. In: Proceedings of the 2022 IEEE Aerospace Conference (AERO), Big Sky, MT, USA, 05–12 March 2022. https://doi.org/10.1109/AERO53065.2022.9843697
11. Truong H. M., Clavel M. Identifying temporary water bodies from drone images at real-time using deep-learning techniques, In: Proceedings of the 2022 International Conference on Advanced Computing and Analytics (ACOMPA), Ho Chi Minh City, Vietnam, 21–23 November 2022, https://doi.org/10.1109/ACOMPA57018.2022.00009
12. Zhang Z., Uchiya T., Takumi I., Speed of autonomous drones in locating survivors after a disaster. In: Proceedings of the IEEE 10th Global Conference on Consumer Electronics (GCCE), Kyoto, Japan, 12–15 October 2021. https://doi.org/10.1109/GCCE53005.2021.9621953
13. Hoshino W., Seo J., Yamazaki Y. A study for detecting disaster victims using multi-copter drone with a thermographic camera and image object recognition by SSD. In: Proceedings of the 2021 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), Delft, Netherlands, 12–16 July 202. https://doi.org/10.1109/AIM46487.2021.9517524
14. Geng X., Peng R. Li M., Liu W., Jiang G., Jiang H., Luo J. A lightweight approach for passive human localization using an infrared thermal camera. IEEE Internet of Things Journal, 2022; 9(24), 24800–24811. https://doi.org/10.1109/JIOT.2022.3194714
15. Zhao X., Xiang M., He J., Huang Ch. Fire detection method in infrared image based on improved YOLO network, In: Proceedings of the 2021China Automation Congress (CAC), Beijing, China, 22–24 October 2021. https://doi.org/10.1109/CAC53003.2021.9728278
16. Munshi A.A. Fire detection methods based on various color spaces and gaussian mixture models. Advances in Science and Technology Research Journal 2021; 15(3), 197–214. https://doi.org/10.12913/22998624/138924
17. Shao Z., Liang Y., Tian F., Song S., Deng R. Constructing 3-D Land Surface Temperature Model of Local Coal Fires Using UAV Thermal Images. IEEE Transactions on Geoscience and Remote Sensing, 2022; 60, 5002309. https://doi.org/10.1109/TGRS.2022.3176854
18. Redmon J., Divvala S., Girshick R., Farhadi A. You Only Look Once: unified, real-time object detection. In: Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 27–30 June 2016. https://doi.org/10.1109/CVPR.2016.91
19. Liu K, Tang H., He S., Yu Q., Xiong Y., Wang N. Performance validation of yolo variants for object detection. In: Proceedings of the 2021 International Conference on Bioinformatics and Intelligent Computing (BIC '21). New York, USA, January 2021. https://doi.org/10.1145/3448748.344878
20. Wu H.-H., Zhou Z., Feng M., Yan Y, Xu H, Qian L. Real-time single object detection on the UAV. In: Proceedings of the 2019 International Conference on Unmanned Aircraft Systems (ICUAS), Atlanta, GA, USA, 11–14 June 2019. https://doi.org/10.1109/ICUAS.2019.8797866
21. Tijtgat N., Ranst W. V., Goedeme T., Volckaert B. De Turck F. Embedded real-time object detection for a UAV warning system. In: Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017. https://doi.org/10.1109/ICCVW.2017.247
22. Subrahmanyam V, Iowa State University ProQuest Dissertations & Theses, 2019. 13896596. https://www.proquest.com/openview/52ebcc2818e1a95e8ee890197e97762b/1
23. Wu Q., Zhou Y., Real-time object detection based on unmanned aerial vehicle. In: Proceedings of the 2019 IEEE 8th Data Driven Control and Learning Systems Conference (DDCLS), Dali, China, 24–27 May 2019. https://doi.org/10.1109/DDCLS.2019.8908984
24. Kim H.J., Shin M.C., Han M.W., Hong Ch., Lee H.W. An Efficient Scheme to Obtain Background Image in Video for YOLO-Based Static Object Recognition. Journal of Web Engineering, 2022; 21(5), 1691–1706. https://doi.org/10.13052/jwe1540-9589.21513
25. Zhang Q., Hu X. MSFFA-YOLO Network: Multiclass Object Detection for Traffic Investigations in Foggy Weather. IEEE Transactions on Instrumentation and Measurement, 2023; 72, 2528712. https://doi.org/10.1109/TIM.2023.3318671
26. Chen X., Peng D. L., Gu Y. Real-time object detection for UAV images based on improved YOLOv5s. Opto-Electron Eng, 2022; 49(3), 210372. https://doi.org/10.12086/oee.2022.210372
27. Carranza-García M., Torres-Mateo J., Lara-Benítez P., García-Gutiérrez J. On the performance of

one-stage and two-stage object detectors in autonomous vehicles using camera data. Remote Sensing. 2021; 13(1), 89. https://doi.org/10.3390/rs13010089

28. J. Lee, J. Wang, D. Crandall, S. Šabanović and G. Fox, Real-time, cloud-based object detection for unmanned aerial vehicles. In: Proceedings of the 2017 First IEEE International Conference on Robotic Computing (IRC), Taichung, Taiwan, 10–12 April 2017. https://doi.org/10.1109/IRC.2017.77
29. Thamm F.-P., Brieger N., Neitzke K.-P., Meyer M., Jansen R., Mönninghof M. Songbird – an innovative UAS combining the advantages of fixed wing and multi rotor UAS, In: Proceedings of the International Conference on Unmanned Aerial Vehicles in Geomatics, Toronto, Canada, 30 Aug–02 Sep 2015. https://doi.org/10.5194/isprsarchives-XL-1-W4-345-2015, 2015
30. Boon M. A., Drijfhout A. P., Tesfamichael, S. Comparison of a fixed-wing and multi-rotor UAV for environmental mapping applications: a case study, In: Proceedings of the International Conference on Unmanned Aerial Vehicles in Geomatics, Bonn, Germany, 4–7 September 2017. https://doi.org/10.5194/isprs-archives-XLII-2-W6-47-2017, 2017
31. Skóra J. The use of unmanned aerial vehicles in the context of ensuring security in the state. Aviation and Security, 2022; 1. https://doi.org/10.55676/asi.v1i1.7