DESIGN OF A SECURITY CONCEPT UTILIZING BVLOS FLIGHTS AND YOLO-BASED OBJECT DETECTION

JAN BACIK[1], LADISLAV HRIC[1], DANIELA PERDUKOVA[1]

[1]Dept. of Electrical Engineering and Mechatronics, Faculty of Electrical Engineering and Informatics, Technical University of Kosice, Kosice, Slovak Republic

DOI: 10.17973/MMSJ.2025_06_2025022

jan.bacik.2@tuke.sk

Recent shifts in Europe's security landscape driven by evolving hybrid warfare tactics require innovative solutions for real-time threat detection. This paper presents a security concept leveraging Beyond Visual Line Of Sight (BVLOS) drone flights combined with state-of-the-art YOLO-based object detection to enhance surveillance capabilities. By integrating advanced drone technology with AI-driven video analysis, the proposed system aims to detect armed individuals and mitigate potential hybrid attacks in critical infrastructure areas. A specialized dataset is prepared to improve detection accuracy and response times, ensuring cost-effective, autonomous monitoring in high-risk scenarios. This approach offers a robust solution for proactive perimeter security, optimizing resource allocation and reinforcing safety measures in an increasingly volatile environment. The results indicate substantial promise for future security implementations.

KEYWORDS

object detection, YOLO11, drone, security management, deep learning

1 INTRODUCTION

The security landscape in Europe has undergone a profound transformation since the outbreak of the war in Ukraine. Initially characterized by traditional military confrontations, the conflict has evolved into multifaceted hybrid warfare, blending conventional tactics with cyber operations and psychological strategies. One striking aspect of this evolution is the extensive use of both commercial and custom-built FPV drones, which have not only intensified battlefield dynamics but also contributed to a broader sense of insecurity across the continent [Pemcak 2024].

Beyond the immediate military implications, this new form of warfare has introduced unconventional threats targeting the public domain. European nations now face coordinated hybrid attacks—often delivered via menacing emails containing bomb threats or warnings of impending armed assaults on public institutions. Educational establishments, from primary schools to universities, have increasingly become targets, forcing authorities to implement disruptive measures such as evacuations, deployment of security forces, and mobilization of specialized response teams. These reactive strategies impose significant financial burdens on states and strain security resources, potentially leaving other critical areas vulnerable.

Amid these escalating challenges, rapid advancements in artificial intelligence offer promising avenues for enhancing public safety. AI-driven systems now enable the autonomous analysis of video footage, opening up new possibilities for early threat detection. In particular, state-of-the-art object detection algorithms—such as the YOLO (You Only Look Once) models from Ultralytics [Redmon 2016] —have emerged as industry benchmarks due to their balanced performance in speed and accuracy. The latest iteration, YOLO11, along with its variants (n, s, m, l, and x), provides a versatile framework suitable for a wide range of detection tasks. Leveraging these cutting-edge models, we are developing a system to detect armed individuals from drone-captured footage by training a model on custom dataset, thereby enhancing real-time detection capabilities during BVLOS (Beyond Visual Line Of Sight) operations and contributing to more effective, proactive security management. However, to fully realize this potential, such technological innovations must be implemented within a robust and evolving regulatory framework. As highlighted in [DAVIES 2018], the rapid advancements in UAV technologies and their expanding operational roles necessitate regular updates to regulatory frameworks. The European Union Aviation Safety Agency (EASA) has proactively developed comprehensive guidelines for BVLOS operations, including the STS-02 [EASA n.d.a] summary of BVLOS flight rules—integral to Regulation (EU) 2019/947—which establish clear operational, technical, and safety standards to harmonize UAV operations across European member states.

In parallel, the commercial drone sector, exemplified by the DJI Matrice 4 series and the DJI Dock 3 charging station introduced in 2025, has spurred significant legislative transformation. On May 22, 2024, Slovakia achieved a historic milestone with its first-ever BVLOS flight, during which Východoslovenska Energetika Holding, a.s. – Východoslovenska distribucna (VSD) successfully conducted an inspection of the traction network. This accomplishment was enabled by adherence to the PDRA – G03 [EASA n.d.b] guidelines, a detailed framework outlining performance data and risk assessment protocols tailored for complex BVLOS scenarios. While PDRA – G03 focuses on technical and risk management aspects, it effectively complements the broader operational guidelines defined by STS-02. Together, these frameworks form a synergistic legislative approach that ensures robust operational safety while fostering innovation in commercial drone technology.

2 RELATED WORKS

Object detection is a widely discussed topic due to its broad application in critical industries such as healthcare, security, manufacturing, and more. Nowadays, powerful GPUs and CPUs provide sufficient computational power for a wide range of machine-learning applications [Hortobagyi 2021, Kuric 2021 & 2022, Hu 2022].

The study [Zhang 2024] describes an innovative implementation of YOLO integrated with autonomous aircraft inspection using a drone. During the inspection, the drone flies around the aircraft and transmits video footage to a ground station where the image processing is performed. By handling the image analysis on the ground station instead of on the drone, the system eliminates the need for high onboard computational power. This automated approach can replace manual inspections carried out by a qualified technician, significantly reducing maintenance and repair costs during test flights and regular aircraft operations.

The study [Di 2023] provides a typical example of employing YOLO for the detection of electrical power lines using drones in hard-to-reach areas. In these settings, traditional manual inspections would incur high costs and pose elevated risks to the distribution company's personnel.

Study [Do 2023] represents a significant contribution by introducing a novel method for human detection using drone footage. The proposed approach achieves an impressive average accuracy of around 90.0% mAP@0.5 on the Human Drones dataset. Detecting humans from aerial imagery is particularly challenging, as the human figure often occupies only a small portion of the image while the remainder is filled with environmental noise. Despite these challenges, the method demonstrates robust performance, highlighting its potential for practical applications in surveillance and security.

From a public safety standpoint in public spaces and institutions, baggage screening is frequently implemented. In addition to traditional manual inspections, security personnel are increasingly exploring automated solutions to enhance efficiency and accuracy. The study [Kundilokovit 2024] focuses on the detection of hazardous objects in X-ray scans of baggage. The authors compare the performance of several models, including CNN, RCNN, Detectron, RetinaNet, and YOLO, to identify the most effective approach for automating threat detection.

Studies [Sumi 2024, Chitravanshi 2024, Pawar 2022] further demonstrate the suitability of YOLO models for hazardous object detection. In these investigations, YOLO-based approaches achieved exceptionally high detection accuracy for various dangerous weapons, including both firearms and pointed weapons. This impressive performance not only underscores the reliability of YOLO in complex security scenarios but also highlights its potential to significantly enhance threat detection capabilities. By providing rapid and precise identification of hazardous items, these models offer a promising solution for improving safety measures in environments where swift response is critical.

As mentioned in the introduction, adherence to applicable legislation, and its continuous adaptation to emerging challenges by relevant authorities, is an indispensable part of all BVLOS drone operations. Study [Lieb 2020] highlights a pan-European operational concept for UAS, centered on the UTM approach known as U- space. U- space is a digital ecosystem and regulatory framework designed to safely and efficiently integrate unmanned aerial systems into European airspace. It offers essential services such as flight planning, airspace monitoring, and real- time traffic management, thereby enabling seamless coordination between drones, manned aviation, and other airspace users. Through standardized procedures and robust technical support, U- space empowers both operators and regulators to manage BVLOS operations with enhanced safety and efficiency while fostering innovation in the rapidly evolving drone industry.

3 SECURITY CONCEPT DESIGN

In the context of our proposed concept, BVLOS operations are indispensable. Visual line of sight (VLOS) flights require the drone operator to maintain continuous visual contact, which can pose a significant security risk when monitoring areas with potentially armed intruders. In 2025, the Chinese company DJI introduced a commercially available solution for safe BVLOS operations, namely the Matrice 4 drone paired with the DJI Dock 3 ground station. This system not only supports autonomous, pre-programmed flight trajectories, reducing operator workload, but also integrates real-time streaming of footage. Together, these features provide an ideal tool for the automated monitoring of designated public spaces and the rapid detection of potential threats, ultimately contributing to more efficient and proactive security management.

Our concept is built on a commercially available DJI solution, extended with a computer powered by NVIDIA Jetson AGX Orin for artificial intelligence (Fig. 1). The NVIDIA Jetson AGX Orin 64GB delivers up to 275 TOPS of computational performance while consuming only 60W, making it exceptionally energy efficient. This powerful edge computing platform is engineered to handle demanding AI applications in real time, enabling the rapid processing of complex neural network and computer vision tasks. Its robust GPU architecture, substantial memory capacity, and optimized power consumption make it an ideal component for enhancing our system's automated threat detection capabilities.

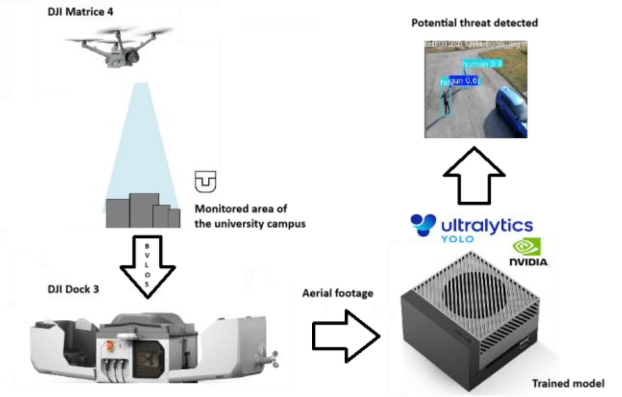

Figure 1. Security concept with trained model to detect potential thread

A key element of the proposed concept is the development of a robust model for detecting potential threats. This process involves creating a sufficiently large dataset, accurately labelling individual objects, and defining the relevant object classes before training a neural network. In the subsequent chapter, we will delve into these steps in greater detail, outlining the methodologies and best practices for each phase of the model development process.

4 DATA COLLECTION

All datasets begin with the collection of images or videos that can be used for further annotation. It is important to consider the necessity of diverse images capturing the target object in different conditions, such as various angles, distances, background variations, and lighting conditions. The quality of the prepared dataset directly affects the quality of the output model. Creating a dataset is one of the most important steps in such a project and will therefore be discussed in this chapter.

Our safety system aims to detect weapons and people to identify potential threats. There are many official and widely used datasets that could be utilized. For instance, the large-scale COCO dataset [Lin 2014] includes 328k images with 91 object categories. However, it is not specialized for aerial detection. On the other hand, one of the most well-known datasets for drone-based applications is the VisDrone dataset [Zhu 2021], which consists of manually annotated images, including over 2.5 million bounding boxes of pedestrians, cars, bicycles, tricycles, and more.

However, none of the available and verified datasets include labelled aerial images of armed individuals. Considering these limitations, as well as the importance of having direct control over the dataset, its modifications, and the nature of the discussed topic, it is suitable to build a dedicated dataset.

None of the well-known datasets were perfect on the first attempt, but they provide valuable insights into the problem and help guide further data collection. In our case, we started with outdoor scenes, where we positioned three different DJI Mini drones at heights ranging from 2 to 5 meters, varying their distances from the location where actors moved with a gas gun. Another set of images was captured using DJI Mini drones in indoor scenarios. After the first sets of video shooting, we collected around 500 images of people, both armed and unarmed.

During data collection, we prioritized safety and took all necessary precautions to avoid causing unintentional panic among bystanders. While walking around the streets with a visible gun would have been a valuable source of data, it would likely cause panic and is not even legal, as such gas guns must be covered in public. This is one of the limitations to consider before collecting footage.

5 IMAGE PRE-PROCESSING

Image pre-processing involves manipulating images and transforming them into a standardized format. This step directly impacts model precision and computational efficiency, and the choice of specific pre-processing techniques should be guided by the application's requirements. In this study, we applied auto-orientation to standardize image orientation based on EXIF metadata. Omitting this step could result in a mismatch between image and bounding box orientations, leading to incorrect model training and inference.
A common practice in deep learning is to start training with lower-resolution images to evaluate model accuracy and performance efficiently. Smaller images reduce training time and speed up inference. However, if the application demands higher resolution, it is advisable to progressively increase image size. In this work, we employed resizing as the next pre-processing step while maintaining the original 16:9 aspect ratio. We reduced the resolution from 1920×1080 to 960×540, ensuring a balance between computational efficiency and model performance.

6 DATA AUGMENTATION

The augmentation process helps generate new images for the dataset by modifying the original ones. In this chapter, the augmentation steps used in our study will be discussed.
To improve the model's ability to recognize objects from different perspectives, we first applied horizontal flipping. We also rotated the images clockwise, counter-clockwise, and within a range of -15° to +15° as shown on Fig.2. Additionally, to cover more perspectives, we applied shear transformations of up to 5° both horizontally and vertically, which helps the model become more resilient to variations in an object's pitch and yaw.

(a) (b) (c)


Figure 2. Rotation augmentation: (a) original image (b) rotated by -15° (c) rotated by +15°

The steps mentioned above help expand the dataset by incorporating objects in various positions, but ensuring diversity in lighting conditions is equally important. Therefore, as the next augmentation step, we created a new subset of images by increasing and decreasing brightness by 25%. Another subset was created by modifying saturation within a range of -30% to +30% and adjusting exposure between -15% and +15% as shown on Fig. 3.

(a) (b) (c)


Figure 2. Exposure adjustment: (a) original image (b) exposure decreased by 15% (c) exposure increased by 15%

Next, we randomly altered image colours by varying the hue between -25° and +25°. Following this, we artificially added noise to the images by converting 2% of the pixels to either black or white (Fig. 4), which helps the model become more resilient to camera artifacts.

(a)
(b)


Figure 4. Noise addition: (a) original image (b) 1.85% of pixels randomly converted to black or white

Finally, as the last augmentation step, we applied mosaic augmentation. This technique helps the model recognize targets in different locations and enhances its robustness to varying surroundings.
In the end, after applying all augmentation techniques, our dataset grew to approximately 1200 pictures.

7 MODEL TRAINING

In this study, we trained models using the state-of-the-art YOLO 11 detector and assessed their performance. To better understand how different YOLO variants perform on our custom dataset, we focused on training the n, m, and x versions. All models were trained with a batch size of 16 for 200 epochs. The models were trained on Google Colab utilizing NVIDIA A100-SXM4-40GB GPUs, ensuring high computational efficiency. Based on the results, we identified the most effective model for further optimization. This assessment helps determine whether YOLO 11-based models are well-suited for our use case. The first metric used is Precision, which indicates the model's ability to reduce false positive detections, ensuring that identified objects are truly present in the image. Precision is defined as:

$$Precision = \frac{TP}{TP+FP} \tag{1}$$

where TP (True Positives) represents the number of correctly detected objects—meaning the model identified objects that exist in the image. FP (False Positives) accounts for cases where the model mistakenly detects objects that are not actually present.
Another crucial evaluation metric is Recall, which assesses the model's ability to correctly identify all relevant objects, minimizing instances of missed detections. Recall is calculated as:

$$Recall = \frac{TP}{TP+FN} \tag{2}$$

where FN (False Negatives) refers to cases where the model fails to detect an object that is actually present in the image.
The final performance metric considered in this study is mAP50-95, which provides a comprehensive evaluation of the model by analysing its precision and recall across various Intersection over Union (IoU) thresholds, ranging from 50% to 95%.

8 RESULTS

As mentioned earlier, we trained three YOLO model variants: n, m, and x. As shown in Table 1, the fastest model is the n variant, achieving an inference speed of 0.7 ms with an mAP50-95 of 0.603. This model would be the optimal choice for applications where speed is the top priority. On the other hand, the m variant demonstrated the highest mAP50-95, reaching 0.675, outperforming the x variant in terms of precision. Based on these results, the m variant appears to be the most suitable for applications prioritizing accuracy, while still maintaining relatively fast inference speed.

Table 1. Performance comparison of models using different YOLO11 variants

Variant	All classes			
	Inference speed	Precision	Recall	mAP50-95
n	0.7 ms	0.864	0.756	0.603
m	2.4 ms	0.858	0.806	0.675
x	4.9 ms	0.879	0.799	0.645

For the proposed safety concept, the m variant was identified as the most suitable model, offering the highest precision while maintaining a good inference speed. Fig 5. illustrates the training process, showing the changes in precision, recall, and mAP50-95 over 200 epochs. Precision exhibited significant oscillations during the initial epochs, followed by a gradual decline throughout the training process. Recall stabilized around 0.75, with fluctuations, while the highest recorded value in the best epoch reached 0.806. The mAP50-95 metric steadily increased over the entire training period, peaking at 0.675, underscoring the model's ability to provide accurate detections across varying Intersection over Union (IoU) thresholds.

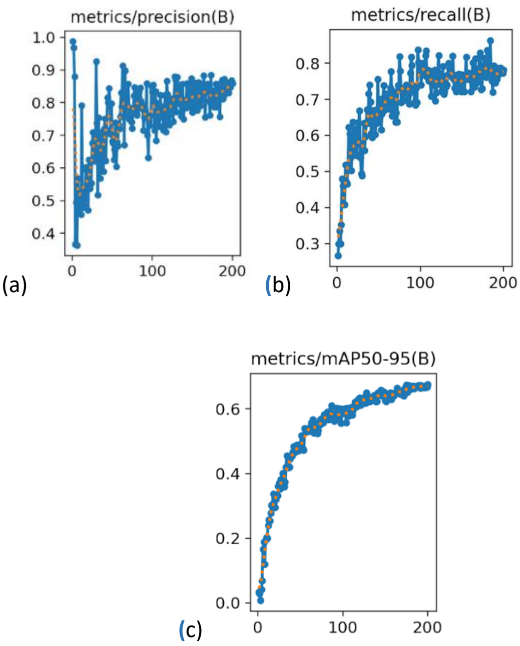

Figure 5. Training progress: (a) Precision (b) Recall (c) mAP50-95

Even though the model achieved satisfactory results, further improvements remain a key objective. As shown in Figure 6, the model demonstrated high confidence in detecting humans, whereas confidence in detecting guns was noticeably lower. This discrepancy may be attributed to an imbalanced dataset, a lower number of gun instances, or insufficient representation of real-world conditions. Additionally, object size poses a challenge, as models generally struggle with detecting smaller objects.

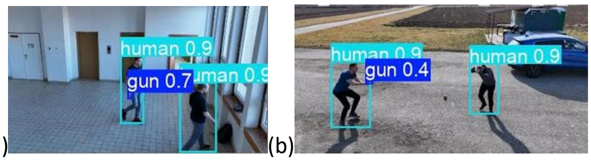

Figure 6. Model Inference: (a) Indoor scenario (b) Outdoor scenario

9 CONCLUSION AND FUTURE WORK

The integration of drones and computer vision for safety applications has gained significant interest in recent years, with further advancements on the horizon. As demonstrated in this study, YOLO models have proven to be highly effective and well-suited for these applications. Successful object detection relies on comprehensive data collection, accurate annotation, and careful dataset preparation, ensuring that the dataset is as diverse and representative of real-world scenarios as possible. In this work, we developed a custom dataset specifically for a security-oriented concept leveraging BVLOS (Beyond Visual Line of Sight) drone operations, covering both indoor and outdoor environments. To enhance diversity, various data augmentation techniques were applied. As discussed in the Results section, the most suitable model for our concept was the n variant of the YOLO model, achieving a precision of 0.858, recall of 0.806, and an mAP50-95 of 0.675. These results are promising, particularly considering the inference speed of 2.4 ms and the potential for further dataset improvements and optimized training configurations. Future research should focus on expanding the dataset with additional real-world scenarios; however, capturing such data presents challenges, as recording footage in public areas is legally restricted, particularly when firearms are involved. Additionally, future work should explore advanced augmentation techniques and alternative training setups to further enhance model efficiency. While developing a robust safety application poses numerous challenges and limitations, ensuring human safety remains a top priority, making these efforts worthwhile.

ACKNOWLEDGMENTS

This work was supported by the Slovak Research and Development Agency under the Contract no. APVV-19-0210 and project VEGA 1/0363/23.

This work was also funded by the EU's NextGenerationEU programme as part of the Recovery and Resilience Plan for Slovakia, under project no. 09I05-03-V02-00018.

REFERENCES

[Chitravanshi 2024] Chitravanshi, D., et al. Weapon Detection from images using YOLO and OpenCV. In: 1st Int. Conf. on Technological Innovations and Advance Computing (TIACOMP 2024), Bali, Indonesia, 2024, pp. 560-565. DOI: 10.1109/TIACOMP64125.2024.00098.

[Davies 2018] Davies, L., et al. Review of Unmanned Aircraft System Technologies to Enable Beyond Visual Line of Sight (BVLOS) Operations. In: 2018 X International Conference on Electrical Power Drive Systems (ICEPDS), Novocherkassk, Russia, 2018, pp. 1–6, doi: 10.1109/ICEPDS.2018.8571665.

[Di 2023] Di, T., et al. Research on Real-Time Power Line Damage Detection Method Based on YOLO Algorithm. In: IEEE 3rd Int. Conf. on Electronic Technology, Communication and Information (ICETCI 2023);

Changchun, China, 2023, pp. 671-676. DOI: 10.1109/ICETCI57876.2023.10176923.

[Do 2023] Do, M.-T., et al. Human Detection Based Yolo Backbones-Transformer in UAVs. In: Int. Conf. on System Science and Engineering (ICSSE); Ho Chi Minh, Vietnam, 2023, pp. 576-580. DOI: 10.1109/ICSSE58758.2023.10227141.

[EASA n.d.a] EASA. Standard Scenario – STS. EASA, n.d. [online]. [Accessed March 9, 2025]. Available from: https://www.easa.europa.eu/en/domains/drones-air-mobility/operating-drone/specific-category-civil-drones/standard-scenario-sts.

[EASA n.d.b] EASA. Predefined Risk Assessment (PDRA). EASA, n.d. [online]. [Accessed March 9, 2025]. Available from: https://www.easa.europa.eu/en/domains/drones-air-mobility/operating-drone/specific-category-civil-drones/predefined-risk-assessment-pdra.

[Hortobagyi 2021] Hortobagyi, A., Pivarciova, E., Koleda, P. Holographic Interferometry for Measuring the Effect of Thermal Modification on Wood Thermal Properties. Applied Sciences, 2021, Vol. 11, No. 6, 2516. DOI: 10.3390/app11062516.

[Hu 2022] Hu, J., et al. Autonomous dynamic line-scan continuous-wave terahertz non-destructive inspection system combined with unsupervised exposure fusion. NDT & E International, 2022, Vol. 132, 102705. DOI: 10.1016/j.ndteint.2022.102705.

[Kundilokovit 2024] Kundilokovit, P., et al. Weapon Detection in X-ray Image of Baggages. In: 23rd Int. Symp. on Communications and Information Technologies (ISCIT 2024); Bangkok, Thailand, 2024, pp. 88-93. DOI: 10.1109/ISCIT63075.2024.10793527.

[Kuric 2021] Kuric, I., et al. Analysis of the Possibilities of Tire-Defect Inspection Based on Unsupervised Learning and Deep Learning. Sensors, 2021, Vol. 21, No. 21, 7073. DOI: 10.3390/s21217073.

[Kuric 2022] Kuric, I., et al. Implementation of Predictive Models in Industrial Machines with Proposed Automatic Adaptation Algorithm. Applied Sciences, 2022, Vol. 12, No. 4, 1853. DOI: 10.3390/app12041853.

[Lieb 2020] Lieb, J., et al. Unmanned Aircraft Systems Traffic Management: A comparsion on the FAA UTM and the European CORUS ConOps based on U-space. In: AIAA/IEEE 39th Digital Avionics Systems Conference (DASC), San Antonio, TX, USA, 2020, pp. 1-6. doi: 10.1109/DASC50938.2020.9256745.

[Lin 2014] Lin, T.Y., et al. Microsoft COCO: Common Objects in Context. Lecture Notes in Computer Science, Vol. 8693. Springer: Cham, 2014. DOI: 10.1007/978-3-319-10602-1_48.

[Pawar 2022] Pawar, S., et al. Fire and Gun Detection Using YOLO v3 For Surveillance System. In: Int. Conf. on Artificial Intelligence and Data Engineering (AIDE 2022); Karkala, India, 2022, pp. 196-201. DOI: 10.1109/AIDE57180.2022.10060055.

[Pemcak 2024] Pemcak, I., et al. Theoretical Vulnerability Analysis of Individual Components of Class I – Micro Drone Against Small Arms Fire. Problems of Mechatronics Armament Aviation Safety Engineering, 2024, 15(4), pp. 9–26. https://doi.org/10.5604/01.3001.0054.9122.

[Redmon 2016] Redmon, J., Divvala, S., Girshick, R., Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In: Proc. of IEEE Conf. on Computer Vision and Pattern Recognition, 2016, pp. 779-788. DOI: 10.1109/CVPR.2016.91.

[Sumi 2024] Sumi, L., et al. A Deep Learning Approach to Gun Detection in Surveillance System using YOLOv9. In: Control Instrumentation System Conference (CISCON 2024); Manipal, India, 2024, pp. 1-5. DOI: 10.1109/CISCON62171.2024.10696152.

[Zhang 2024] Zhang, D., et al. Research on Aircraft Patrol Inspection Method Using UAV Based on YOLO Algorithm. In: 4th Int. Conf. on Electronic Information Engineering and Computer (EIECT 2024); Shenzhen, China, 2024, pp. 412-415. DOI: 10.1109/EIECT64462.2024.10866925.

[Zhu 2021] Zhu, P., et al. Detection and Tracking Meet Drones Challenge. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, Vol. 44, No. 11, pp. 7380-7399.

CONTACTS:

Jan Bacik, Ing. PhD.; Ladislav Hric, Ing.
Daniela Perdukova, Prof. Ing. PhD.
Technical university of Kosice, Faculty of Electrical Engineering and Informatics, Dept. of Electrotechnics and Mechatronics
Letna 9, Kosice, 042 00, Slovak Republic
+421 55 602 2251, jan.bacik.2@tuke.sk; ladislav.hric@tuke.sk
+421 55 602 2256, daniela.perdukova@tuke.sk