HADMÉRNÖK
VÉDELEMINFORMATIKA
20. évfolyam (2025) 3. szám 71–83.
DOI: 10.32567/hm.2025.3.4

Attila Hammas[1]

More Data, More Problems?

Exploring the Impacts of Using External Datasets when Training Deep Learning Models for Remote Sensing Applications

Abstract

The resurgence of large-scale conventional warfare, exemplified by the ongoing conflict in Ukraine, has highlighted the critical importance of modern technologies in enhancing situational awareness and target acquisition on the battlefield. In particular, the integration of unmanned aerial vehicles (UAVs) and artificial intelligence (AI)-driven computer vision systems has emerged as a key enabler of real-time intelligence and precision engagement. This paper presents an approach to enhance object detection models on aerial imagery for military applications. Initial experiments revealed shortcomings in detecting certain object classes, particularly in complex environments and under variable lighting conditions. To address these issues, the paper investigates impacts of cross-dataset training to improve the robustness and accuracy of object detection models. Through selective label integration and careful dataset curation, the paper demonstrates that incorporating assets from external sources significantly enhances generalisation and detection performance. The results underline the potential of leveraging large-scale annotated datasets to augment domain-specific applications with minimal additional labelling cost.

Keywords: reconnaissance, IMINT, artificial intelligence, machine learning

Introduction

The full-scale invasion of Ukraine by Russia in 2022 sent shockwaves across the international community, not only due to its humanitarian and geopolitical implications

[1] PhD student, Ludovika University of Public Service Doctoral School of Military Engineering, e-mail: hammas.attila@stud.uni-nke.hu

but also because it signified a dramatic resurgence of high-intensity, conventional warfare between sovereign states – something Europe had not witnessed on such a scale since World War II. For decades, armed conflicts in the post-Cold War era were predominantly characterised by asymmetrical warfare, insurgencies and proxy battles. The re-emergence of large-scale, state-versus-state conflict – complete with tanks, artillery and mass mobilisations – was seen as an anachronism by many, a relic of 20th-century military history that had no place in modern geopolitics. Yet, the events in Ukraine challenged that assumption. This war has prompted a re-evaluation of European defence strategies, NATO readiness and global power dynamics. It has also raised urgent questions about the role of deterrence, the importance of military alliances and the stability of the international rules-based order. As such, the invasion is not just a regional crisis, it is a pivotal moment in global history, redefining how war is perceived and prepared for in the 21st century.

In addition to its geopolitical and strategic consequences, the war in Ukraine has also served as a proving ground for a new era of military technology and innovation. One of the most striking features of the conflict has been the widespread use of drones – not only by state actors but also by irregular forces – ranging from inexpensive commercial quadcopters used for reconnaissance and targeting, to advanced loitering munitions capable of precision strikes. The fusion of Intelligence, Surveillance and Reconnaissance (ISR) capabilities has dramatically increased situational awareness on both sides, with real-time battlefield intelligence. Moreover, artificial intelligence has begun to play a significant role, from the automatisation of image analysis and the identification of targets to enhancing electromagnetic warfare systems and optimising logistics. This blending of traditional combat with digital innovation has underscored a new model of hybrid warfare, where software, data and autonomous systems can be as decisive as tanks and artillery. The Ukraine conflict is not just a return to conventional war; it is a glimpse into the future of how wars will be fought from now on.

ISTAR (Intelligence, Surveillance, Target Acquisition and Reconnaissance) stands at the forefront of this technological evolution, serving as a critical military capability that integrates diverse intelligence streams and sensor inputs to generate comprehensive situational awareness, enabling faster and more informed decision-making.[2] In the context of modern warfare, ISTAR capabilities have become increasingly intertwined with disruptive technologies such as drones, autonomous weapon systems and artificial intelligence. These advancements hold the promise of not only strengthening operational effectiveness but also significantly reducing risks for human personnel.[3] By leveraging autonomous systems to perform dangerous reconnaissance missions or precision strikes, human soldiers are kept at safer distances from frontline threats. AI-driven analysis of vast datasets can rapidly identify threats, improve targeting accuracy and reduce collateral damage.

During the progress of research and development of a platform for processing and analysing data provided by Unmanned Aerial Vehicles (UAV) to intensify situational awareness and target acquisition capabilities, notable shortcomings in object

[2] Boda 2024.
[3] Balogh 2012.

detection performance were identified, particularly regarding specific object classes. This study investigates whether the integration of external datasets improves accuracy and robustness of the employed object detection models. Accordingly, the research addresses whether supplementing synthetic and limited real-world data with large-scale external datasets can enhance accuracy and generalisation in UAV-based ISTAR applications. It further examines the extent to which cross-dataset training contributes to improved performance. In the following sections, the impact of incorporating annotated data from additional sources on model performance is evaluated.

Theoretical overview

Machine learning, particularly through Convolutional Neural Networks (CNN), has revolutionised computer vision by enabling machines to recognise patterns in images. CNNs utilise multiple layers, including convolution, pooling and fully connected layers, to process and classify images. Deeper and more complex CNN architectures require significantly more computational resources for training and inference.[4] Deep networks are also prone to overfitting, especially when trained on small datasets, this issue can be addressed using techniques such as dropout and data augmentation.[5]

In the field of computer vision, several important applications can be identified (Figure 1):

- *Image classification:* CNNs have achieved state-of-the-art results in image classification tasks by effectively learning hierarchical representations of images.[6]
- *Object detection, recognition and tracking:* CNNs are widely used in object detection frameworks such as YOLO (You Only Look Once)[7] and Faster R-CNN (Region based Convolutional Neural Network),[8] which can identify and localise objects within images.[9]
- *Image segmentation:* CNNs divide an image into multiple segments, often with pixel-level precision, allowing computer vision systems to more accurately detect and interpret objects by outlining their exact shapes rather than using simple bounding boxes.
- Other applications include video analysis, face recognition and human pose estimation, showcasing the versatility of CNNs in handling various visual tasks.[10]

While CNNs have become a cornerstone of computer vision, they are not without limitations. The complexity of CNNs can lead to high computational costs, making them less accessible for applications with limited resources. Additionally, the reliance on large, labelled datasets for training poses challenges in domains where such data is scarce.

4 Van Doorn 2014; Zhao et al. 2024.
5 Zhao et al. 2024.
6 Shetty et al. 2022; Zhao et al. 2024.
7 Redmon et al. 2015; Redmon–Farhadi 2016; 2018; Bochkovskiy et al. 2020.
8 Ren et al. 2017.
9 Shetty et al. 2022; Rana–Chauhan 2021.
10 Shetty et al. 2022; Voulodimos et al. 2018.

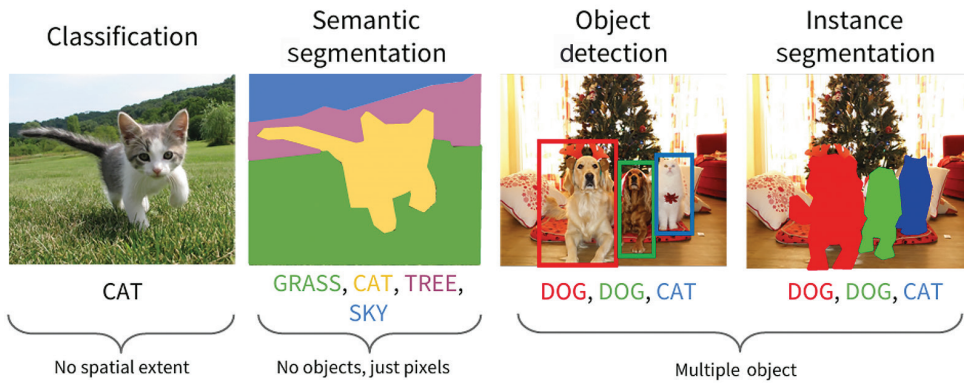

Figure 1: Visualisation of different computer vision tasks
Source: Li–Adeli 2024: 14.

The size and quality of a training dataset are pivotal factors that significantly influence the performance of machine learning models. The interplay between these two elements determines the model's ability to generalise its accuracy and its efficiency. While larger datasets generally provide more learning opportunities, the quality of the data is often more critical in ensuring reliable deep learning model outputs. The scarcity of data is a core issue in training Machine Learning (ML) application. Big preexisting datasets like MS COCO (Microsoft Common Objects in Context)[11] and PASCAL VOC (Pascal Visual Object Classes)[12] contain a wide range of object classes and scenarios, which can significantly boost the diversity of training data. This diversity helps in improving the generalisation ability of models, makes them more robust to different environments and conditions. In those cases when the label provided by these large datasets do not exactly cover the required label set, cross-dataset training can be applied, allowing models to learn from a union or an intersection of different object classes without the need for exhaustive labelling across all datasets.[13] This approach is efficient for industrial applications where object classes frequently increase. By leveraging the work done for the already existing datasets, we can reduce the costs of labelling.

Methodology

The original dataset consisted of synthetic images generated using Unreal Engine based AirSim simulator complemented with real-world images collected by using drones and downloaded from the internet. Based on the analysis of our previous results, our trained models fail to generalise enough, and they underperform in detecting people not in uniforms or certain cars. As shown in Figure 2, the model can detect soldiers in uniforms but struggles detecting people in civilian clothing. The synthetic datasets generated using the simulator and other 3D game engines are designed to emphasise objects of military relevance.[14]

[11] Lin et al. 2014.
[12] Everingham et al. 2010.
[13] Yao et al. 2020.
[14] Hammas 2023; Brassai et al. 2024.

Figure 2: Example outputs of our models before including external datasets
Note: Left: soldiers in uniforms; right: people in civilian clothing.
Source: compiled by the author

To extend the dataset while minimising acquiring and labelling costs, preexisting datasets containing people and different vehicle categories were employed:

- *Microsoft Common Objects in Context:*[15] Includes images from everyday scenes with objects in *natural and complex contexts,* helping models generalise better to real-world environments compared to synthetic or simplified datasets. It contains over 200 thousand images with 80 different labels. As illustrated in Figure 3, the dataset includes people in various poses, which can help improve robustness when they are crouching or crawling on frames.

Figure 3: Example images from MS COCO dataset
Source: Lin et al. 2014

[15] Lin et al. 2014.

- *VisDrone:*[16] Contains drone footage in urban areas with complex scenes, varying light conditions and perspectives specifically tailored for object detection tasks in aerial imagery as can be seen in Figure 4. The dataset presents video sequences and still images predominantly from urban environments, which are characterised by their high level of scene complexity. It includes a diverse range of labelled object categories such as pedestrian, people, bicycle, car, van, truck, tricycle, bus and motorcycle.

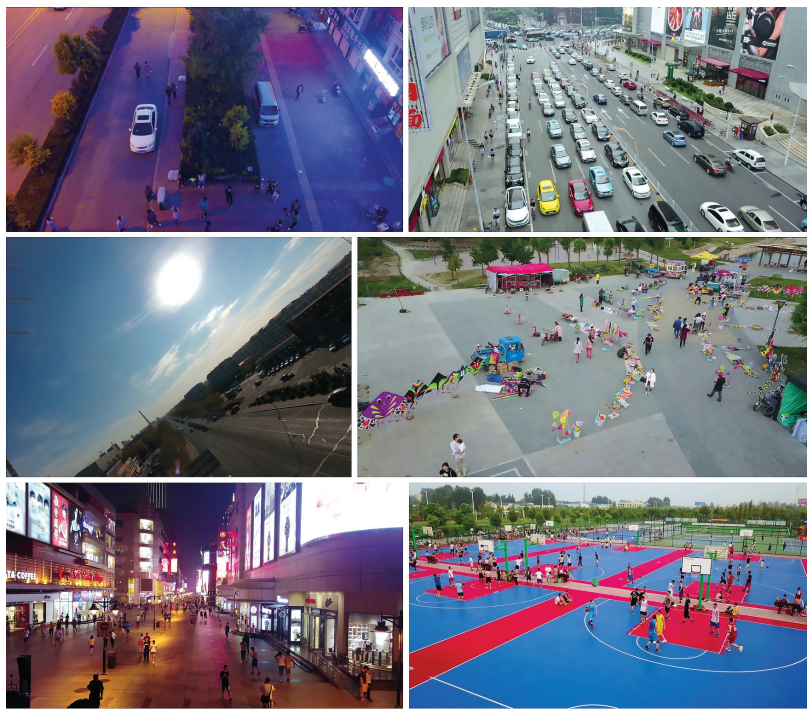

Figure 4: Example images from VisDrone dataset
Source: Zhu et al. 2018

These datasets also contain irrelevant object classes (e.g. food items, toys, household objects). These labels were removed from the images which resulted in many images remaining without any annotation.

More datasets were considered, then excluded due to conflicts:

- *DOTA (Dataset for Object Detection in Aerial Images):*[17] Provides high-resolution aerial imagery with objects annotated in arbitrary orientations. It mainly uses rotated bounding boxes but it is also available with horizontal bounding boxes. It does not include people since it was created for object detection on satellite imagery.

[16] Zhu et al. 2018.
[17] Xia et al. 2018.

- *UAVDT (Unmanned Aerial Vehicle Detection and Tracking):*[18] It is very similar to VisDrone, but with emphasis on vehicle detection and tracking. It conflicts with our classes since the images contain people, but they are not annotated. If the dataset is combined with VisDrone without further annotation it may reduce the recall and accuracy of detecting people.

The classes selected from the MS COCO and VisDrone datasets were carefully adapted to align with the objectives of this study, which focus on improving situational awareness through the reliable detection of humans and vehicles relevant to UAV-based ISTAR operations. After filtering out irrelevant categories, the following classes were retained and harmonised across the combined dataset: (index: class name):

- 0: Human
- 1: Car
- 2: Truck
- 3: Bus
- 5: Train
- 31: Unknown Aircraft

The rules for adapting the classes found in this external dataset were as follows:

Table 1: The rules for transferring labels from external dataset into the combined dataset

Dataset	Class name in external dataset	Class name in combined dataset
MS COCO	Person	Human
	Car	Car
	Bus	Bus
	Train	Train
	Truck	Truck
	Other classes were removed.	
VisDrone	Pedestrian	Human
	People	Human
	Car	Car
	Van	Car
	Truck	Truck
	Bus	Bus
	Other classes were removed.	

Source: compiled by the author

For testing purposes, the Ultralytics[19] framework was used for training and evaluation. We trained both models for 240 epochs with the following parameters: learning rate: 0.01 (with 3 warmup epochs – the value was gradually increased in the first 3 epochs), momentum: 0.9, and with an input image size of 1280 × 1280.

[18] Du et al. 2018.
[19] Ultralytics s. a.

Results and discussion

Due to the size of the selected datasets the final combined dataset is unbalanced. This is a limitation of this research. The number of instances is orders of magnitude larger compared to the other classes which are not in the selected external datasets. As shown in Figure 5, the Human (0) and Car (1) classes are overrepresented in this dataset. This issue is not addressed in this research but as can be seen in Figure 7 it does not impact precision and recall.

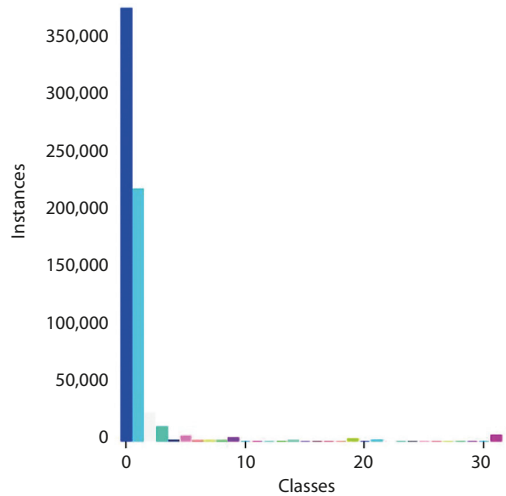

Figure 5: Distribution of classes in the training dataset
Source: compiled by the author

As was emphasised in the introduction, the model trained solely on our internal dataset performed poorly when detecting people in civilian clothing or civilian vehicles (e.g. cars, trucks). The extended dataset samples from the validation and testing subset of the selected external datasets were added in unchanged form. This is also evident in Figure 6, where the model's precision is substantially lower.

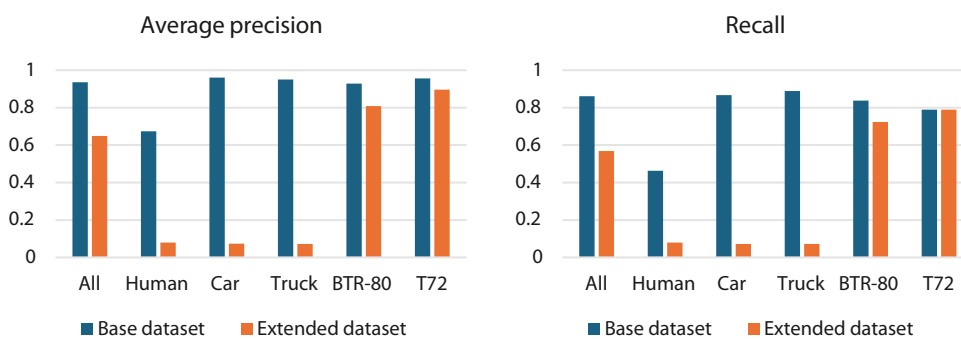

Figure 6: Comparison of the initially trained model on the base dataset and validated on the extended dataset
Source: compiled by the author

Compared to the models trained in the original base dataset, the newly trained models show significant improvements, as shown in Figure 7. The precision and recall values are not negatively impacted by the imbalance between the classes of the dataset.

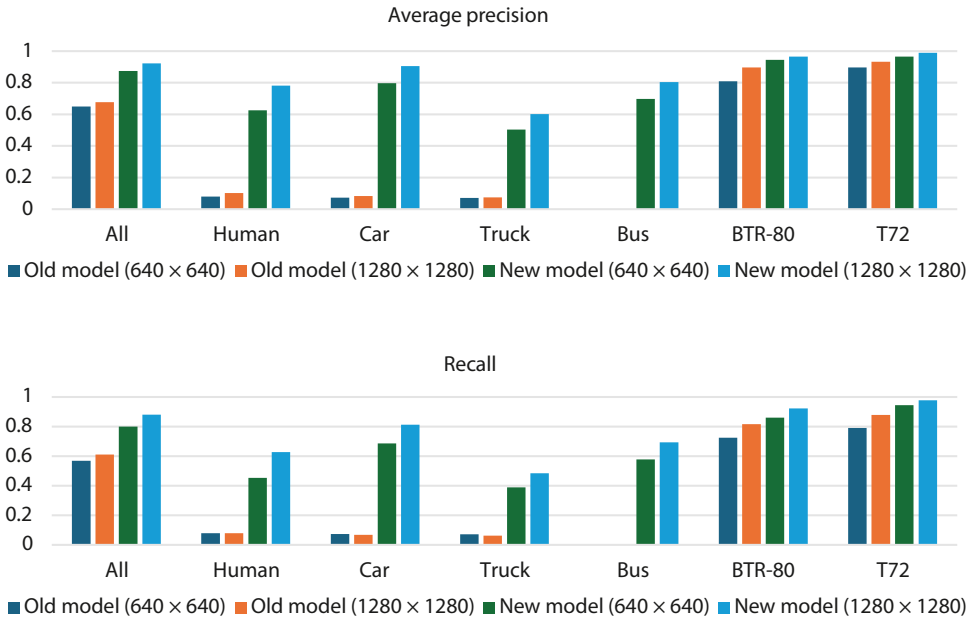

Figure 7: Comparison of the base model trained on the initial dataset and the new model trained on the extended dataset with multiple input resolutions (640 × 640 and 1280 × 1280)

Source: compiled by the author

In some rare cases, the newly trained model detected and identified object instances which were not labelled by the authors of the dataset, as shown on Figure 8. In the example on Figure 8, the image had no relevant labels but was left in as a negative example. The phone class were removed since they do not align with the objectives of this work (as explained in the methodology section). Surprisingly, the object detector still managed to identify people in the image. As shown in the left-hand image of Figure 8, the screen displays pictures of people.

Figure 8: One of the images in the validation set with desired output on the top and predicted output on the bottom

Source: compiled by the author based on image found in the dataset presented by LIN et al. 2014

The image used for example comes from the validation subset of the MS COCO dataset, where phones were labelled from various angles, but the people visible on the screen were not annotated. Thanks to the inclusion of numerous small-scale human instances in both the VisDrone and our internal datasets, the model was able to learn to detect these as well.

Another issue identified involves false positive detections, particularly during takeoff, landing, or when the drone encounters certain textures, as shown in Figure 9. The VisDrone dataset contains numerous small objects, as illustrated in Figure 10. These objects often occupy only a few dozen pixels and lack significant detail, appearing mostly as dark shapes on a gray background. During takeoff and landing, the video stream tends to be unstable, and compression artifacts increase, which occasionally results in false positives.

Figure 9: Example of false positive instances on roof with a texture
Note: Detections on the left and zoomed-in texture on the right.
Source: compiled by the author

Figure 10 further illustrates a zoomed-in image from the VisDrone dataset. A magnified section can be seen on the left; the full image is on the right, where a car is shown in a blue frame, a human in a red frame and the zoomed-in section in a yellow frame.

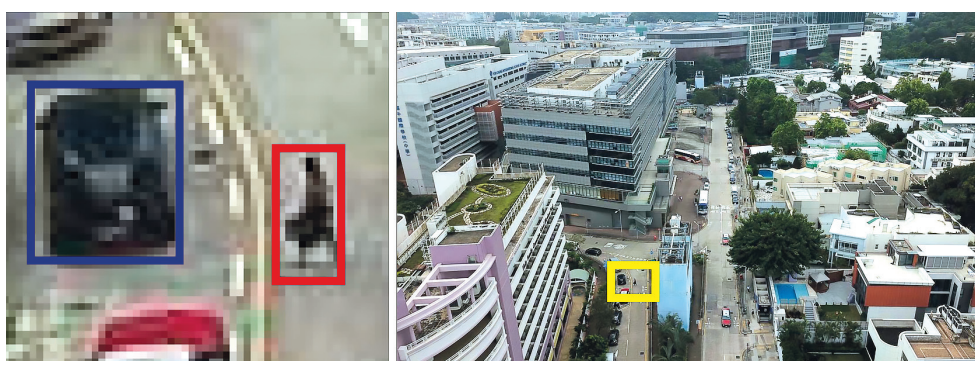

Figure 10: Example of a zoomed-in image from VisDrone dataset
Source: compiled by the author based on image found in the dataset presented by ZHU et al. 2018

Conclusion

This study highlights the challenges and opportunities of improving object detection models for UAV-based ISTAR applications by leveraging external datasets and cross-dataset training. While the initial models, trained solely on synthetic and limited real-world data collected by us, underperformed in recognising civilian objects under varied conditions, integrating large-scale datasets such as MS COCO and VisDrone significantly improved generalisation and detection performance.

However, this approach also introduced certain limitations, notably dataset imbalance and an increased risk of false positives in complex or noisy environments. Despite these challenges, the expanded training data enhanced the model's ability to detect people and vehicles in more realistic and diverse scenarios. Negative examples – images with no valid labels – further contributed to reducing false detections, demonstrating the importance of including such samples in training. This research demonstrates that supplementing domain-specific data with curated external datasets is an effective, low-cost strategy to improve the robustness and accuracy of object detection models for UAV operations. Future work may focus on addressing dataset imbalance, refining label consistency across combined datasets, and developing techniques to mitigate false positive detections caused by sensor artifacts or environmental noise. These improvements will advance further deployment of reliable AI-powered situational awareness systems in real-world defence and security applications.

References

Balogh, Péter (2012): A magyar honvédség ISTAR (ISR) képességei, a fejlesztés lehetséges irányai, különös tekintettel az elektronikai hadviselésre. *Hadmérnök,* 7(4), 75–94. Online: http://hadmernok.hu/2012_4_baloghp.pdf

Bochkovskiy, Alexey – Wang, Chien-Yao – Liao, Hong-Yuan Mark (2020): YOLOv4: Optimal Speed and Accuracy of Object Detection. *arXiv*. Online: https://doi.org/10.48550/arXiv.2004.10934

Boda, Mihály (2024): A kockázatkerülő háború és a bátorság a 20–21. század fordulóján. *Honvédségi Szemle,* 152(3), 113–125. Online: https://doi.org/10.35926/HSZ.2024.3.9

Brassai, Sándor Tihamér – Szántó, Norbert – Bajka, Adorján – Bárdi, Olivér – Németh, András – Hammas, Attila (2024): *Simulation Environment Implementation for Generation of Training Samples.* 2024 25th International Carpathian Control Conference (ICCC), Krynica Zdrój, Poland, 22–24 May 2024. Online: https://doi.org/10.1109/ICCC62069.2024.10569502

Du, Dawei – Qi, Yuankai – Yu, Hongyang – Yang, Yifan – Duan, Kaiwen – Li, Guorong – Zhang, Weigang – Huang, Qingming – Tian, Qi (2018): The Unmanned Aerial Vehicle Benchmark: Object Detection and Tracking. *arXiv*. Online: https://doi.org/10.48550/arXiv.1804.00518

Everingham, Mark – Van Gool, Luc – Williams, Christopher K. I. – Winn, John – Zisserman, Andrew (2010): The Pascal Visual Object Classes (VOC) Challenge.

International Journal of Computer Vision, 88(2), 303–338. Online: https://doi.org/10.1007/s11263-009-0275-4

Hammas, Attila (2023): *Harcjárművek észlelése szintetikusan előállított mintákon tanított mélytanuló algoritmusok segítségével.* Budapest: Nemzeti Közszolgálati Egyetem.

Li, Fei-Fei – Adeli, Ehsan (2024): *CS231n: Deep Learning for Computer Vision.* Online: https://cs231n.stanford.edu/slides/2024/lecture_1_part_2.pdf

Lin, Tsung-Yi – Maire, Michael – Belongie, Serge – Hays, James – Perona, Pietro – Ramanan, Deva – Dollár, Piotr – Zitnick, C. Lawrence (2014): Microsoft COCO: Common Objects in Context. In Fleet, David – Pajdla, Tomas – Schiele, Bernt – Tuytelaars, Tinne (eds.): *Computer Vision – ECCV 2014. Lecture Notes in Computer Science.* Cham: Springer, 740–755. Online: https://doi.org/10.1007/978-3-319-10602-1_48

Rana, Ajay – Chauhan, Kuldeep (2021): Computer Vision and Machine Learning for Image Recognition: A Review of the Convolutional Neural Network (CNN) Model. *Asian Journal of Multidimensional Research,* 10(10), 1023–1029. Online: https://doi.org/10.5958/2278-4853.2021.00920.4

Redmon, Joseph – Divvala, Santosh – Girshick, Ross – Farhadi, Ali (2015): You Only Look Once: Unified, Real-Time Object Detection. *arXiv*. Online: https://doi.org/10.48550/ARXIV.1506.02640

Redmon, Joseph – Farhadi, Ali (2016): YOLO9000: Better, Faster, Stronger. *arXiv*. Online: https://doi.org/10.48550/arXiv.1612.08242

Redmon, Joseph – Farhadi, Ali (2018): YOLOv3: An Incremental Improvement. *arXiv*. Online: https://doi.org/10.48550/arXiv.1804.02767

Ren, Shaoqing – He, Kaiming – Girshick, Ross – Sun, Jian (2017): Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 39(6), 1137–1149. Online: https://doi.org/10.1109/TPAMI.2016.2577031

Shetty, Ksheera R. – Soorinje, Vaibhav S. – Dsouza, Prinson – Swasthik (2022): Deep Learning for Computer Vision: A Brief Review. *International Journal of Advanced Research in Science, Communication and Technology,* 2(2), 450–463. Online: https://doi.org/10.48175/IJARSCT-2898

Ultralytics (s. a.): *Ultralytics Documentation.* Online: https://docs.ultralytics.com/

Van Doorn, Joost (2014): *Analysis of Deep Convolutional Neural Network Architectures.* Proceedings of the Twenty First Twente Student Conference on IT, Enschede, The Netherlands, 1–7.

Voulodimos, Athanasios – Doulamis, Nikolaos – Doulamis, Anastasios – Protopapadakis, Eftychios (2018): Deep Learning for Computer Vision: A Brief Review. *Computational Intelligence and Neuroscience,* 2018(1), 1–13. Online: https://doi.org/10.1155/2018/7068349

Xia, Gui-Song – Bai, Xiang – Ding, Jian – Zhu, Zhen – Belongie, Serge – Luo, Jiebo – Datcu, Mihai – Pelillo, Marcello – Zhang, Liangpei (2018): *DOTA: A Large-Scale Dataset for Object Detection in Aerial Images.* 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, 18–23 June 2018. Online: https://doi.org/10.1109/CVPR.2018.00418

YAO, Yongqiang – WANG, Yan – GUO, Yu – LIN, Jiaojiao – QIN, Hongwei – YAN, Junjie (2020): Cross-dataset Training for Class Increasing Object Detection. *arXiv*. Online: https://doi.org/10.48550/arXiv.2001.04621

ZHAO, Xia – WANG, Limin – ZHANG, Yufei – HAN, Xuming – DEVECI, Muhammet – PARMAR, Milan (2024): A Review of Convolutional Neural Networks in Computer Vision. *Artificial Intelligence Review,* 57(4). Online: https://doi.org/10.1007/s10462-024-10721-6

ZHU, Pengfei – WEN, Longyin – BIAN, Xiao – LING, Haibin – HU, Qinghua (2018): Vision Meets Drones: A Challenge. *arXiv*. Online: https://doi.org/10.48550/arXiv.1804.07437