

YOLO-Based Vehicle Detection in Drone Aerial Imagery Using Transfer Learning

¹Siddiqui Arshad Hussain Abdul Bari, ²Bhavna Pancholi

Submitted: 17/04/2022 Revised: 20/05/2022 Accepted: 16/06/2022

Abstract: Vehicle detection in drone aerial imagery has become an important research area due to the growing use of Unmanned Aerial Vehicles (UAVs) in intelligent transportation systems, surveillance, disaster monitoring, border security, and urban traffic management. Traditional vehicle detection methods based on handcrafted features and conventional machine learning techniques often struggle to achieve high accuracy under complex aerial imaging conditions such as varying illumination, occlusion, small object size, dense traffic regions, and complex backgrounds. This study presents a YOLO-based deep learning framework integrated with transfer learning for efficient vehicle detection in aerial drone images. The proposed approach utilizes pretrained COCO dataset weights along with aerial datasets including VEDAI, DOTA, COWC, and VAID for model training and evaluation. Data preprocessing, annotation conversion into YOLO format, and optimized network parameter configuration are employed to improve detection performance and computational efficiency. The customized YOLO model effectively detects vehicles in real-time aerial scenes while operating with moderate GPU resources. Experimental results demonstrate improved detection accuracy, reduced training time, and robust performance under challenging environmental conditions. Evaluation metrics such as Precision, Recall, Intersection over Union (IoU), Average Precision (AP), and Mean Average Precision (mAP) validate the reliability and effectiveness of the proposed framework for UAV-based intelligent transportation and surveillance applications.

Keywords: YOLO, Vehicle Detection, Drone Aerial Imagery, Transfer Learning, UAV Surveillance

Introduction

The rapid advancement of Unmanned Aerial Vehicles (UAVs) and deep learning technologies has significantly transformed the field of aerial surveillance and intelligent transportation systems. Drone aerial imagery provides high-resolution visual data that can efficiently monitor large geographical areas in real time, making it highly useful for traffic management, border security, disaster monitoring, urban planning, and surveillance applications. Among various computer vision tasks, vehicle detection from aerial imagery has emerged as a challenging research problem due to factors such as small object size, varying illumination conditions, complex backgrounds, shadows, occlusion, and different viewing angles. Traditional machine learning approaches based on handcrafted features and region proposal methods often suffer from limited accuracy and high computational complexity when applied to aerial scenes. Therefore, advanced deep learning-based object detection techniques are increasingly being adopted to improve detection performance and processing speed.

YOLO (You Only Look Once) is a state-of-the-art real-time object detection algorithm that has gained widespread popularity due to its high detection speed and

accuracy. Unlike conventional object detection methods, YOLO performs object localization and classification simultaneously in a single neural network pass, enabling efficient real-time detection. In this research, a YOLO-based vehicle detection framework using transfer learning is proposed for drone aerial imagery analysis. Transfer learning utilizes pretrained weights from the COCO dataset, allowing the model to learn generalized object features while reducing training time and computational requirements. The proposed system is trained and evaluated using aerial image datasets such as VEDAI and VAID to detect vehicles under diverse environmental and traffic conditions. Experimental results demonstrate that the customized YOLO model effectively identifies vehicles with high confidence and robust performance, making it suitable for intelligent transportation, aerial surveillance, and smart city applications.

The motivation behind this research lies in developing an efficient and cost-effective aerial vehicle detection framework using transfer learning and YOLO architecture. Instead of training a deep neural network from scratch, pretrained weights from the COCO dataset are utilized to reduce computational requirements and training duration. The proposed system customizes publicly available YOLO models and adapts them for aerial imagery applications using specialized datasets such as VEDAI, DOTA, COWC, and VAID. This approach enables accurate vehicle detection even with

¹Research Scholar, Department of Electrical Engineering, Faculty of Technology and Engg, The M S University of Baroda.

²Assistant Professor, Department of Electrical Engineering, Faculty of Technology and Engg, The M S University of Baroda

limited hardware resources and moderate GPU computational power.

The primary objectives of this research are:

- To study the effectiveness of YOLO-based deep learning algorithms for aerial vehicle detection.
- To preprocess and convert aerial datasets into YOLO-compatible annotation formats.
- To implement transfer learning using pretrained COCO weights.
- To evaluate detection performance under varying aerial imaging conditions.
- To reduce computational complexity while maintaining real-time detection capability.

The proposed research contributes toward intelligent transportation systems, automated traffic analysis, urban surveillance, disaster response systems, and military reconnaissance applications. The study also highlights the importance of efficient preprocessing, annotation techniques, and optimized model training strategies for aerial image analysis.

Literature Review

The rapid advancement of deep learning and computer vision technologies has transformed intelligent transportation systems, aerial surveillance, and autonomous monitoring applications. Among the various object detection algorithms, the YOLO (You Only Look Once) framework has emerged as one of the most efficient and accurate approaches for real-time object recognition and tracking. Neupane, Horanont, and Aryal (2022) developed a transfer learning-improved deep learning model for real-time vehicle classification and tracking. Their study demonstrated that transfer learning significantly enhances the accuracy and efficiency of vehicle recognition systems by utilizing pretrained convolutional neural networks. The proposed framework improved detection speed while reducing computational complexity, making it suitable for traffic monitoring applications in dynamic urban environments. Similarly, Wang et al. (2021) proposed a YOLO-based target detection model specifically designed for offshore UAV datasets. Their research focused on detecting moving targets in maritime environments where varying illumination, sea reflections, and object motion create challenges for traditional detection methods. The findings indicated that optimized YOLO architectures can successfully identify and track targets with high precision under complex environmental conditions. These studies collectively highlight the increasing adoption of deep learning models in intelligent monitoring systems due to their capability to perform

accurate real-time analysis while maintaining computational efficiency.

The application of UAV imagery for surveillance and transportation monitoring has become an important research area because drones provide wide-area coverage, operational flexibility, and cost-effective data collection. Zhang Sr et al. (2019) investigated YOLO-based pedestrian detection for UAV monitoring systems and demonstrated that deep learning algorithms can accurately detect pedestrians from aerial viewpoints despite scale variations and motion blur. Their work emphasized the importance of feature extraction and multiscale learning for improving pedestrian recognition in real-time UAV applications. Uzar et al. (2021) conducted a detailed performance analysis of multiple YOLO versions for automatic vehicle detection from UAV images. The researchers compared detection accuracy, processing speed, and computational efficiency across different YOLO models and concluded that newer YOLO versions provided better balance between speed and precision. The study highlighted the suitability of advanced YOLO architectures for smart surveillance and traffic management systems. Yu et al. (2021) further contributed to UAV-based vehicle detection by introducing a spatial cognition-driven deep learning framework for car detection in aerial imagery. Their approach incorporated contextual spatial relationships to improve the recognition of small and partially occluded vehicles in dense traffic environments. The model achieved improved detection accuracy by integrating spatial cognition mechanisms with deep neural networks. These studies collectively demonstrate that UAV-assisted intelligent monitoring systems have significantly benefited from deep learning advancements, particularly in the areas of multiscale object recognition, contextual learning, and real-time processing capabilities.

Vehicle and object detection under challenging environmental conditions has also become a major focus of intelligent transportation research. Mahmood, Ahmed, and Ahmed (2020) proposed a YOLO computational mechanism for vehicle detection using infrared road traffic images. Their study addressed the limitations of traditional RGB-based imaging systems in nighttime and low-light environments. The researchers demonstrated that infrared imaging combined with YOLO algorithms significantly improves detection accuracy during adverse weather and poor visibility conditions. This approach provides enhanced reliability for intelligent traffic surveillance systems operating continuously throughout the day and night. Martinez-Alpiste et al. (2020) similarly investigated real-time low-pixel infrared human detection from unmanned aerial vehicles. Their work focused on identifying humans in low-resolution thermal imagery captured by UAVs. The researchers optimized

detection performance through lightweight deep learning strategies capable of handling limited pixel information. The proposed method achieved reliable detection accuracy while maintaining low computational requirements suitable for UAV deployment. These studies highlight the growing importance of infrared sensing technologies in surveillance applications, especially in scenarios where visible-spectrum imaging becomes ineffective. The integration of thermal imaging with deep learning algorithms has expanded the capabilities of intelligent monitoring systems in security, transportation, and emergency response applications.

Apart from infrared imaging, radar-assisted deep learning approaches have also contributed significantly to intelligent object detection systems. Kim et al. (2020) proposed a YOLO-based simultaneous target detection and classification framework using automotive FMCW radar systems. Their research combined radar signal processing with deep neural networks to enhance target recognition in autonomous driving applications. The proposed framework demonstrated strong performance in detecting and classifying multiple objects simultaneously under adverse environmental conditions such as rain, fog, and low visibility. Radar-based sensing provides reliable distance and motion information, which complements visual imaging systems and improves overall detection robustness. In another related study, Al-Shakarji et al. (2019) developed a multi-cue vehicle detection framework for semantic video compression in georegistered aerial videos. Their approach integrated appearance cues, motion analysis, and contextual spatial information to improve vehicle detection while simultaneously optimizing video compression efficiency. The proposed system enabled efficient aerial surveillance by reducing bandwidth requirements without sacrificing detection performance. Li, Wang, and Ding (2019) further addressed the challenge of tiny vehicle detection in UAV imagery. Since vehicles often appear as very small objects in high-altitude aerial images, traditional detection algorithms struggle to identify them accurately. The researchers introduced improved feature extraction and multiscale learning mechanisms to enhance small object recognition. Their findings demonstrated that specialized deep learning architectures significantly improve the detection of tiny vehicles in UAV-based surveillance systems. Together, these studies illustrate the importance of integrating multiple sensing modalities and feature enhancement techniques to improve the robustness and scalability of intelligent detection frameworks.

Deep learning techniques developed for vehicle detection have also shown effectiveness in broader aerial object

recognition applications. Hong et al. (2019) explored the application of deep learning methods for bird detection using unmanned aerial vehicle imagery. Although the study focused on wildlife monitoring rather than transportation systems, it provided valuable insights into aerial object detection challenges such as scale variation, background complexity, and object motion. The researchers demonstrated that convolutional neural networks can successfully distinguish small flying objects from complex natural environments with high accuracy. Their findings highlighted the adaptability of deep learning frameworks for diverse aerial monitoring tasks. Overall, the reviewed literature confirms that YOLO-based deep learning models have become essential tools for real-time vehicle, pedestrian, and object detection across UAV, infrared, radar, and aerial imaging applications. Researchers have consistently focused on improving detection accuracy, computational efficiency, and robustness against environmental challenges including low illumination, occlusion, scale variation, and background clutter. Transfer learning, multiscale feature extraction, contextual learning, and multimodal sensing integration have significantly improved the performance of intelligent monitoring systems. Despite these advancements, several challenges remain unresolved, including efficient detection of extremely small objects, limited onboard computational resources in UAV platforms, and the requirement for large annotated datasets. Future research is expected to emphasize lightweight neural networks, edge computing, transformer-based architectures, and sensor fusion techniques to further enhance the performance and scalability of intelligent aerial surveillance and transportation monitoring systems.

Methodology

The proposed methodology employs a YOLO-based deep learning framework integrated with transfer learning for vehicle detection in drone aerial imagery. The system architecture consists of dataset collection, preprocessing, annotation conversion, model training, transfer learning, and performance evaluation stages.

3.1 Dataset Collection

The research utilizes publicly available aerial image datasets including:

- VEDAI
- COWC
- DOTA
- VAID

No.	Dataset	Image	Images	Classes	Image Size	Annotations
-----	---------	-------	--------	---------	------------	-------------

		Format				
01	VEDAI	JPEG	1250	9	512×512, 1024×1024	Sample class, ground truth (GT) center point coordinates, direction, and coordinates of GT's four corners
02	COWC	TIFF	53	4	2000×2000 to 19,000×19,000	Sample class and GT center point coordinates
03	DOTA	JPEG	2086	15	Approximately 4000×4000	Sample class and coordinates of GT's four corners
04	VAID	JPEG	6000	7	1137×640	XML annotation files

These datasets contain aerial images captured from drones, satellites, and surveillance systems under varying environmental conditions. The datasets include multiple vehicle categories such as cars, trucks, buses, trailers, pickups, and vans. The VAID dataset particularly contains 6000 aerial images captured in Taiwan with a resolution of 1137×640 pixels.

Data Preprocessing

Preprocessing is essential for improving dataset quality and ensuring compatibility with YOLO architecture. The preprocessing stage includes:

- Image resizing
- Noise removal
- Data normalization
- Annotation correction
- Dataset augmentation

The aerial image annotations are converted into YOLO format where each object is represented using normalized coordinates:

- Class ID
- x-center
- y-center
- width
- height

The normalized annotation format improves detection consistency across varying image dimensions.

COCO Dataset Weights use in YOLO Algorithm

YOLO Object Detection Algorithm

YOLO (You Only Look Once) is a real-time object detection algorithm based on Convolutional Neural Networks (CNNs) that revolutionized computer vision by performing object detection in a single pass through the neural network. Unlike traditional region proposal-based methods such as Faster R-CNN, YOLO divides an image into grids and simultaneously predicts bounding boxes and class probabilities, making it significantly faster and suitable for real-time applications. YOLO v1, introduced by Joseph Redmon in 2015, established the foundation for single-stage object detection with high processing speed and multi-object detection capability.

YOLO v2 improved the limitations of YOLO v1 by introducing batch normalization, anchor boxes, and multi-scale training. These enhancements increased recall accuracy and improved detection performance for varying image resolutions. The model dynamically adjusted input image sizes during training, making it more robust for different scales of objects.

YOLO v3 further enhanced object detection performance by using the Darknet-53 architecture and multiple detection layers. It introduced logistic loss and nine anchor boxes, significantly improving the detection of small, dense, and overlapping objects. YOLO v4 optimized both speed and accuracy, enabling real-time detection on standard GPUs while achieving higher Average Precision (AP) and Frames Per Second (FPS), making it highly effective for aerial vehicle detection applications.

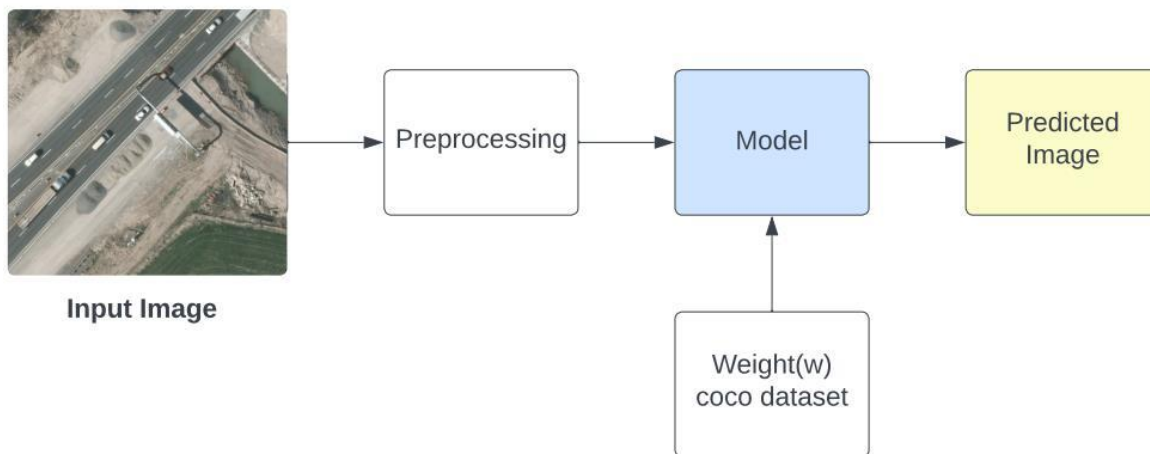


Figure 1 COCO Dataset Weights use in YOLO Algorithm

The figure illustrates the workflow of using pretrained COCO dataset weights in the YOLO object detection algorithm for aerial image analysis. Initially, the input aerial image undergoes preprocessing to improve image quality and normalize dimensions before being passed to the YOLO model. The model utilizes pretrained weights derived from the COCO dataset, which contains a large

collection of labeled objects. These pretrained weights enable the network to recognize general object features such as edges, shapes, and textures efficiently. Finally, the model generates the predicted output image with detected objects highlighted, improving detection accuracy while reducing training time and computational requirements.

Experiment Result & Observation

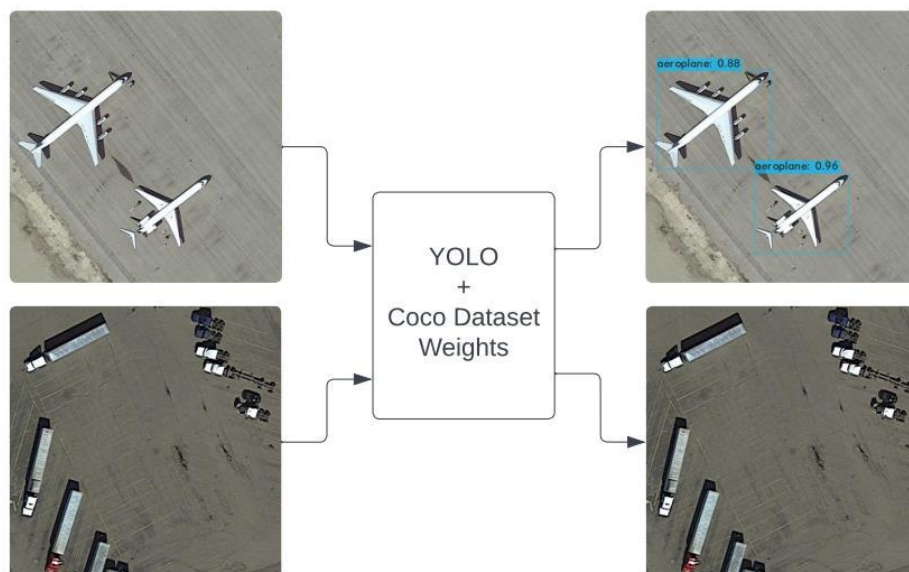


Figure 2 Experiment Result & Observation, Case Study – 1

This Google Colab snapshot demonstrates the implementation of the YOLO v4 algorithm for aerial image object detection. The upper section presents the original input aerial image containing roads, vegetation, and surrounding land areas, while the lower section shows the predicted output image with detected objects highlighted using bounding boxes and labels. The model

Test Case - ii

successfully identifies vehicles and other objects with corresponding confidence scores, reflecting prediction accuracy. The example highlights the real-time detection capability, efficiency, and reliability of YOLO v4 for aerial surveillance and intelligent transportation applications.



Figure 3 Case Study - ii, Input image, Output Image

Custom YOLO model Experiments result & Observation

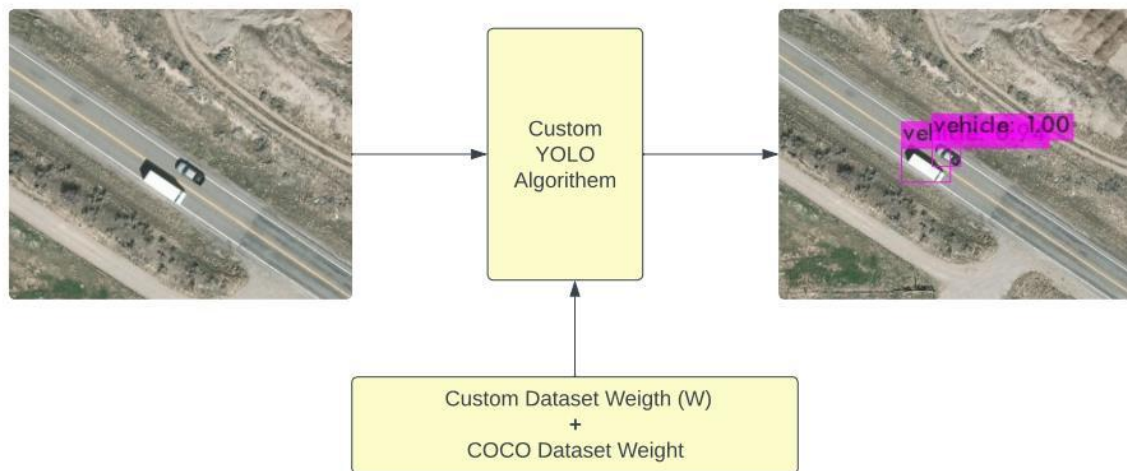


Figure 4 Our Model with custom dataset weight(w), predicting the output of VEDAI dataset

The figure presents the experimental workflow and observation of the customized YOLO model for vehicle detection in aerial imagery using the VEDAI dataset. The process begins with an input aerial image containing vehicles captured from a top-view perspective. The image is processed through a custom YOLO algorithm trained using a combination of custom dataset weights and pretrained COCO dataset weights. This transfer

learning approach enhances the model's ability to recognize vehicle features accurately while reducing computational requirements. The output image demonstrates successful vehicle detection with bounding boxes and confidence scores, highlighting the effectiveness, precision, and real-time capability of the proposed custom YOLO-based detection framework.



Figure 5 Input image of VAID Dataset (Left Side), Predicted Output Image from the Our Model (Right Side), Showing three vehicles



Figure 6 Input image of VAID Dataset (Left Side), Predicted Output Image from the Our Model (Right Side) showing four vehicles

The figure demonstrates the vehicle detection capability of the proposed custom YOLO model on the VAID aerial image dataset. The left image represents the original drone-captured input scene containing a highway and multiple moving vehicles, while the right image shows the predicted output generated by the trained model. The algorithm successfully identifies and localizes four vehicles using bounding boxes and confidence labels.

The detection results indicate that the model performs effectively under aerial viewing conditions with varying object orientations and complex road backgrounds. This experiment validates the robustness, accuracy, and real-time detection performance of the customized YOLO framework for intelligent transportation and aerial surveillance applications.



Figure 7 Input image to Predicted Output Image from the Out Model

The figure illustrates the transformation of an input aerial image into a predicted output image using the proposed custom YOLO-based vehicle detection model. The upper image represents the original drone-captured road intersection containing multiple vehicles under real-world traffic conditions. The lower image shows the detection results generated by the trained model, where vehicles are accurately identified with bounding boxes, labels, and confidence scores. The algorithm successfully detects several vehicles of different sizes and positions despite the complexity of the traffic environment. These results demonstrate the effectiveness, robustness, and real-time object detection capability of the proposed YOLO framework for aerial traffic monitoring and intelligent transportation applications.

Conclusion

This research presented a YOLO-based vehicle detection framework for drone aerial imagery using transfer learning techniques. The study demonstrated that integrating pretrained YOLO models with custom aerial image datasets significantly improves the efficiency and accuracy of vehicle detection in UAV-based surveillance systems. Transfer learning using COCO pretrained weights reduced computational complexity, training time, and hardware requirements while maintaining strong detection performance. The proposed framework successfully detected vehicles in challenging aerial conditions such as varying illumination, complex

backgrounds, shadows, dense traffic regions, and small object sizes.

The experimental results obtained from datasets including VEDAI and VAID confirmed the capability of the customized YOLO model to perform accurate real-time object detection with high confidence scores and stable mean Average Precision (mAP). The gradual reduction in training loss and consistent prediction accuracy validated the robustness and convergence of the model. Additionally, the system effectively localized vehicles using bounding boxes under different viewing angles and environmental conditions.

The research highlights the practical importance of deep learning and transfer learning in intelligent transportation systems, traffic monitoring, border surveillance, and urban management applications. Future improvements may include lightweight edge-based implementations, multi-object tracking, enhanced small object detection, and integration with IoT-enabled smart city infrastructure for advanced real-time aerial surveillance systems.

References

- [1] Neupane, B., Horanont, T., & Aryal, J. (2022). Real-time vehicle classification and tracking using a transfer learning-improved deep learning network. *Sensors*, 22(10), 3813.
- [2] Wang, Z., Zhang, X., Li, J., & Luan, K. (2021). A YOLO-based target detection model for offshore

- unmanned aerial vehicle data. *Sustainability*, 13(23), 12980.
- [3] Zhang Sr, D., Shao, Y., Mei, Y., Chu, H., Zhang, X., Zhan, H., & Rao, Y. (2019, May). Using YOLO-based pedestrian detection for monitoring UAV. In *Tenth International Conference on Graphics and Image Processing (ICGIP 2018)* (Vol. 11069, pp. 1141-1145). SPIE.
- [4] Mahmood, M. T., Ahmed, S. R. A., & Ahmed, M. R. A. (2020, November). Detection of vehicle with Infrared images in Road Traffic using YOLO computational mechanism. In *IOP Conference Series: Materials Science and Engineering* (Vol. 928, No. 2, p. 022027). IOP Publishing.
- [5] Uzar, A., Öztürk, Ş., Bayrak, O., Arda, T., & Tunalioglu Öcalan, N. (2021). Performance analysis of YOLO versions for automatic vehicle detection from UAV images. *Advanced Remote Sensing*, 1(1).
- [6] Hong, S. J., Han, Y., Kim, S. Y., Lee, A. Y., & Kim, G. (2019). Application of deep-learning methods to bird detection using unmanned aerial vehicle imagery. *Sensors*, 19(7), 1651.
- [7] Kim, W., Cho, H., Kim, J., Kim, B., & Lee, S. (2020). YOLO-based simultaneous target detection and classification in automotive FMCW radar systems. *Sensors*, 20(10), 2897.
- [8] Yu, J., Gao, H., Sun, J., Zhou, D., & Ju, Z. (2021). Spatial cognition-driven deep learning for car detection in unmanned aerial vehicle imagery. *IEEE transactions on cognitive and developmental systems*, 14(4), 1574-1583.
- [9] Al-Shakarji, N., Bunyak, F., Aliakbarpour, H., Seetharaman, G., & Palaniappan, K. (2019). Multi-cue vehicle detection for semantic video compression in georegistered aerial videos. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops* (pp. 56-65).
- [10] Li, J., Wang, R., & Ding, J. (2019, April). Tiny vehicle detection from uav imagery. In *Chinese Conference on Image and Graphics Technologies* (pp. 370-381). Singapore: Springer Singapore.
- [11] Martinez-Alpiste, I., Golcarenenji, G., Wang, Q., & Alcaraz-Calero, J. M. (2020, November). Real-time low-pixel infrared human detection from unmanned aerial vehicles. In *Proceedings of the 10th ACM symposium on design and analysis of intelligent vehicular networks and applications* (pp. 9-15).