SETSCI Conference Proceedings 21, 12-17, 2024

International Trend of Tech Symposium

7-8 December 2024, İstanbul - Türkiye

https://doi.org/10.36287/setsci.21.3.012

Integrating UAVs and YOLO Deep Learning for Early-Stage Forest Fire Detection

Hakan Mert Didis[1*], Firuze Adibeli[2], Ilayda Boz[3], Nesli S. Azdavay[1]

[1]Computer Engineering Department, Mersin University, Mersin, Turkey
[2]Department of Environmental Engineering, Mersin University, Mersin, Turkey
[3]Aerospace Engineering Department, Tarsus University, Mersin, Turkey
[*](h.didis@mersin.edu.tr)

Abstract – The uncontrolled spread of forest fires, driven by climate change and increasing human activities, poses a significant global challenge, disrupting the delicate balance of ecosystems. These fires, whether caused by natural disasters or human factors, have severe and often irreversible economic, social, and environmental consequences. The rising frequency of forest fires due to both artificial and natural causes is becoming a major international concern. Fire detection methods range from human observation and satellite-based systems to optical smoke detection and watchtowers. While each method offers advantages depending on the application, traditional approaches often suffer from delayed response times, making early intervention difficult. Unmanned aerial vehicles (UAVs) have emerged as a powerful alternative in fire detection, offering the mobility to access difficult terrain and perform rapid, detailed observations over large forested areas. This study explores how forest fires are detected using UAVs integrated with YOLOv4, YOLOv7, and YOLOv9 models, while also examining the limitations of these algorithms. The performance comparison reveals that YOLOv9 outperforms the other models in terms of precision, recall, and speed. With a precision of 0.922, mAP@50 of 0.915, and mAP@50:95 of 0.872, YOLOv9 demonstrates superior performance, exceeding YOLOv7 by 2.78% and YOLOv4 by 9.33%. These findings offer valuable insights into the use of UAVs for fast and accurate fire detection, highlighting their critical role in combating forest fires effectively.

Keywords – forest fire, fire detection, deep learning, unmanned aerial vehicle, YOLOv4, YOLOv7, YOLOv9

I. INTRODUCTION

Forest fires, which destroy thousands of hectares of forests each year, have become a global concern, exacerbated by climate change. These fires can occur naturally due to dry weather conditions or lightning strikes, but human activities such as campfires, picnics, and discarded cigarette butts significantly contribute to their frequency. As one of the most devastating natural disasters, forest fires cause immense economic losses, severely damage the environment, and, most critically, endanger human lives. The impact of these fires can be analyzed across economic, environmental, and social dimensions. Forest ecosystems, rich in biodiversity, suffer significant harm, while the trees' ability to absorb atmospheric carbon dioxide diminishes, leading to increased carbon emissions. As a result, forest fires contribute to global warming and further climate deterioration. In economic terms, these fires damage agricultural lands and settlements, while socially, they affect communities both physically and psychologically. Given the rapid spread of forest fires, often referred to as the "lungs of the planet," effective detection and control are crucial for protecting all life on Earth. Various fire detection methods have been developed, ranging from human observation to satellite, camera, and sensor-based techniques [1]. However, traditional detection methods have notable limitations, especially in terms of rapid response across large forested areas [2]. Human observation, once a primary method, is now impractical due to labor constraints and its inefficiency in covering vast areas. Satellite-based detection, while effective for wide-area monitoring, lacks the capability for real-time and early-stage fire detection [3]. Although satellites offer advantages in detecting large-scale fires, they fall short in identifying smaller regional fires at their onset.

In contrast, unmanned aerial vehicles (UAVs) offer significant advantages due to their superior mobility, high-resolution imaging, and real-time data transmission capabilities [4]. UAVs can be deployed quickly and provide detailed information by focusing on specific areas, making them a promising solution for early fire detection [5]. Their flexibility and speed allow for more effective monitoring, potentially preventing the devastating effects of wildfires [6]. Accurate and timely detection is critical in mitigating fire-related losses, highlighting the importance of fast, reliable fire detection systems. With the growing integration of computer vision object detection methods across multiple fields, their application in fire and smoke detection is gaining attention [7]. In security and surveillance, for example, object detection systems alert authorities to potential threats, allowing for timely intervention. Similarly, in autonomous systems, unmanned vehicles rely on cameras and sensors to safely navigate their environments by detecting surrounding objects [8]. Fire and smoke detection pose unique challenges due to their dynamic nature, but the development of advanced machine vision algorithms offers an economical and effective solution for early wildfire detection. There are two primary types of object detection algorithms used in wildfire detection: two-stage and single-stage detectors. Two-stage algorithms,

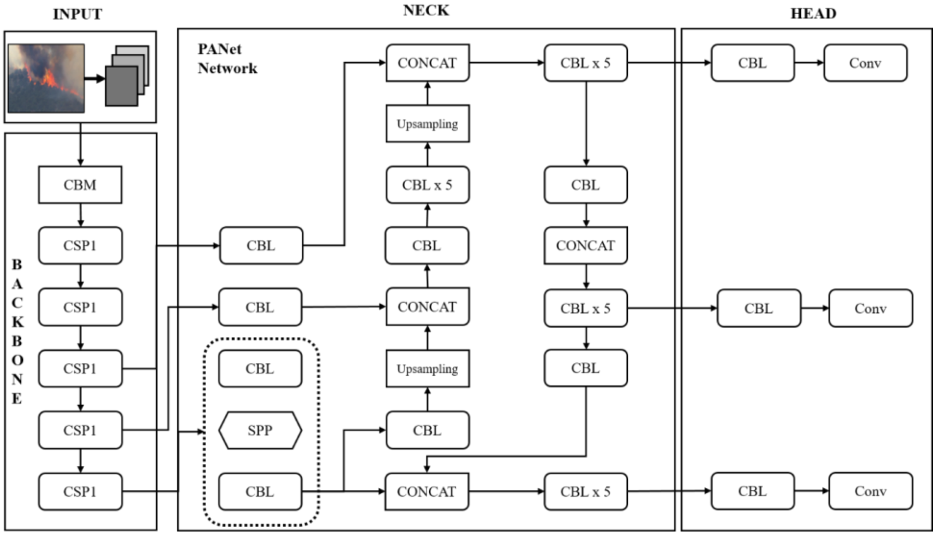

Fig. 1. Network architecture of YOLOv4.

such as Region-based Convolutional Neural Networks (R-CNN) [9] and Faster R-CNN [10], are highly accurate but slower due to their use of a region proposal network (RPN), which first scans the image for candidate objects before performing detailed classification. On the other hand, single-stage algorithms such as SSD [11], R-SSD [12], CenterNet [13], and YOLO [14, 15] are faster but generally less accurate, as they perform object detection and classification in one step.

In this study, we compare the performance of YOLOv4, YOLOv7, and YOLOv9 algorithms in detecting forest fires, particularly in the context of climate-induced changes, using UAV-captured images. As members of the YOLO family, these models simultaneously predict bounding boxes and class probabilities, making them well-suited for real-time applications such as wildfire detection, where a trade-off between speed and accuracy is crucial. This study aims to evaluate the strengths and weaknesses of these models through comprehensive performance measurements, contributing to the ongoing effort to mitigate the immense damage caused by forest fires.

II. Methodology

YOLO is a highly efficient and widely used object detection algorithm, renowned for its real-time performance. Unlike other object detection algorithms that rely on region proposal methods, YOLO predicts both the class and location of all objects in an image in a single pass through the neural network, which allows for fast detections. However, despite its speed, YOLO tends to exhibit lower accuracy when detecting small objects [16].

A. YOLO Algorithms

Introduced in April 2020 [17], YOLOv4 is a deep learning-based object detector known for its balance between speed and accuracy. YOLOv4's architecture, as shown in Fig. 1, is divided into four main components: the input, backbone, neck, and head. In the input component, the image is fed into the algorithm in this stage. For the backbone, YOLOv4 uses the CSP connections combined with Darknet-53 for feature extraction. This backbone efficiently captures essential image features. The neck layer is responsible for multi-scale feature aggregation. YOLOv4 introduces innovative modules here, such as the Path Aggregation Network (PAN), Spatial Pyramid Pooling (SPP), and Spatial Attention Module (SAM). PAN enhances detection by combining global and local information. SPP improves the network's spatial awareness by pooling features at different scales. SAM, on the other hand, helps the model focus on the most important parts of the input by applying spatial attention mechanisms. In the head layer, the final detection is performed in this stage, where bounding boxes are predicted and objects are classified based on the features extracted by the previous layers. This combination of innovations enables YOLOv4 to provide a well-rounded performance, making it suitable for various real-time detection tasks.

With its architecture shown in Fig. 2, YOLOv7 represents a significant improvement in single-stage object detection over its predecessors, especially in balancing speed and accuracy [18]. It is designed to operate efficiently across a wide range of scenarios, delivering frame rates between 5 FPS and 160 FPS, depending on the model configuration. Similar to other YOLO models, the input consists of images or video frames. The backbone incorporates convolutional layers, max-pooling layers, and advanced modules such as Extended Efficient Layer Aggregation Networks (ELAN) and SPPCSPC. ELAN is used to design an efficient network structure that strengthens learning and optimizes gradient flow. SPPCSPC combines the strengths of CSP with spatial pyramid pooling, enhancing the network's ability to perceive objects at multiple scales. The neck of YOLOv7 introduces the CAT (concatenate) module, which combines feature maps from different levels to support multi-scale detection. This allows the model to detect objects of various sizes more accurately. The final layer, head, generates bounding boxes and classifies objects based on the aggregated feature maps. YOLOv7 uses three distinct loss functions: bounding box loss, objectness loss, and classification loss. The bounding box loss measures the overlap between the predicted and true object location, the objectness loss assesses the confidence of the detection, and the classification loss evaluates the accuracy of the object class prediction. These loss functions work together to optimize the model for accurate and efficient object detection.

YOLOv9, introduced in February 2024, builds on the strengths of its predecessors while introducing cutting-edge features [19, 20]. It excels in both object detection and segmentation, making it suitable for tasks requiring high precision and speed. YOLOv9 addresses the slow convergence problem with its two main architectural innovations:

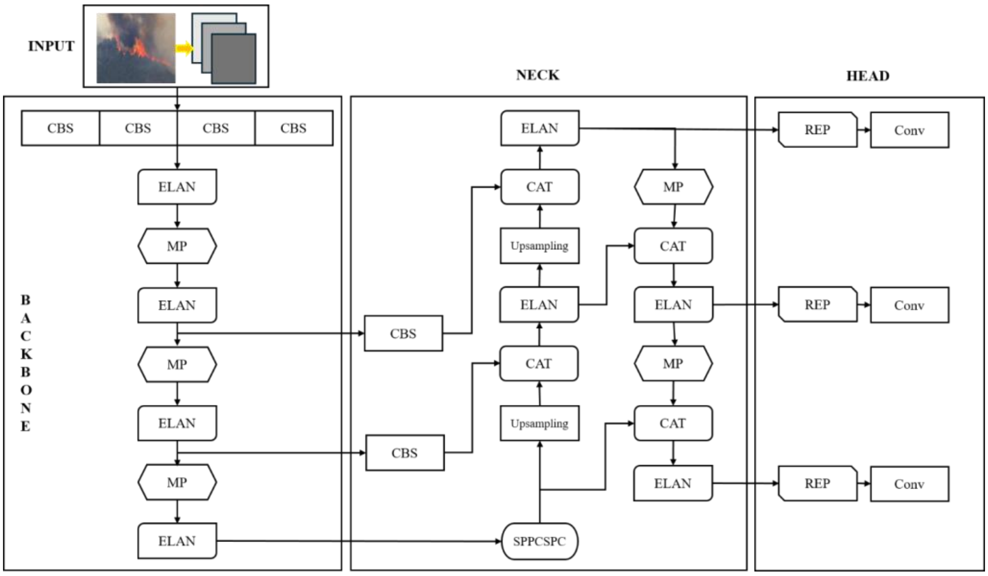

Fig. 2. YOLOv7's network architecture.

Programmable Gradient Information (PGI) and Generalized Efficient Layer Aggregation Network (GELAN). PGI enhances the efficiency of gradient computation, leading to faster and more reliable convergence during training. GELAN optimizes parameter usage, reducing computational redundancy while maintaining high performance. Together, these architectures prevent information bottlenecks and improve the model's accuracy by minimizing information loss during training. YOLOv9's architecture is grounded in the Information Bottleneck Principle, which ensures that the most relevant information is preserved during data transformation in deep neural networks. The mutual information between the input and output is maximized, preventing the loss of critical information. This principle is expressed mathematically as:

$$I(X, X) \geq I(X, f_\theta(X)) \geq I(X, g_\varphi(f_\theta(X))) \tag{1}$$

where I represents mutual information, f and g are transformation functions, and θ and ϕ are their respective parameters. YOLOv9 applies invertible functions to neural networks to further reduce information loss during data transformations:

$$X = v_\zeta\left(r_\psi(X)\right) \tag{2}$$

where r represents forward transformations and v represents backward transformations, with ψ and ζ being their respective parameters. YOLOv9's improvements in gradient computation and feature aggregation make it a powerful tool for real-time detection, maintaining a balance between speed and accuracy that is crucial for applications like wildfire detection.

B. Performance Metrics

Several key performance metrics were employed to compare the algorithms' effectiveness. Recall represents the proportion of true positives that are correctly identified out of all actual positives, as expressed in (3):

$$\text{Recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{3}$$

where TP denotes true positives, and FN represents false negatives. Precision quantifies the accuracy of the positive predictions by measuring the proportion of true positives out of all predicted positives, as shown in (4):

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{4}$$

where FP stands for false positives. The F1 Score calculates the harmonic mean of Precision and Recall, offering a balanced measure between the two, as illustrated in (5):

$$\text{F1 Score} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall+Precision}} \tag{5}$$

These expressions rely on the values of True Positives (TP), False Positives (FP), and False Negatives (FN) to provide an accurate assessment of detection performance.

C. Dataset

To evaluate the performance of YOLOv4, YOLOv7, and YOLOv9 in detecting forest fires, a comprehensive dataset was compiled from a variety of challenging conditions. As forest fires did not occur in the region during the study period, UAV flights could not be conducted to capture real-time data. Instead, forest fire images and videos were sourced from web-based repositories. This dataset includes images from various fires captured at different times, terrain types, and environmental conditions, offering diverse scenarios for the models to handle. The robustness of the models was further tested by incorporating images taken from various altitudes and angles, enhancing the models' ability to detect fires from multiple perspectives. The resulting dataset consists of 4,730 images, which were split into training (62%), validation (14%), and testing (16%) sets. The validation and testing sets included images that the models had not encountered during training to ensure a reliable evaluation of their performance. Model training was conducted on Google Colaboratory

(Colab), a cloud-based platform that integrates Jupyter notebooks and offers powerful GPU and TPU accelerators, making it highly suitable for deep learning tasks.

III. Results

Fig. 3. Examples of forest fire detection using the YOLOv9 algorithm, showcasing both successful detections and areas with partial or missed detections.

Fig. 3 provides an in-depth analysis of YOLOv9's performance in detecting forest fires, highlighting both its successes and limitations. Image 1 shows that YOLOv9 struggled with drawing bounding boxes of adequate size for fires visible through dense trees. Although the fire was not entirely covered, the detection was reasonably accurate. Image 2 demonstrates the impact of smoke on detection accuracy. The wind direction caused smoke to obscure flames, leading to partial detection. While YOLOv9 detected fires in some areas, it failed to capture others due to the dense smoke cover. The third example shows a large, visible fire, where YOLOv9 performed successfully, detecting the flames with high accuracy. Image 4 reveals a limitation: while flames were detected, the bounding boxes only partially enclosed them, missing some flame regions. Additionally, thick smoke may have concealed other flames, lowering the detection success. Image 5 illustrates the challenges posed by dense smoke, which created blurriness in the image spectrum. Although distinct flames were detected, YOLOv9 failed to identify ember flames, and flames behind the smoke were also missed. Image 6 shows a distinct flame in the lower left corner, which YOLOv9 detected without difficulty. Image 7 depicts a hazy scene caused by smoke, yet YOLOv9 was still able to detect flames amidst the smoke. Image 8 presents a scenario similar to Image 5. YOLOv9 correctly identified distinct flames but missed embers that were partially extinguished. Image 9 reveals that dark smoke obscured some flames, preventing their detection, even though fires in other regions were successfully identified. Image 10 shows successful detection in the upper left region, but false positives were introduced due to visual similarities between city lights and flames. Additionally, a road resembling fire in color and texture led to an incorrect detection. Image 11 shows a single-strip flame that was detected with multiple bounding boxes, introducing unnecessary computational overhead. A single bounding box would have sufficed for this scenario. Image 12 demonstrates the challenges posed by thick smoke, which masked flames to a significant degree. Despite this, YOLOv9 achieved a certain level of detection, but flames behind the smoke remained undetected. Image 13 shows the successful detection of prominent flames, though faint flames went unnoticed. Image 14 presents a heavily smoke-obscured scene. YOLOv9 detected distinct flames, but flames concealed by thick smoke were missed. Image 15 depicts widespread smoke that reduced image quality. No distinct flames were visible, and small flames on tree branches remained undetected.

These observations underscore the complexity of detecting forest fires in challenging environments. Environmental factors such as dense vegetation, thick smoke layers, and rugged terrain significantly hinder detection accuracy, leading to either incorrect classifications or partial failures. Furthermore, dynamic factors like the intensity and spread of the fire, as well as the presence of false positives in the field (e.g., city lights or roads resembling flames), pose additional challenges, increasing detection errors. Despite some deficiencies, these visuals reflect the inherent difficulty of detecting forest fires in such complex conditions.

Table 1 presents the comparative performance metrics for YOLOv4, YOLOv7, and YOLOv9. These models were evaluated based on sensitivity, recall, F1 score, mAP, and inference speed. YOLOv9 outperformed both YOLOv4 and YOLOv7 across all metrics, demonstrating superior precision, recall, mAP, and F1 score. Its sensitivity of 0.922 indicates the lowest error rate in positive predictions, surpassing YOLOv4 (0.866) and YOLOv7 (0.901). In terms of recall, YOLOv9 also excelled, achieving the highest success rate among the models. The F1 Score, which balances precision and recall, further reflects YOLOv9's superior performance. Speed is another critical factor, especially for real-time applications. YOLOv9 runs at 91 FPS, significantly faster than YOLOv7 (65 FPS) and YOLOv4 (37 FPS). This faster processing time makes YOLOv9 particularly advantageous in time-sensitive scenarios. The mAP@50 metric highlights YOLOv9's strong detection performance, with a value of 0.899, compared to 0.837 for YOLOv4 and 0.811 for YOLOv7. Similarly, YOLOv9 achieved the best mAP@50:95 score of 0.872, while YOLOv7 and YOLOv4 scored lower at 0.811 and 0.727, respectively. When considering both detection accuracy and processing speed, YOLOv9 stands out as the most robust option for real-time forest fire detection. While YOLOv7 exhibits solid performance, particularly in recall and F1 score, it falls short in speed and precision compared to YOLOv9. For scenarios that demand quick decision-making and accurate fire detection, the faster inference time and higher accuracy of YOLOv9 make it a strategically advantageous choice.

Table 1. Comparison of the performance metrics of the YOLO versions.

Model	Metrics					
	Precision	Recall	mAP@50	mAP@50:95	F1 Score	Speed
YOLOv4	0.866	0.809	0.837	0.727	0.836	37 FPS
YOLOv7	0.901	0.895	0.899	0.811	0.898	65 FPS
YOLOv9	0.922	0.925	0.915	0.872	0.923	91 FPS

IV. CONCLUSION

This study highlights the critical role of early and accurate detection strategies in controlling forest fires, which are increasingly influenced by climate change and human negligence, leading to widespread ecological damage. By integrating unmanned aerial vehicles (UAVs) equipped with high-capacity sensors and advanced cameras with YOLOv4, YOLOv7, and YOLOv9 algorithms, this research demonstrates significant advancements in detecting fires at their early stages. The high sensitivity and rapid detection capabilities of these algorithms, particularly in successful test cases, underline their potential for real-time forest fire monitoring. Despite the challenges posed by low visibility, dense smoke, and complex environmental conditions, YOLOv9 has shown particular promise due to its superior detection accuracy and processing speed. Its performance in real-time scenarios is especially noteworthy, making it a strong candidate for practical applications in forest fire prevention and management. However, the study also identifies areas for improvement, particularly regarding algorithmic errors and detection limitations in difficult conditions, such as false positives and missed detections in smoke-obscured regions. In conclusion, the integration of powerful computer vision algorithms, such as YOLOv9, with UAV technology and advanced hardware provides an effective and scalable approach to forest fire detection and management. The balanced optimization of these tools can significantly improve early fire detection, offering a vital solution to mitigate the devastating impact of forest fires on ecosystems and human communities alike. Future work should focus on refining these models to enhance their robustness in complex and rapidly changing fire environments.

REFERENCES

[1] P. Barmpoutis, P. Papaioannou, K. Dimitropoulos, N. Grammalidis, “A review on early forest fire detection systems using optical remote sensing,” *Sensors*, vol. 20, no. 6442, 2019.

[2] R.S. Priya, K. Vani, "Deep Learning Based Forest Fire Classification and Detection in Satellite Images," 2019 11th International Conference on Advanced Computing (ICoAC), Chennai, India, 2019, pp. 61-65.

[3] P. Mittal, R. Singh, A. Sharma, “Deep learning-based object detection in low-altitude UAV datasets: a survey,” *Image and Vision Computing*, 104, 104046, 2020.

[4] M. Bakirci, "A drone-based approach to enhance spatial insight into surrounding air pollutant distributions for healthier indoor environments," Journal of Building Engineering, vol. 87, no. 109023, 2024. https://doi.org/10.1016/j.jobe.2024.109023

[5] M. Bakirci, "Efficient air pollution mapping in extensive regions with fully autonomous unmanned aerial vehicles: A numerical perspective," Science of The Total Environment, vol. 909, no. 168606, 2024. https://doi.org/10.1016/j.scitotenv.2023.168606

[6] M. Bakirci, "Enhancing air pollution mapping with autonomous UAV networks for extended coverage and consistency," Atmospheric Research, vol. 306, no. 107480, 2024. https://doi.org/10.1016/j.atmosres.2024.107480

[7] M. Bakirci, I. Bayraktar, "Harnessing UAV technology and YOLOv9 algorithm for real-time forest fire detection," 2024 International Russian Automation Conference (RusAutoCon), pp. 95-100, Sochi, Russian Federation, 2024. https://doi.org/10.1109/RusAutoCon61949.2024.10694663

[8] M. Bakirci, B. Toptas, "Kinematics and autoregressive model analysis of a differential drive mobile robot," 4th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), pp. 1-6., Ankara, Turkey, 2022. https://doi.org/10.1109/HORA55278.2022.9800071

[9] Girshick, R. (2015). Fast R-CNN. *ArXiv*. https://arxiv.org/abs/1504.08083

[10] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *ArXiv*. https://arxiv.org/abs/1506.01497

[11] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., & Berg, A. C. (2015). SSD: Single Shot MultiBox Detector. *ArXiv*. https://doi.org/10.1007/978-3-319-46448-0_2

[12] Jeong, J., Park, H., & Kwak, N. (2017). Enhancement of SSD by concatenating feature maps for object detection. *ArXiv*. https://arxiv.org/abs/1705.09587

[13] Duan, K., Bai, S., Xie, L., Qi, H., Huang, Q., & Tian, Q. (2019). CenterNet: Keypoint Triplets for Object Detection. *ArXiv*. https://arxiv.org/abs/1904.08189

[14] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2015). You Only Look Once: Unified, Real-Time Object Detection. *ArXiv*. https://arxiv.org/abs/1506.02640

[15] P. Soviany and R. T. Ionescu, "Optimizing the Trade-Off between Single-Stage and Two-Stage Deep Object Detectors using Image Difficulty Prediction," 2018 20th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing (SYNASC), Timisoara, Romania, 2018, pp. 209-214.

[16] T. Diwan, G. Anirudh, J.V. Tembhurne, "Object detection using YOLO: challenges, architectural successors, datasets and applications," Multimed Tools Appl, vol. 82, pp. 9243–9275, 2023. https://doi.org/10.1007/s11042-022-13644-y

[17] Bochkovskiy, A., Wang, C., & Liao, H. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. *ArXiv*. https://arxiv.org/abs/2004.10934

[18] R. Niu, Y. Qu and Z. Wang, "UAV Detection Based on Improved YOLOv4 Object Detection Model," *2021 2nd International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE)*, Zhuhai, China, 2021, pp. 25-29.

[19] Wang, C., Yeh, I., & Liao, H. (2024). YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. *ArXiv*. https://arxiv.org/abs/2402.13616

[20] M. Bakirci, I. Bayraktar, "Transforming aircraft detection through LEO satellite imagery and YOLOv9 for improved aviation safety," 2024 26th International Conference on Digital Signal Processing and its Applications (DSPA), pp. 1-6, Moscow, Russian Federation, 2024. https://doi.org/10.1109/DSPA60853.2024.10510106