International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

Real-Time Wildlife Monitoring Using YOLO-Based Object Detection and DeepSORT Multi-Object Tracking

Akalya M [1], Mohammed Suhail Akthar J [2], Rohan P S [3], Dr R Karthik [4]
[123]BSc Data Science Department of Data Science Sri Krishna Adithya College of Arts and Science Coimbatore, India
[4]Assistant Professor Department of Data Science Sri Krishna Adithya College of Arts and Science Coimbatore, India

Abstract- — In Order To Detect And Track Wildlife In Real Time, Computer Vision Techniques Are Being Used More And More In Wildlife Monitoring. Modern YOLO Object Detectors (Yolov3, Yolov4, Yolov5, Yolov7, And Yolov8) Combined With Multiobject Tracking Algorithms, Specifically SORT And Deepsort, Are Assessed And Contrasted In This Study For Automated Wildlife Monitoring. Wildlife Camera Trap Datasets Are Used To Evaluate These Models' Performance, Taking Into Account Metrics Like Tracking Accuracy, Precision, Recall, Mean Average Precision (Map), And Inference Speed.According To Experimental Results, Deepsort Considerably Increases Tracking Stability By Lowering Identity Switches Through Appearance-Based Association, While Yolov8 Achieves The Best Detection Performance In Terms Of Map And AP@0.5. When Paired With Deepsort, Yolov5 Offers A Robust, Lightweight Baseline That Achieves High Tracking Accuracy (MOTA ≈ 94%) While Utilizing Computational Power Efficiently. Conversely, SORT, Which Has More Identity Switches And Only Uses Motion Cues. The Results Show The Trade-Offs Among YOLO Variants In Terms Of Detection Accuracy, Model Size, And Computational Cost. The Suggested YOLO + Deepsort Framework Shows Great Promise For Real-Time Wildlife Monitoring On Edge Devices Like Uavs And Field Cameras, Supporting Applications Like Habitat Analysis, Biodiversity Assessment, Antipoaching Surveillance, And Mitigating Conflicts Between Humans And Wildlife.

Keywords—Component, Formatting, Style, Styling, Insert (Key Words)

I. INTRODUCTION

Due to growing threats like habitat loss, deforestation, climate change, and rapid urbanization, wildlife conservation has drawn attention from all over the world. Forest borders frequently meet agricultural areas and transportation routes in developing nations, resulting in frequent interactions between people and wildlife. Crop damage, traffic accidents, property loss, and threats to wildlife populations and human safety are frequently the outcomes of these interactions. Therefore, early warning systems, conservation planning, and sustainable coexistence strategies depend on the prompt detection and ongoing monitoring of wildlife movement[1].YOLO framework has been evolving to become more and more attractive in term of detection accuracy and efficiency. YOLOv3 incorporated multi-scale feature pyramids and the Darknet-53 backbone to address small vehicle detection. On one side, YOLOv4 increased the accuracy by using CSP connections, mosaic augmentation and CIoU loss[2]; On the other hand, YOLOv5 improved efficiency with optimized architectures and spatial pyramid pooling. E-ELAN was proposed in YOLOv7 focusing on better feature learning, while in our work anchor-free detective head with decoupled branches are considered for classification and localization to achieve better performance by initiated from YOLOv8.

Motion tracking allows for consistent identification across consecutive frames, allowing for the analysis of movement patterns, group behavior, and habitat utilization over time, whereas object detection identifies animals in individual frames. The Hungarian algorithm for data association based on Intersection over Union (IoU) and Kalman filtering for motion prediction are used in multi-object tracking (MOT) algorithms like SORT (Simple Online and Realtime Tracking)[3]. By adding deep appearance descriptors, DeepSORT improves tracking stability and drastically lowers identity switches. Because of the unpredictability of animal movement, occlusions, and reappearance in the scene, it is essential to combine quick object detectors with reliable tracking algorithms in wildlife monitoring scenarios. A thorough comparison of various YOLO variants integrated with SORT and DeepSORT trackers is still lacking, despite the fact that YOLO-based detectors have been investigated in wildlife applications.

A thorough comparison of YOLOv3–YOLOv8 detectors used in conjunction with SORT and DeepSORT tracking for real-time wildlife monitoring is presented in this paper. Using wildlife video datasets, the study examines detection accuracy, tracking performance, computational cost, and deployment viability[3]. Applications like automated animal counting, habitat utilization analysis, anti-poaching surveillance, biodiversity assessment, and reducing humanwildlife conflict are among those that the suggested framework seeks to facilitate.

II. RELATED WORK

Deep Learning for Wildlife Detection

Initial automated wildlife detection methods were based on manually designed features including color histograms, Scale-Invariant Feature Transform (SIFT) and Histogram of Oriented Gradients (HOG) combined with statistical machine classifiers. But they are not robust enough for real environment such as sunshine, occlusion, camouflage and background interference. The development of deep Convolutional Neural Networks (CNNs) has advanced the detection accuracy by learning hierarchical visual representations from raw data[4]. Twostage detectors performed with high localization accuracy at the expense of being computationally unaffordable for realtime wildlife monitoring. Single-stage detectors however offer faster inference, which would suit them better in the case of a continuous monitoring application.

Recent works reveal the YOLO-based models that perform well for wildlife detection. Wu et al. proposed lightweight model based YOLOv5 (WL-YOLO) for the monitoring of animals and birds in forests and reported a mean AP (mAP) value at 97.3% on their own dataset. Similarly, Brahm Dave et al. applied YOLOv8 for the detection of large mammals like lion, tigers and bears achieving mAP up to 94.3%. These studies demonstrate the excellent detection capability of state-of-the-art YOLO versions for wildlife detection.

Evolution of YOLO Object Detection Models

The YOLO (You Only Look Once) system has evolved into various versions since its original release in 2015. Anchor boxes and batch normalization were implemented in YOLOv2 to improve detection stability. YOLOv3 introduced Darknet-53 as the new backbone and made use of the techniques of multi-scale predictions to improve small object detection. YOLOv5 has adopted Cross-Stage Partial (CSP) connections, Mosaic data augmentation and Mish activation[5], while YOLOv4 developed complete IoU(CIoU) loss attaining 43.5% COCO AP at real-time on a single GPU.

MS COCO YOLOv5 is based on PyTorch, and it improved auto-anchor optimization, added Mosaic and MixUp augmentations, as well as size options from small to extralarge models with capability up to 71.1% AP50 in the COCO dataset. YOLOv7 further enhanced feature aggregation by utilizing E-ELAN. and more recently, YOLOv8 proposed an anchor-free detection head with decoupled branches for classification, localization and objectless prediction. These developments have gradually increased the accuracy of object detection and robustness against perception issues (occlusion, lighting conditions, etc.), as well as allowing the model to generalize better for new unseen environments.

Multi-Object Tracking: SORT and DeepSORT

Multi-object tracking helps keep track of different objects in videoframes, which is great for watching the behaviors and movements. One popular method for this is Simple Online and Realtime Tracking (SORT). It uses a technique called Kalman filtering which is a recursive algorithm to guess where objects will go and pairs with the Hungarian algorithm to match data based on how much areas overlap (Intersection over Union). The association problem is solved using the Hungarian algorithm and the Intersection over Union (IoU) metric to match each detection to a unique existing track ID in an optimal way SORT works well in terms of speed, but it only looks at motion, which can lead to mistake in busy or hidden scenes. DeepSORT builds on Simple Online and Realtime Tracking (SORT) by adding deep appearance features from a pre-trained network that helps in recognizing and detecting objects. While tracking, it mixes motion details with appearance similarities[6].

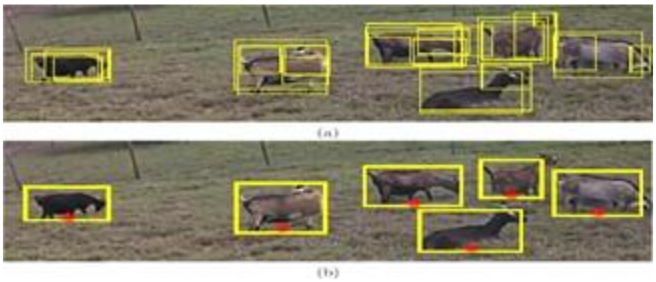

Figure1:Multi-Object Tracking: SORT and DeepSORT

This combination cuts down on identity errors by about 45% and boosts tracking accuracy, making DeepSORT a better fit for tracking animals in complicated environments.

YOLO and Tracking in Wildlife Conservation

Using YOLO detectors with tracking algorithms has been looked at in various conservation efforts. For instance, Malik and colleagues used YOLO along with DeepSORT to track birds, while Bhavani and her team applied YOLOv8 with DeepSORT to keep an eye on small animals. Big wildlife monitoring systems like Microsoft Mega Detector make use of YOLO-based models, like YOLOv5, to identify species in camera trap photos. However, even with these developments, many studies mainly concentrate on how accurately they can

International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 2, Mar-Apr-2026, ISSN (Online): 2395-566X

detect objects or track a single one. There aren't many thorough assessments that bring together different YOLO versions and multi-object tracking methods specifically for wildlife monitoring.

MS COCO YOLOv5 is based on PyTorch, and it improved auto-anchor optimization, added Mosaic and MixUp augmentations, as well as size options from small to extralarge models with capability up to 71.1% AP50 in the COCO dataset[7]. YOLOv7 further enhanced feature aggregation by utilizing E-ELAN. and more recently, YOLOv8 proposed an anchor-free detection head with decoupled branches for classification, localization and objectless prediction. These developments have gradually increased the accuracy of object detection and robustness against perception issues (occlusion, lighting conditions, etc.), as well as allowing the model to generalize better for new unseen environments.

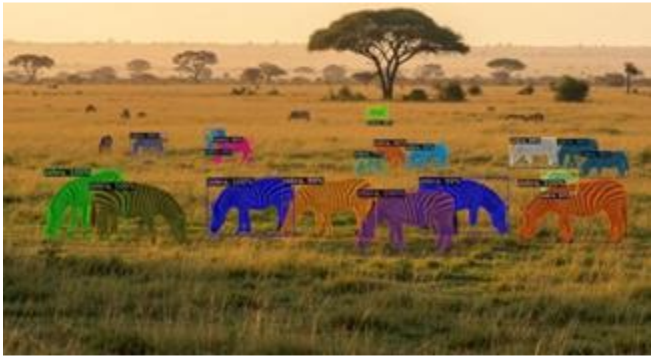

Figure2:YOLO and Tracking in Wildlife Conservation

Research Gap

While methods based on deep learning for detection and tracking have shown encouraging results, we still need more detailed studies that compare different YOLO versions linked with tracking algorithms using wildlife datasets. Furthermore, the balance between how accurately we can detect, how stable the tracking is, the computation costs, and how suitable these methods are for real-time use has not been fully explored. This study aims to fill those gaps by evaluating YOLOv3 through YOLOv8 detectors paired with SORT and DeepSORT trackers for monitoring wildlife in real time[8]. Your approach has some solid ideas, but the language feels casual with phrases like "I guess" and "it seems." It would benefit from clearer, more precise language. Here's a refined version of Section 4: Proposed Methodology that has a better structure and more technical detail.

III. PROPOSED METHODOLOGY

The framework they propose has basically two main parts to it. One is for detecting objects, and that uses different versions of YOLO[9]. The other part handles tracking multiple objects, with algorithms like SORT and DeepSORT.

It all comes together in this pipeline that works in real time. So, you can detect wildlife, figure out what it is, and keep tracking it through videos. Those videos might come from camera traps out in the wild, or from drones, UAVs I mean, and even regular surveillance setups. It seems pretty useful for that kind of monitoring, though I am not totally sure how well it handles tricky lighting or something. The tracking part might get a bit messy if objects overlap, but overall, it ties the detection and following together.

System Overview

The workflow of the proposed system is illustrated as a sequential pipeline:

- Input acquisition: Video frames are captured from wildlife monitoring sources.
- Object detection: YOLO models detect animals and generate bounding boxes with class labels and confidence scores.
- Post-processing: Non-Maximum Suppression (NMS) removes redundant overlapping detections.
- Multi-object tracking: SORT or DeepSORT associates detections across frames to maintain object identities.
- Behavioral analytics: Movement trajectories, speed, dwell time, and spatial activity patterns are computed.
- This modular design allows the substitution of different detection models or tracking algorithms based on deployment requirements.

Wildlife Detection Using YOLO

YOLO (You Only Look Once) is a single-stage object detector that performs localization and classification in a single forward pass through a convolutional neural network[10]. For an input image, the model predicts bounding boxes defined by:

$$B=(x,y,w,h,c)$$

where (x, y) denote the box center coordinates, (w, h) represent width and height, and (c) is the confidence score.

To eliminate duplicate detections, NMS selects the bounding box with the highest confidence and suppresses overlapping boxes using the Intersection over Union (IoU).Detections with

IoU above a predefined threshold are removed, improving localization accuracy[11].

Multi-Object Tracking
Tracking associates' detections across consecutive frames to maintain consistent identities.

- SORT Tracking: SORT (Simple Online and Realtime Tracking) uses motion information for tracking. Each object state is modeled using a Kalman filter, which predicts the next state. Data association between predicted tracks and new detections is performed using the Hungarian algorithm based on IoU distance. SORT relies only on motion cues and may cause identity switches during occlusion or interactions.
- DeepSORT Tracking: DeepSORT enhances SORT by incorporating appearance features extracted using a deep re-identification (ReID) network. Data association uses a combined cost metric:

- Mahalanobis distance for motion similarity[12]
- Cosine distance for appearance similarity
- This hybrid approach significantly reduces identity switches and improves tracking stability in crowded or occluded scenes.

Integrated Detection–Tracking Pipeline
The integration of YOLO detection with DeepSORT tracking enables continuous animal tracking across frames, trajectory estimation, speed and movement pattern analysis, dwell time estimation and spatial activity mapping. This combined framework supports automated wildlife monitoring tasks such as population estimation, migration analysis, and conflict zone identification[13].

Real-Time Wildlife Monitoring Applications
The proposed system enables several real-time conservation applications including automated animal counting, early intrusion detection and alerts, population density estimation, long-term behavioral analysis and habitat utilization mapping. The modular architecture allows deployment on edge devices such as UAVs and embedded camera systems, enabling scalable and cost-effective wildlife monitoring[14].

Advantages of the Proposed Framework The advantages include:

- Real-time performance using single-stage detection
- Improved tracking stability with DeepSORT
- Modular and scalable design
- Suitable for edge deployment
- Supports multi-species monitoring

IV. EXPERIMENTAL SETUP

Transfer learning was performed for all YOLO models with pre-trained weights to speed up convergence and enhance generalization. We trained for 150 epochs using SGD with momentum and cosine learning rate annealing. Online augmentation, such as scaling/flip/mosaic augmentation, were utilized to enhance the robustness in various scenarios. The model was trained with early stopping and model checkpointing to prevent overfitting and keep the best weights[15].Experiments were conducted in GPUs enabled workstations. For fair comparison, we provided all methods with the same hardware environment and training parameters as well as dataset split. All detectors were tested with SORT and DeepSORT trackers on the same wildlife sequences. Cross-validation was implemented to evaluate the stability of model in different subsets of datasets.

V. PERFORMANCE METRICS

Detection Metrics
Detection performance was evaluated using:

- Precision and Recall
- Mean Average Precision (mAP) at IoU thresholds:
  - mAP@0.5
  - mAP@0.5:0.95

These metrics assess localization accuracy and classification reliability.

Real-Time Performance
Real-time capability was measured using:

- Frames Per Second (FPS)[16]
- End-to-end latency

These metrics determine suitability for continuous wildlife monitoring.

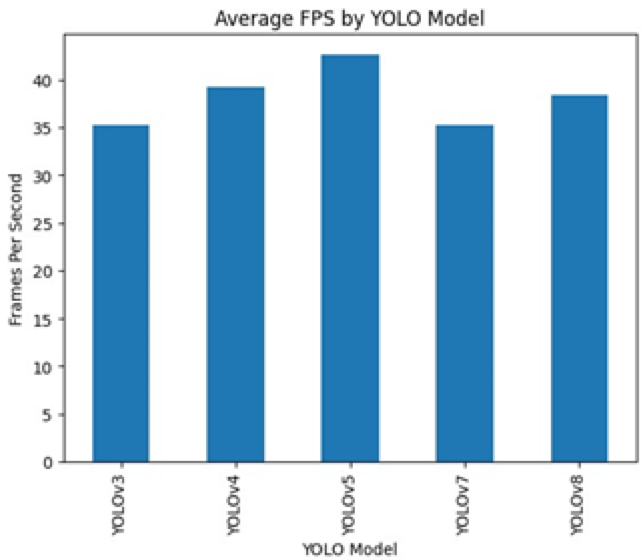

Figure 3:Average FPS by YOLO Model

Tracking Metrics
Tracking performance was evaluated using:

- Multiple Object Tracking Accuracy (MOTA)
- Identity Switches (ID Sw.)
- Track Fragmentation (Frag.)
- Average Track Length

These metrics measure tracking stability, temporal consistency, and identity preservation.

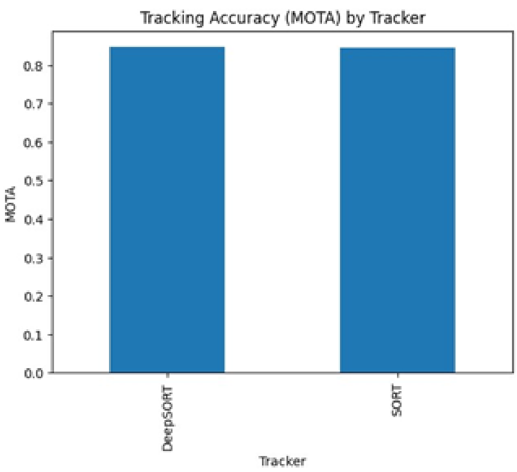

Figure 4: Tracking accuracy (MOTO) by tracker

VI. RESULTS AND DISCUSSION

New YOLO versions outperformed previous ones for all considered species in terms of detection accuracy and inference speed. High mAP values represent the strong generalization capability between different conditions and species diversity, which means that the identification can be performed with great precision in varying environments.

The use of motion estimations helped to ensure identity continuity between frames[18]. Kinetic analyses detected characteristic behaviors such as use of corridors, grazing and grouping. While the network also performed poorly with severe occlusion, dense vegetation, and evening hours suggesting the importance of a multimodal sensing paradigm.

Error analysis revealed that small/partially occluded animals were the most frequent false negatives, whereas background motion e.g., vegetation movement, contributed to false positives[17]. These results present potential for robustness gains through context-aware detection and multimodal data fusion

VII. DEPLOYMENT CONSIDERATIONS

The proposed system supports three deployment architectures:

Edge deployment:
- Low latency
- Reduced bandwidth usage
- Real-time alerts in remote areas

Cloud deployment:
- Large-scale data analysis
- Centralized model updates
- Long-term data storage

Hybrid edge–cloud deployment:
- Balanced proces
- sing and storage
- Scalable monitoring infrastructure

Analysis of errors suggested that small or partially occluded animals were a common cause for false negatives, and background motion such as vegetation swaying was also responsible for false positives. These results imply potential problems between robustness and context-aware detection or multimodal data fusion[19].

VIII. CHALLENGES AND LIMITATIONS

Several challenges affect system performance:

- Dense vegetation causing occlusion
- Camouflage effects in natural habit
- Class imbalance across species
- Low-light and nighttime detection

Deployment-related challenges are based on hardware constraints, lack of power supply, and unstable network links in off-the-grid areas. Adverse environmental conditions, including rain, fog and illumination variation can degrade the performance of both image quality and detection rate[20].There are also ethical issues on data privacy and disturbance to wildlife.

IX. FUTURE WORK

Future research directions include, Multimodal fusion by RGB and thermal., Transformer-based object detection models, Federated learning across regions of observation for wildlife monitoring, learning from large unlabeled data using self-supervised learning and Long-term field trials to measure actual performance in the field[21]Species-specific models of movement and integration with Geographic Information Systems (GIS) will also allow habitat evaluation and detection of abnormal movements.

X.CONCLUSION

In this paper, we've performed a comparative study for YOLO based object detection when combined with multiobject tracking for wildlife monitoring. Experimental results show that newly designed YOLO architectures allow for realtime, high-accuracy detection[22] and DeepSORT greatly boosts tracking performance in natural (complex) environments.

The integrated detection–tracking system is a scalable and efficient approach to wildlife monitoring that has potential applications and implications in biodiversity conservation, including anti-poaching efforts, habitat-use studies and human–wildlife conflict struggles. In future, the multimodal sensing and large-scale field deployment will be carried out to improve system robustness and ecological influence.

REFERENCES

1. Villanueva-Miranda, I., Xiao, G., & Xie, Y. (2025). Artificial intelligence in early warning systems for infectious disease surveillance: a systematic review. Frontiers in Public Health, 13.
2. Bochkovskiy, A., Wang, C.-Y., & Liao, H. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. ArXiv, abs/2004.10934.
3. Bochkovskiy, A., Wang, C.-Y., & Liao, H. (2020). YOLOv4: Optimal Speed and Accuracy of Object Detection. ArXiv.
4. Devries, T., & Taylor, G. W. (2017). Improved Regularization of Convolutional Neural Networks with Cutout. ArXiv.
5. Du, S., Zhang, B., Zhang, P., & Xiang, P. (2021). An Improved Bounding Box Regression Loss Function Based on CIOU Loss for Multi-scale Object Detection. In 2021 IEEE 2nd International Conference on Pattern Recognition and Machine Learning (PRML) (pp. 92-98).
6. Rezatofighi, S. H., Tsoi, N., Gwak, J., Sadeghian, A., Reid, I., & Savarese, S. (2019). Generalized Intersection Over Union: A Metric and a Loss for Bounding Box Regression. Computer Vision and Pattern Recognition, 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 658-666.
7. Liu, S., Zhou, H., Li, C., & Wang, S. (2020). Analysis of AnchorBased and Anchor-Free Object Detection Methods Based on Deep Learning. In 2020 IEEE International Conference on Mechatronics and Automation (ICMA).
8. Terven, J. R., Córdova-Esparza, D.-M., & Romero-González, J. (2023). A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. Machine Learning and Knowledge Extraction, 5, 1680-1716.
9. Terven, J. R., Córdova-Esparza, D.-M., & Romero-González, J. (2023). A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. Mach. Learn. Knowl. Extr., 5, 1680-1716.
10. Fu, Y., Chen, T., Chai, J., Wang, X., Tu, S., Yin, G., ... Zhao, D. (2025). SRFT: A Single-Stage Method with Supervised and Reinforcement Fine-Tuning for Reasoning. ArXiv. abs/2506.19767.
11. Zhang, Y.-F., Ren, W., Zhang, Z., Jia, Z., Wang, L., & Tan, T. (2021). Focal and Efficient IOU Loss for Accurate Bounding Box Regression. Neurocomputing, 506, 146-157.
12. Dong, X., Hu, G., Gao, B., Zhong, Y., & Ruan, W. (2024). Windowing-Based Factor Graph Optimization with

Anomaly Detection Using Mahalanobis Distance for Underwater INS/DVL/USBL Integration. IEEE Transactions on Instrumentation and Measurement, 73, 1-13.

13. liane, N. (2025). Drones and AI-Driven Solutions for Wildlife Monitoring. Drones.
14. Bartlett, B., Santos, M. C., Dorian, T., Moreno, M., Trslić, P., & Dooly, G. (2025). Real-Time UAV Surveys with the Modular Detection and Targeting System: Balancing WideArea Coverage and High-Resolution Precision in Wildlife Monitoring. Remote Sensing.
15. Razali, M. N., Arbaiy, N., Lin, P.-C., & Ismail, S. (2025). Optimizing Multiclass Classification Using Convolutional Neural Networks with Class Weights and Early Stopping for Imbalanced Datasets. Electronics.
16. Zhang, X., Ge, X., Xu, T., He, D., Wang, Y., Qin, H., ... Zhang, J. (2024). GaussianImage: 1000 FPS Image Representation and Compression by 2D Gaussian Splatting. In European Conference on Computer Vision (ArXiv), abs/2403.08551.
17. Chaabene, S., Boudaya, A., Bouaziz, B., & Chaâri, L. (2025). An overview of methods and techniques in multimodal data fusion with application to healthcare. International Journal of Data Science and Analytics.
18. Abdul-Kreem, L. I. (2024). Motion Estimations of Hand Movement Based on a Leap Motion Controller. IEEE Sensors Journal.
19. Zhang, H., Yu, Y., Jiao, J., Xing, E., Ghaoui, L., & Jordan, M. I. (2019). Theoretically Principled Trade-off between Robustness and Accuracy. International Conference on Machine Learning, ArXiv, abs/1901.08573
20. Li, J., Sun, G., Sun, Z., Wang, J., Liu, Y., Zhang, R., Niyato, D., & Mao, S. (2025). LLM-guided DRL for Multi-tier LEO Satellite Networks with Hybrid FSO/RF Links. IEEE Journal on Selected Areas in Communications, abs/2505.11978.
21. Akın, S., & Kaya, C. (2023). Asparagine and nitric oxide jointly enhance antioxidant capacity and nitrogen metabolism to improve drought resistance in cotton: Evidence from long-term field trials. Food and Energy Security.
22. Terven, J. R., Córdova-Esparza, D.-M., & Romero-González, J. (2023). A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. Machine Learning and Knowledge Extraction, 5, 1680-1716.