International Journal of Web of Multidisciplinary Studies
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
website: https://ijwos.com
Volume 3 Issue 4 , April 2026
E-ISSN : 3049-2424
DOI: https://doi.org/10.71366/ijwos

YOLO-Based Object Detection and Deep SORT Tracking for Remote Sensing Video Analysis

Prasad A[1], Rama Krishna K[2], Rahul G[3], Dr. B. Kalpana[4]
[1,2,3] *Student, Computer Science and Engineering, JNN Institute of Engineering, 1J.N.N Institute of Engineering, Chennai, India*
[4]*Head of the Department, Computer Science and Engineering, JNN Institute of Engineering, 1J.N.N Institute of Engineering, Chennai, India.*

Article Info

Article History:

Published:1 April 2026

Publication Issue:
Volume 3, Issue 4
April-2026

Page Number:
1-8

Corresponding Author:
Rahul G

Abstract:

Remote sensing video analysis has become an important research area in computer vision due to its wide applications in surveillance, traffic monitoring, and environmental monitoring. Traditional object tracking techniques struggle with complex scenes, occlusion, and dynamic backgrounds present in aerial videos. Recent advances in deep learning have enabled more accurate and efficient object detection and tracking methods.This research presents a framework for remote sensing video tracking using deep learning techniques. The proposed system integrates the YOLO (You Only Look Once) object detection model with the Deep SORT multi-object tracking algorithm. YOLO is used to detect objects in each frame of the video, while Deep SORT is used to track detected objects across consecutive frames while maintaining consistent identities.The system processes remote sensing videos by extracting frames, detecting objects, and associating them across frames using motion prediction and appearance feature matching. Experimental results demonstrate that the proposed method improves object detection accuracy and tracking performance in complex environments.This framework can be applied in several domains including UAV surveillance, traffic monitoring, disaster management, and security systems.

Keywords: Remote Sensing, Object Detection, Multi-Object Tracking, Deep Learning, YOLO, Deep SORT, Computer Vision, Video Surveillance, Aerial Video Analysis, Object Tracking

1. INTRODUCTION

Remote sensing technology plays a significant role in monitoring and analyzing large-scale environments. Remote sensing videos captured by satellites, drones, or unmanned aerial vehicles provide valuable information for surveillance, traffic monitoring, and environmental analysis.

However, analyzing remote sensing videos presents several challenges. Objects in aerial videos often appear small and may move rapidly across frames. In addition, environmental factors such as lighting variations, background clutter, and camera motion make object detection and tracking difficult.

Object tracking is an important task in computer vision that involves identifying and following objects across multiple video frames. Traditional tracking techniques rely on handcrafted features and motion estimation algorithms. These methods often fail in complex scenarios involving multiple objects or occlusions.

Copyright © 2025 The Author(s) : This is an open access article under the CC BY license (http://creativecommons.org/licenses/by-nc/4.0/)

Deep learning has significantly improved object detection and tracking capabilities. Convolutional neural networks have demonstrated excellent performance in detecting objects within images and videos. Among these models, YOLO has gained popularity for its ability to perform real-time object detection with high accuracy.

Multi-object tracking algorithms such as Deep SORT have further enhanced tracking performance by combining motion prediction and appearance feature extraction. By integrating YOLO and Deep SORT, it is possible to build an efficient system for detecting and tracking multiple objects in remote sensing videos.

The objective of this research is to develop a deep learning-based framework capable of detecting and tracking multiple objects in aerial video sequences

2. LITERATURE REVIEW

Several research studies have been conducted in the field of object detection and tracking using deep learning techniques.

Early object detection methods relied on traditional machine learning techniques such as Support Vector Machines and handcrafted feature extraction. These methods required significant manual feature engineering and often struggled with large-scale datasets.

With the advancement of deep learning, convolutional neural networks have become the dominant approach for object detection tasks. One of the most influential object detection frameworks is YOLO (You Only Look Once). YOLO treats object detection as a regression problem and processes the entire image in a single forward pass through a neural network. This allows the model to achieve high detection speed and real-time performance.

Researchers have proposed several improved versions of YOLO, including YOLOv3, YOLOv5, and YOLOv8, which provide better detection accuracy and faster processing speed.

In addition to object detection, multi-object tracking has received significant attention in computer vision research. Multi-object tracking aims to track multiple targets across video frames while maintaining consistent identities.

The SORT (Simple Online and Realtime Tracking) algorithm was one of the early approaches used for multi-object tracking. However, SORT relies only on motion information and cannot handle occlusion effectively.

To address this limitation, the Deep SORT algorithm was introduced. Deep SORT extends the SORT algorithm by incorporating appearance feature extraction using deep neural networks. This enables the system to distinguish between objects with similar motion patterns and reduces identity switching.

Recent studies have demonstrated that combining YOLO object detection with Deep SORT tracking provides accurate and efficient multi-object tracking in video sequences.

Despite these advances, several challenges remain in remote sensing video tracking, including small object detection, occlusion handling, and real-time performance constraints.

3. PROBLEM STATEMENT

Remote sensing videos contain complex visual environments that make object detection and tracking difficult.

Some of the major challenges include:

- Objects appearing at different scales
- Small object sizes in aerial images
- Background clutter and noise
- Camera motion from drones or satellites
- Occlusion between objects
- Maintaining consistent object identities

Traditional tracking algorithms often fail to handle these challenges effectively. Therefore, an efficient system that combines deep learning-based object detection with robust tracking algorithms is required.

The main goal of this research is to develop a framework capable of accurately detecting and tracking multiple objects in remote sensing videos.

4. Objectives

The primary objectives of this research are:

- To detect objects in remote sensing video frames using deep learning techniques.
- To track multiple objects across consecutive frames.
- To maintain consistent object identities during tracking.
- To reduce identity switching between tracked objects.
- To analyze object trajectories in aerial video sequences.

By achieving these objectives, the proposed system can improve the accuracy and efficiency of remote sensing video analysis.

5. PROPOSED METHODOLOGY

The proposed system integrates deep learning-based object detection with multi-object tracking algorithms.

The system processes remote sensing videos in several stages:

1. Video preprocessing
2. Frame extraction
3. Object detection using YOLO
4. Feature extraction
5. Motion prediction
6. Data association
7. Multi-object tracking using Deep SORT
8. Trajectory generation

Each stage plays an important role in ensuring accurate detection and tracking of objects across frames.

6. YOLO OBJECT DETECTION MODEL

YOLO is a deep learning-based object detection algorithm designed for real-time detection tasks.

Unlike traditional detection models that process images in multiple stages, YOLO processes the entire image in a single neural network pass. This significantly improves detection speed.

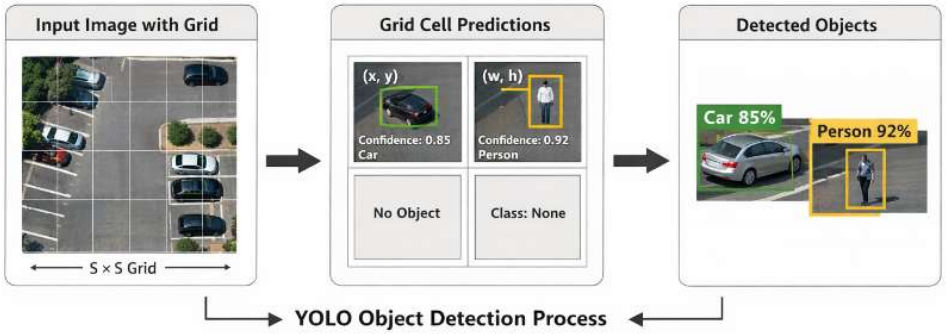
The YOLO algorithm divides the input image into a grid of cells. Each grid cell predicts bounding boxes and confidence scores indicating the presence of objects.

If an object is detected within a grid cell, the model predicts:

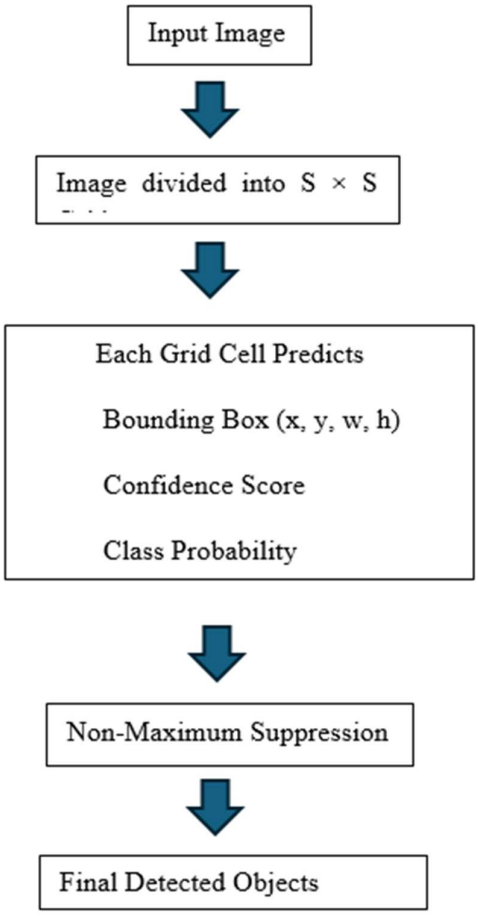
- Bounding box coordinates
- Object class
- Confidence score

After detection, a technique called Non-Maximum Suppression is applied to remove duplicate bounding boxes and retain the most accurate detection.

YOLO has become one of the most widely used object detection algorithms due to its speed and accuracy.



YOLO Grid Detection Architecture



Explanation

In the YOLO detection process, the input image is divided into multiple grid cells. Each grid cell predicts bounding boxes and object class probabilities. If the confidence score exceeds a predefined threshold, the object is detected. Non-Maximum Suppression removes overlapping bounding boxes to ensure accurate detection.

7. DEEP SORT TRACKING ALGORITHM

Deep SORT is a multi-object tracking algorithm that builds upon the SORT framework.

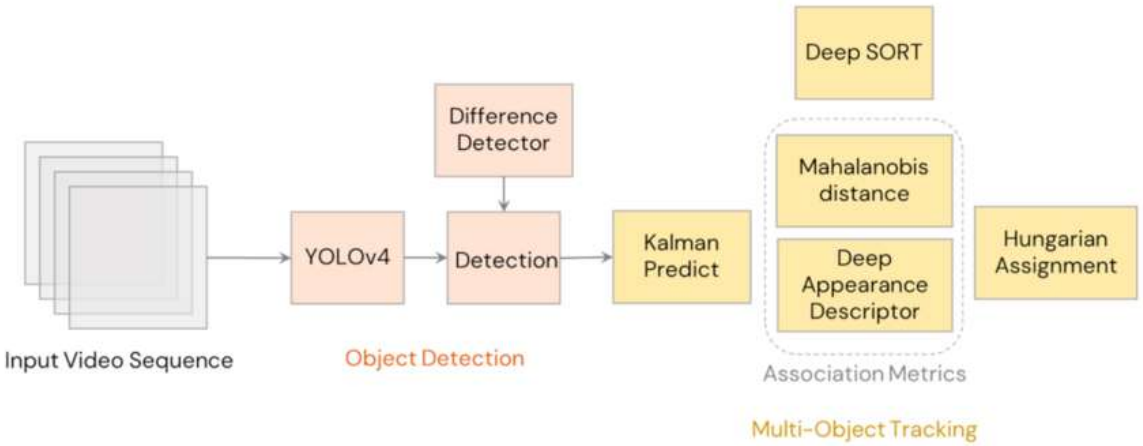
The algorithm combines motion prediction with appearance feature extraction to maintain consistent object identities across frames.

The tracking process consists of several steps:

1. Predict object positions using a Kalman Filter.
2. Extract appearance features using a deep neural network.
3. Match detected objects with existing tracks using the Hungarian Algorithm.

4. Assign unique IDs to tracked objects.

By combining motion and appearance information, Deep SORT improves tracking performance and reduces identity switching.

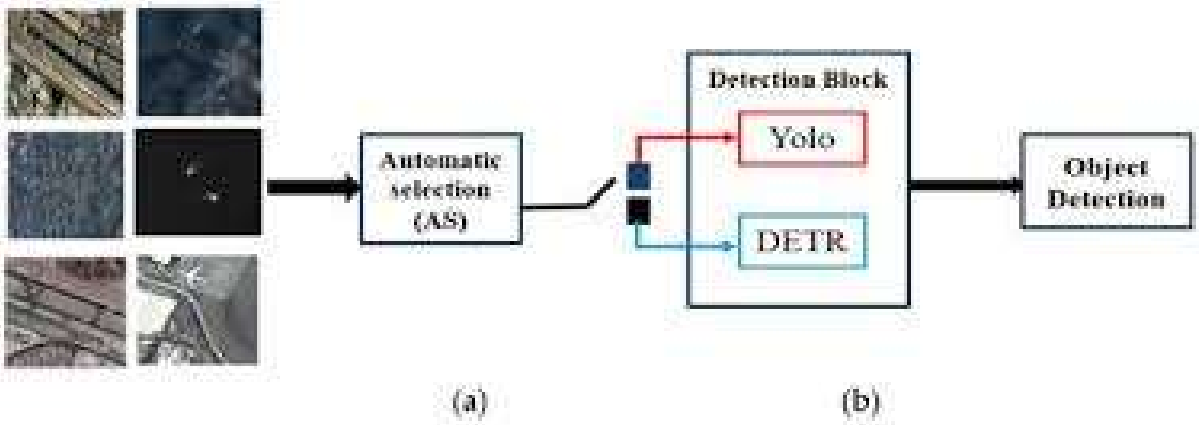


8. PROPOSED SYSTEM ARCHITECTURE

The proposed system integrates object detection and tracking modules to analyze remote sensing videos efficiently.

The system begins by receiving remote sensing video input captured by drones or satellites. The video is divided into frames and processed individually. Each frame is analyzed using the YOLO detection model to identify objects.

Detected objects are passed to the Deep SORT tracking module, which associates them across frames and assigns unique IDs. The system then generates trajectories representing the movement paths of tracked objects.



Explanation

The proposed architecture integrates YOLO object detection with Deep SORT multi-object tracking. YOLO detects objects in each video frame, while Deep SORT maintains object identities across frames using motion prediction and appearance feature matching.

9.APPLICATIONS

The proposed remote sensing video tracking system can be applied in several real-world scenarios.

Some important applications include:

- Traffic monitoring and vehicle tracking
- UAV surveillance systems
- Border security monitoring
- Disaster management and rescue operations
- Environmental monitoring
- Smart city surveillance systems

These applications demonstrate the importance of accurate object detection and tracking in remote sensing video analysis.

10.CONCLUSION

This research presented a deep learning-based framework for remote sensing video tracking using YOLO and Deep SORT algorithms.

The proposed system combines fast object detection with robust multi-object tracking to analyze aerial video sequences. YOLO provides efficient detection of objects in each video frame, while Deep SORT maintains consistent object identities across frames.

The integration of these techniques enables accurate tracking of multiple objects in complex environments.

Future research can focus on improving detection accuracy for small objects and optimizing the system for real-time deployment in large-scale surveillance applications.

References

[1] G. Rahul,K.Rama Krishna,A.Prasad,Kalpana.B "A Literature Survey on Remote Sensing Video Tracking Using Deep Learning Techniques," *International Journal of Engineering Development and Research (IJEDR)*, vol. 14, no. 2, pp. 115–118, Feb. 2026.

[2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 779–788.

[3] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv preprint arXiv:1804.02767, 2018.

[4] A. Bochkovskiy, C. Wang, and H. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv preprint arXiv:2004.10934, 2020.

[5] N. Wojke, A. Bewley, and D. Paulus, "Simple Online and Realtime Tracking with a Deep Association Metric," in *IEEE International Conference on Image Processing (ICIP)*, 2017, pp. 3645–3649.

[6] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple Online and Realtime Tracking," in *IEEE International Conference on Image Processing (ICIP)*, 2016.

[7] Y. Zhang, P. Sun, Y. Jiang, et al., "ByteTrack: Multi-Object Tracking by Associating Every Detection Box," in *European Conference on Computer Vision (ECCV)*, 2022.

[8] W. Liu, D. Anguelov, D. Erhan, et al., "SSD: Single Shot MultiBox Detector," in *European Conference on Computer Vision*, 2016.

[9] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137–1149, 2017.

[10] H. Wang and C. Schmid, "Action Recognition with Improved Trajectories," in *IEEE International Conference on Computer Vision (ICCV)*, 2013.

[11] D. Held, S. Thrun, and S. Savarese, "Learning to Track at 100 FPS with Deep Regression Networks," in *European Conference on Computer Vision*, 2016.

[12] M. Milan, L. Leal-Taixé, I. Reid, S. Roth, and K. Schindler, "MOT16: A Benchmark for Multi-Object Tracking," arXiv preprint arXiv:1603.00831, 2016.

[13] Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun, "YOLOX: Exceeding YOLO Series in 2021," arXiv preprint arXiv:2107.08430, 2021.

[14] L. Wen, D. Du, Z. Lei, et al., "UA-DETRAC: A New Benchmark and Protocol for Multi-Object Detection and Tracking," *Computer Vision and Image Understanding*, vol. 193, 2020.