African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

REAL-TIME VEHICLE MONITORING: A UNIFIED FRAMEWORK FOR DETECTION, TRACKING, AND BEHAVIOURAL CLASSIFICATION

Hussain Jumah, Z. [1], and Chyaid, A. M. [2]
[1,2] Department of Computer Science, College of Computer Science and Information Technology, University of Basra, Basra, Iraq.
[1]zahraa.h.jumah@uobasrah.edu.iq
[2]adala.gyad@uobasrah.edu.iq

ABSTRACT

Purpose: This paper proposes a unified framework that integrates YOLOv8s for accurate object detection and classification, DeepSORT for robust multi-object tracking, and an attention-based LSTM model for analysing temporal vehicle behaviours in urban environments.

Design/Methodology/Approach: The proposed framework was evaluated using the UAVDT dataset through a structured methodology. Initially, YOLOv8s was trained to detect and classify vehicles using appropriate preprocessing and training configurations. Subsequently, DeepSORT was employed to associate detected objects across frames and maintain consistent tracking identities. Temporal features extracted from object trajectories were then fed into the LSTM-Attention model to recognise vehicle behaviour patterns.

Research Limitations: The system's performance may be affected by class imbalance in the dataset and challenges in recognising transitional or ambiguous behaviours in highly complex traffic scenarios. Additionally, deployment on resource-constrained UAV platforms requires further optimisation.

Findings: Experimental results demonstrate strong performance, achieving an overall detection precision of 89.8%, a recall of 75.3%, and an mAP@50 of 82.2%. The DeepSORT tracker achieved robust identity preservation with an IDF1 score of 87.9%, even in dense urban environments. Furthermore, the behaviour recognition module achieved an overall F1-score of 0.93, confirming the effectiveness of the proposed system across various behavioural scenarios.

Practical Implications: The proposed framework can be effectively deployed in intelligent transportation systems and UAV-based monitoring platforms to enhance traffic management, improve surveillance efficiency, and support real-time decision-making.

Social Implications: The system helps reduce traffic accidents by enabling early detection of risky driving behaviours and supporting smart city surveillance systems, thereby improving public safety.

Originality/Value: The novelty of this work lies in integrating detection, tracking, and temporal behaviour analysis within a single unified framework, along with the use of an attention-based LSTM for improved behaviour recognition in real-world urban traffic scenarios.

Keywords: *Behaviour classification. DeepSORT. real-time monitoring. vehicle detection.YOLOv8*

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

INTRODUCTION

Context and Motivation

Road traffic accidents are still one of the main causes of death and invalidism for young people in developed as well as developing countries. Hence, timely and accurate accident detection is necessary to reduce the death toll and optimise emergency response operations (Kumar et al., 2020). In many industrialised countries, about 40% of daily accidents are fatal; in developing countries, it is even around 70% (Pemila et al., 2024). For example, the expansion of drivers in India has been sustained by inadequate surveillance and suboptimal driver registration systems (Sekhar et al., 2022). These are worrying statistics that demand advanced, scalable traffic monitoring.

In the last few years, large numbers of Applications Unmanned Aerial Vehicles (UAVs) have revolutionised smart transportation systems and urban surveillance (Wang et al., 2019). Their ability to record long-distance, high-resolution traffic videos from flexible aerial viewpoints offers the possibility of reducing collection time and overcoming the limitations of in-ground static cameras in collecting around-the-clock features. With respect to conventional sensing instruments and traditional manual investigations, UAV platform monitoring has the advantages of relatively low cost, large coverage area, and better suitability for real situations (Bartlett et al., 2025).

The growing use of UAVs has driven the global market's growth. It is reported by China/Financial Times that, as a regulatory authority, there are approximately 2.2 million registered drones in the Civil Aviation Administration of China (CAAC) (Financial Times, 2025)by the end of 2024. The drone world market, estimated at 30.7 billion dollars in 2024, will reach 74.8 billion dollars by 2033 with an annual growth rate of 10.41% based on AI-based navigation, smart-city traffic analytics and IoT(illustration) (IMARC Group, 2025).

Few studies provide a unified pipeline that integrates YOLOv8 detection, DeepSORT tracking, and LSTM-based temporal modelling. Behaviour recognition, particularly in identifying changes in vehicle behaviour, remains underexplored, with most work focusing only on static classification. Class imbalance in UAV datasets still dominates the performance of minority classes (bus, van, and truck). Most works heavily rely on over-sampling or augmentation techniques to address this problem, which is harmful to generalisation capacity. Achieving real-time efficiency (>30 FPS) while preserving high precision and recall across diverse conditions remains a significant challenge.

To overcome these limitations, this paper develops a new end-to-end integrated framework that fuses YOLOv8 for detection, DeepSORT for multi-object tracking, and LSTM with Attention for temporal behaviour modelling. This framework is developed to meet the requirements of detection, classification, tracking and behaviour recognition under UAV (Unmanned Aerial Vehicle) traffic surveillance constraints: high accuracy and real-time performance.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

Limitations of Current UAV Traffic Analysis Systems

Despite the substantial benefits of UAVs for traffic monitoring, there are a few technical barriers which hinder the current surveillance systems' performance (Wu et al., 2021).
Deterioration of detection and tracking performance: The accuracy of computer vision systems often decreases because of intensive lighting variation, motion blur, partial occlusion problems, as well as the diverse sizes of vehicles in a scene (Yusuf et al., 2024).

Insensitive to task integration: It is one of the largest bottlenecks because the integration of all tasks (detection, classification, tracking, and behaviour analysis) is poor. The majority of this work only targets one or two of these tasks, and/or focuses on real-time processing speed rather than accuracy. However, this compromise comes at the cost of limiting the ability to accommodate underrepresented classes or traffic situations due to its pooling operation (Rishika et al., 2023).

Dependence on traditional moving-object processing: Some existing detection and tracking systems treat moving areas as "motion pixels" without further differentiation of vehicle types or actions. This limited capability restricts the solution space for handling high-level (semantic) queries, such as "recognising illegal parked cars" or "recognising trucks waiting at traffic lights" (Wu et al., 2019).

Furthermore, class imbalance remains one of the main issues yet to be addressed. There are 66.5% of cars and less than 15% of trucks, buses and vans combined in the UAVDT dataset. This bias makes detectors biased towards the majority category, resulting in higher detection accuracy on cars and weaker performance on less representative vehicles (Ghosh et al., 2024).

Proposed Hybrid Framework

To tackle these challenges, this paper presents an integrated system that addresses the main drawbacks of current approaches and applies a customised hybrid model that combines the YOLOv8 detector for object detection, the DeepSORT algorithm for multi-object tracking, and an LSTM network enriched with an attention mechanism for real-time vehicle behaviour classification. By using balanced training techniques and level-wise temporal modelling, the system's objective is to reduce data imbalance effects without relying heavily on synthetic augmentation. The system is designed to achieve competitive accuracy across all vehicle classes, with real-time performance (>30 FPS) and low processing latency (<50 ms per frame).

The architecture integrates strong YOLOv8 object detection, identity-maintaining tracking with DeepSORT, and LSTM-based temporal reasoning to model the sequential behaviour of vehicles. As a result, it robustly copes with partial occlusion, varying scales and abundant behaviour transitions in complex traffic scenes. A thorough experiments on the UAVDT benchmark illustrate that our framework perform stably under various environment settings, which further verifies the effectiveness for intelligent traffic monitoring and autonomous navigation systems.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

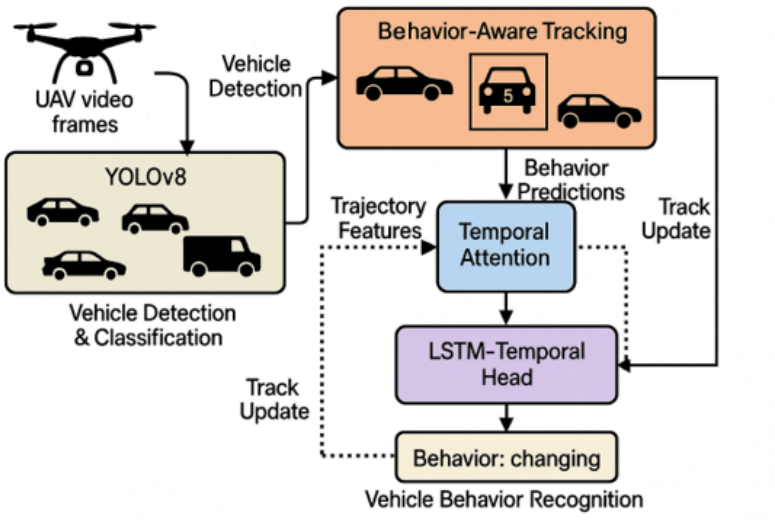

Figure 1: The Proposed Hybrid Framework

The main scientific and practical contributions of this paper, in the context of UAV-based traffic monitoring and intelligent transportation systems, are as follows:
1) Unified End-to-End Framework for Spatio-Temporal UAV Traffic Analysis
We present a fully integrated end-to-end framework that combines vehicle detection, classification, multi-object tracking, and temporal behaviour recognition in a single pipeline for UAV-based traffic monitoring. Unlike current methods, which deal with these tasks separately, the presented system builds a unified spatiotemporal representation that supports coherent reasoning from perception at the frame level to understanding the behaviour of trajectories at the sequence level in complex traffic scenes above.

2) Attention-Enhanced Temporal Modelling for Vehicle Behaviour Recognition
To reliably recognise dynamic vehicle movements from UAV trajectories, we also propose an LSTM-based temporal modelling module incorporating an attention mechanism. This architecture allows critical motion segments to be weighted adaptively so that the system can focus on behaviour-transition-related motion events, such as acceleration, deceleration, and changing direction. Thus, the proposed model performs much better on challenging or transitional behaviours than traditional temporal modelling methods.

3) Class-Imbalance-Aware Training Strategy without Data Oversampling
Extreme class imbalance was addressed by proposing loss reweighting and class-aware fine-tuning for underrepresented vehicle categories (bus, truck, van) in the UAV traffic dataset. In contrast to previous works that are either heavily dependent on oversampling or on synthetic

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

data augmentation, the proposed method enhances minority-class performance while remaining generalizable and unbiased.

4) Comprehensive Multi-Level Evaluation on UAVDT Benchmark
Extensive experiments are carried out on the UAVDT dataset, and to further analyse object detection performance, we use standard detection, tracking, and classification metrics: Precision, Recall, mAP (mean Average Precision), MOTA (multiple Object Tracking Accuracy), IDF1, and F1-score. The analysis reports quantitative class- and behaviour-specific performance information to establish an empirical standard for future UAV-based traffic analysis work.

5) Real-Time UAV Traffic Monitoring with Balanced Accuracy–Efficiency Trade-off
It can be implemented easily and delivers up to 30+ FPS in real time, with high sound detection accuracy, robust identity preservation, and dependable action recognition performance. Such a trade-off between accuracy and efficiency could make the system feasible for practical deployment in real-world UAV-assisted ITS and smart-city surveillance.
The efficacy of individual components is further verified by a thorough ablation study, which specifically shows the performance gain from attention-based temporal modelling, motion features, and class-imbalance-aware training separately.

OVERVIEW OF UAV-BASED VEHICLE ANALYSIS

With recent advances in computer vision, unmanned aerial vehicles (UAVs) have become powerful tools for traffic monitoring and intelligent transportation systems. However, vehicle detection and tracking and behaviour recognition in UAV videos are still challenging owing to scale variation, occlusion, illumination change, dynamic traffic, and severe class imbalance, where cars dominate over buses/vans/trucks. Issues of real-time and robustness to diverse real-world conditions need to be addressed with models that are both accurate and computationally efficient (Yang et al., 2022a).

Technical Challenges in Processing UAV Imagery

Recently, object detection from Unmanned Aerial Vehicle (UAV) imagery has gained more and more popularity due to its wide range of applications, including security surveillance (Loukinas, 2022), traffic monitoring, precision agriculture (Srivastava & Prakash, 2023), disaster management and aerial imaging (Hung et al., 2019). This increasing interest follows the natural progression of computer vision research, but traditional object detectors fine-tuned on ground-level images often experience a significant drop in accuracy when operating directly on aircraft imagery.

This degradation is largely caused by the extremely complex and dynamic natures of UAV scenes (e.g. changeable flight heights, diverse camera angles, motion blur issues and unpredictable light or weather conditions), which bring great challenges to detection models'

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

robustness and generalisation (Wu et al., 2019). While object detection and tracking have been studied for decades, and current detectors perform well in static ground-based environments, aerial videos present an additional set of challenges due to drone movement and large variations in object scale and viewpoint (Yu et al., 2020).

These challenges are summarised as follows:

- Hypsometric and scale differences: The scale of objects in the video scene from UAV relies directly on the flight height of the drone. For instance, if a DJI Inspire 2 is flying at approximately 500 meters in height, the vehicles look very tiny, and it becomes challenging to recognise them accurately and track (Li et al., 2021).
- Changes of viewpoint: As UAVs are able to fly in three-dimensional space, the same target may be observed from different viewpoints, front view, top view and side views, either vice versa, within a very short time and as a result, it is hard to maintain spatial consistency between frames (Bouguettaya et al., 2021).
- Lighting and weather conditions: As the data collected by aerial vehicles is in an open, on controll ed environment, it is influenced by varying lighting and weather (e.g., day/night, sunlight, fog or rain), which can alter the visual appearance of objects dramatically (Xia et al., 2018).

To address these problems, some annotated datasets of aerial image analysis have been released (e.g., DOTA (Zou et al., 2024), NWPU VHR-10 (Kumar et al., 2022) and VEDAI (Sommer et al., 2017)). These data are typically obtained from near-vertical (top-down) views, such as those seen in satellite images. But, they fail to model the diversity of altitude, viewpoint and scene dynamics in UAV footage, and fall behind real-world aerial complexities (Du et al., 2018).

To fill this gap, two main UAV-based benchmarks are designed:

- UAVDT (Zhu et al., 2018)Contains 100 video sequences (approximately 80K frames) collected under complex traffic conditions and provides weather, altitude, camera view , and bounding box annotations.
- VisDrone2018 (Zhu et al., 2018) : It is a large-scale dataset that contains more than 10,000 still images and over 179,000 frames extracted from 263 video clips of different scenarios (many objects to very crowded scenes) of a city, with diversity regarding both spaces and compositions, which calls for the detection and tracking problem.

Evolution of Object Detection and Tracking Methods

Early algorithms, such as Faster R-CNN and SSD, achieved excellent detection performance; however, they are computationally intensive and cannot be used in real time for UAVs (Zhu et al., 2024). The success of the YOLO family series of single-stage detectors has advanced the field in terms of speed and accuracy trade-off (Alrayes et al., 2025a). More recently, YOLOv8 was proposed with an anchor-free, decoupled detection head and a resize-invariant topology based on C2f modules to enhance precision (both precision and recall) without increasing latency (Sommer et al., 2017).

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

Meanwhile, multi-object tracking has developed from motion-based methods such as SORT to the most popular DeepSORT algorithm (Hanzla et al., 2024), which incorporates appearance embeddings to alleviate ID switches and enhance long-term identity retention in crowded traffic scenes, given their dynamic behaviour.

Deep Learning Approaches in UAV Traffic Surveillance

Recent works have extensively utilised deep learning (Deep Learning-based approaches) methodologies to improve the detection, tracing, and recognition of vehicles in aerial scenarios.

- (Al Mudawi et al., 2023)Combined, they fused YOLOv8 with a Deep Belief Network to detect and classify vehicles in aerial image sequences and achieved high accuracy across multiple classes, with the robustness of their approach verified in an urban UAV environment.
- (Yang et al., 2022b) introduced an advanced vehicle detection framework based on double-layer LSTM and explained how the use of temporal information could help maintain a consistent classification under dynamic traffic conditions.
- (Zhu et al., 2024)proposed a real-time YOLOv8 +SORT pipeline for UAV-based traffic monitoring on mobile platforms and obtained Precision =98.27%; Recall =87.93% at 30 frames per second (FPS), by also covering an analysis of urban traffic behaviours.
- (Alrayes et al., 2025b) proposed a fusion-based framework constructed from YOLOv8 to enhance long-term target tracking with large-scale variations, and realised robust detections and tracking accuracy in a UAV surveillance scene.
- (Hanzla et al., 2024) studied a vehicle recognition pipeline based on DeepSORT in aerial imagery and stressed the significance of consistent identity to preserve properties when vehicles overlap or occlude each other in dense traffic.

METHODOLOGY

Overview of The Proposed Approach

We would like to design an integrated framework for real-time vehicle detection, classification, and tracking from aerial videos captured by Unmanned Aerial Vehicles (UAVs) to achieve a suitable trade-off between high accuracy and computational cost. The framework developed integrates YOLOv8 for vehicle detection and classification in a single frame, DeepSORT to track objects across video frames, and an LSTM network that analyses temporal motion sequences to recognise vehicle behaviours (i.e., static, moving, or changing).

The general procedure is to first detect a person in each frame of the video using an object detector, then calculate the optical flow for the detected region, track it across frames, and finally let an LSTM learn these dynamics. Such a design allows the system to work reliably in challenging urban environments, cope with changes in lighting conditions, different camera viewpoints, and varying traffic densities, and maintain real-time performance.

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

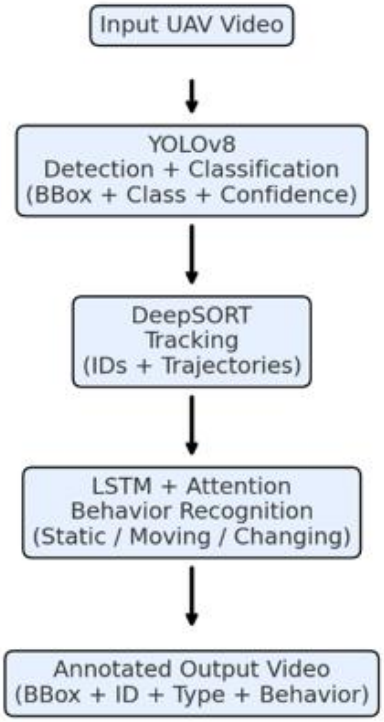

Figure 2: Overall architecture of the proposed system, which is based on YOLOv8 for detection and classification, DeepSORT for multi-object tracking and LSTM with Attention for behaviour recognition in UAV videos

In order to improve the effectiveness of object detection, this work utilises the YOLOv8 algorithm as a basic model for detection and classification, which is proven to be capable of real-time computer vision tasks while balancing between accuracy and processing speed (Zhao et al., 2023).There are several versions in the YOLOv8 family - i.e., YOLOv8n, YOLOv8s, YOLOv8m, YOLOv8l and YOLOv8x – which vary in complexity, accuracy and inference speed (Hermens, 2024). Unlike many prior works, which have been focused on light YOLOv8n (aimed at high speed at the expense of detection accuracy), we conducted a comprehensive assessment of YOLOv8s and the advanced version, YOLOv8x, on the UAVDT dataset.

According to the experimental results, while YOLOv8x achieved the highest accuracy, it was found to be too slow for real-time aerial applications. On the other hand, YOLOv8s achieved a more balanced performance, with much better accuracy than all prior works and maintaining real-time processing speed, hence being selected for the proposed framework.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

The architecture of YOLOv8 is based on single-stage object detectors that directly predict the bounding boxes and class probabilities in one pass without relying upon region proposal networks, and as a result, it produces more efficient inference (You et al., 2024). The YOLOv8 architecture contains three primary parts (Chang & Wang, 2024):

1) *Backbon*
   Extracts the hierarchical spatial features conditioned on C2f modules consists of Cross-Stage Partial (CSP) connections.
   This structure promotes the flow of gradients and the reusability of features while still achieving high computational efficiency.
2) *Neck*
   Based on an improved PANet, which is capable of aggregating feature maps from different scales to enhance detection robustness for small and partially occluded objects in UAV scenes.
3) *Head*
   Employs a Decoupled Head architecture, which decouples the regression and classification branches so that the network estimates localisation and class prediction in two independent paths for faster convergence and better accuracy.
   YOLOv8 model predicts the objectness score, bounding box coordinates and class probabilities at every grid location as:

$$p(obj) = \sigma(S\_o)$$
$$b = (x, y, w, h) = f\_anchor\text{-}free(S\_b)$$
$$p(c|obj) = softmax(S\_c)$$

where:
p(obj) denotes the probability that an object exists in a given cell, obtained through a sigmoid activation ***σ(·)*** applied to the detection score ***S_o***.

b = (x, y, w, h) represents the predicted bounding box parameters (center coordinates, width, and height), calculated via the anchor-free regression function ***f_anchor-free(·)***.

p(c|obj) represents the conditional probability distribution over object classes obtained using the ***softmax*** function on ***S_c*** (Bochkovskiy et al., 2020).

The **final detection confidence** for a class **c** is computed as:

$$Confidence(c) = p(obj) \times p(c|obj)$$

During training, YOLOv8 optimizes a composite loss function comprising three components:

$$L\_total = L\_box + L\_cls + L\_obj$$

where:
L_box is the bounding box regression loss, computed using **Complete IoU (CIoU)** or **Generalized IoU (GIoU)**:

$$L\_box = 1 - IoU(b\_pred, b\_gt)$$

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

L_cls is the classification loss (cross-entropy), and **L_obj** is the objectness loss (binary cross-entropy) (Zheng et al., 2020).

This unified design enables YOLOv8 to efficiently detect multiple vehicle classes (car, truck, bus, van) while maintaining high accuracy and frame rate in UAV-based aerial imagery.

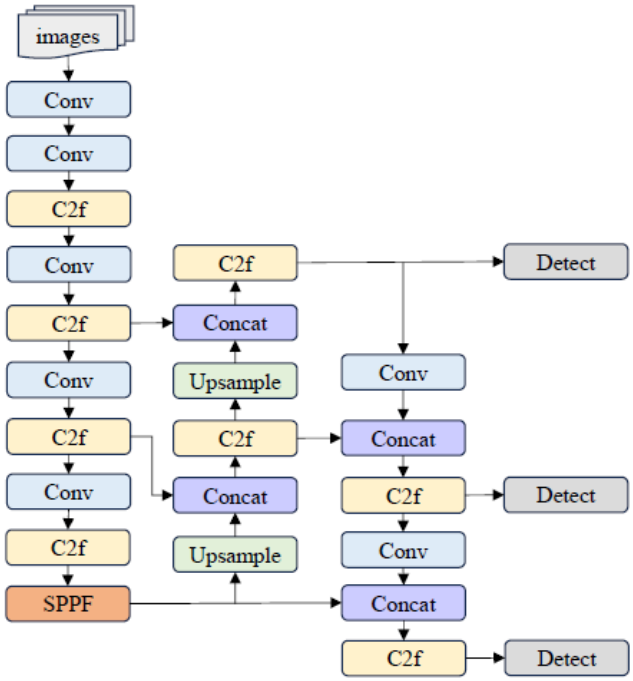

Figure 3: YOLOv8 Architecture Diagram(Liu et al., 2025a).

To demonstrate the practical detection capability of the proposed model, Figure 4 presents examples of detection results produced by the YOLOv8 model on the UAVDT dataset, where each vehicle is enclosed by a bounding box along with its corresponding class label and confidence score.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

Figure 4: Example of YOLOv8 detection and classification results on the UAVDT dataset

Tracking Module: Deepsort

The **DeepSORT (Deep Simple Online and Realtime Tracking)** algorithm is a highly efficient multi-object tracking method that combines motion prediction and appearance matching to maintain consistent object identities across video frames.

The **Kalman Filter** predicts the object's next position based on linear state transition equations:

$$x_k = F \times x_{k-1} + w_k$$
$$z_k = H \times x_k + v_k$$

where $\mathbf{x_k}$ represents the state vector (location, velocity, aspect ratio), **F** is the state transition matrix, and $\mathbf{w_k}$ and $\mathbf{v_k}$ are process and measurement noise, respectively(Kalman, 1960) .
The motion similarity between predicted and detected objects is measured using the **Mahalanobis Distance**(Wojke et al., 2017a):

$$d^1(i, j) = (y_j - \hat{y}_i)^T \times S_i^{-1} \times (y_j - \hat{y}_i)$$

and the **Cosine Distance** measures appearance similarity between deep feature embeddings (Du et al., 2023):

$$d^2(i, j) = 1 - (r_i^T \times r_j)$$

The two distances are fused into a single matching cost(Wojke et al., 2017b):

$$d(i, j) = \lambda \times d^1(i, j) + (1 - \lambda) \times d^2(i, j)$$

where $\lambda \in [0, 1]$ controls the trade-off between motion and appearance.

The obtained cost matrix is then minimised using the Hungarian Algorithm to match detections with existing tracks, thereby maintaining identity across frames. By the mathematical representation, it allows DeepSORT to achieve robust tracking performance with occlusions, vehicle congestions and rapid motion changes, as well as real-time efficiency (Yu et al., 2019).

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

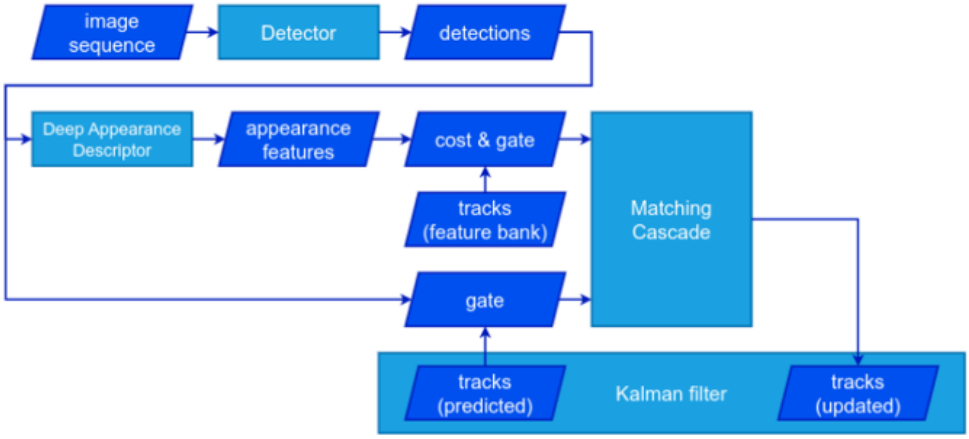

Figure 5: DeepSORT Architecture Diagram(Meimetis et al., 2023).

The following figure shows the capability to consistently track vehicle identity over more than 90 frames, highlighting its support for long-sequence tracking. Each vehicle is assigned a unique identifier that is preserved across multiple frames, even when occluded and interacts with other vehicles.

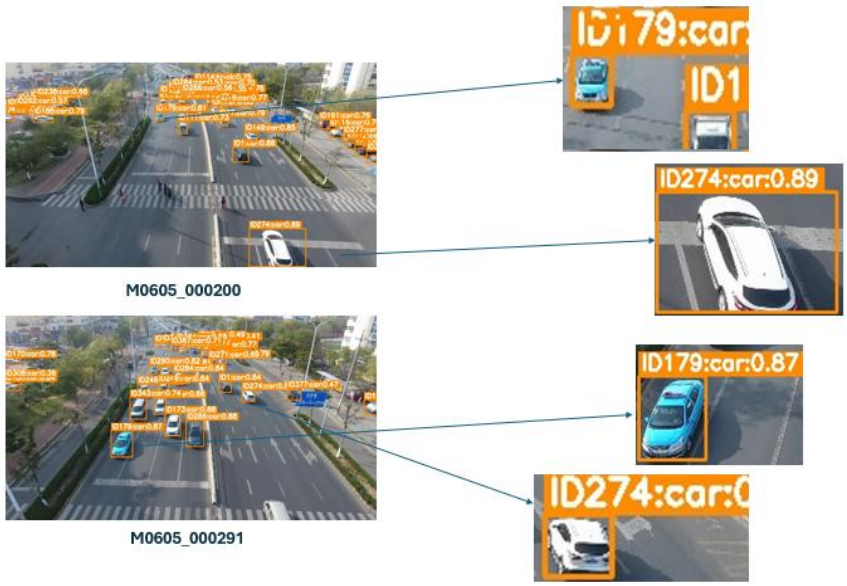

Figure 6: Example of video tracking using DeepSORT on the UAVDT benchmark.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

Behaviour Recognition: LSTM with Attention

Long Short-Term Memory **(LSTM)** network is one of the most powerful architectures for modelling sequences, because it is capable of learning long-term temporal dependencies without suffering from vanishing gradient problems found in traditional RNNs (Al-Selwi et al., 2024).

In the proposed scheme, an LSTM is used to capture temporal dependencies among frames from a vehicle's motion trajectory to learn a good representation of behaviour evolution.

The LSTM architecture is based on a gated-cell structure, which includes three main gates (output, input and forget) controlling the information propagation along time (Krichen & Mihoub, 2025).

The gates in each time step t are calculated as follows :

$$f_t = \sigma(W\_f \times [h_{t-1}, x_t] + b\_f)$$
$$i_t = \sigma(W\_i \times [h_{t-1}, x_t] + b\_i)$$
$$\hat{g}_t = tanh(W\_g \times [h_{t-1}, x_t] + b\_g)$$
$$o_t = \sigma(W\_o \times [h_{t-1}, x_t] + b\_o)$$

The cell state and hidden state are updated using:

$$c_t = f_t \odot c_{t-1} + i_t \odot \hat{g}_t$$
$$h_t = o_t \odot tanh(c_t)$$

where $\odot$ denotes element-wise multiplication, and $\sigma(\cdot)$ and $\tanh(\cdot)$ are nonlinear activation functions(Hochreiter & Schmidhuber, 1997).

This gating mechanism enables the network to selectively retain significant information while discarding irrelevant signals, making it particularly suitable for analysing temporal vehicle movement patterns.

To further improve classification accuracy and interpretability, an Attention Mechanism is integrated atop the LSTM layer.

While the LSTM captures the overall temporal dependencies, the Attention mechanism allows the model to selectively emphasise time steps that contribute most significantly to the final behaviour decision.

Given the sequence of hidden states from the LSTM, represented as:

$H = [h_1, h_2, \ldots, h\_T]$,

The attention weights are computed as:

$$e_t = v^T \times tanh(W_a \times h_t + b_a)$$
$$\alpha_t = exp(e_t) / \Sigma_i exp(e_i)$$

The context vector is then obtained as a weighted sum of all hidden states:

$$c = \Sigma_t \alpha_t \times h_t$$

This context vector is concatenated with the final hidden state h_T to produce a rich, attention-aware representation:

$$r = [c \oplus h\_T]$$

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

Finally, this vector is passed through a fully connected layer followed by a Softmax activation to generate the final behavioural class probabilities:

$$\hat{y} = softmax(W\_y \times r + b\_y)$$

By combining LSTM with the Attention mechanism, the model achieves adaptive temporal focusing, assigning higher importance to critical moments within a vehicle's motion sequence—such as sudden acceleration or directional changes—while down-weighting redundant or static frames(Bahdanau et al., 2014).

This hybrid LSTM–Attention structure provides an optimal balance between long-term dependency modelling and short-term selective reasoning, thereby enhancing both the accuracy and
interpretability of behaviour recognition in real-time UAV-based traffic analysis(Teo et al., 2024).

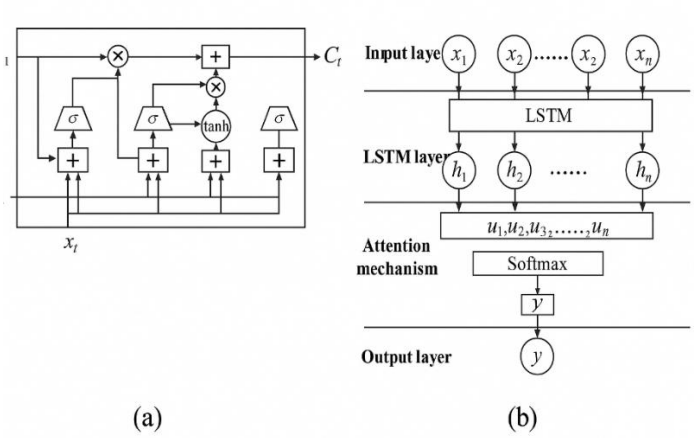

Figure 7 (A) illustrates the internal architecture of a standard LSTM network, which relies on input, output, and forget gates to regulate the flow of sequential information within the cell state. In contrast, Figure (B) presents an Attention-based LSTM model, where the attention mechanism enables the network to focus on the most relevant parts of the input sequence, enhancing performance in tasks that require long-term contextual understanding

Proposed Integrated Architecture:YOLOV8+ DEEPSORT + LSTM–Attention

The proposed system combines three parts, i.e., YOLOv8, DeepSORT and LSTM–Attention to provide a general pipeline for real-time vehicle detection, tracking and behaviour recognition in UAV-based aerial videos.

Each module abstractly carries out a sub-task and, as a result, composes a continuous pipeline of spatial-temporal information from the raw video frames to high-level behaviour understanding.

At each frame, the YOLOv8 model looks for all vehicles visible in that frame and predicts their bounding boxes, confidence scores, and class labels (car, truck, bus or van) in the first

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

stage. The detection outputs are fed to DeepSORT, which enforces temporal consistency by assigning each vehicle an individual identity ID in subsequent frames. DeepSORT uses a combination of Kalman filtering for trajectory prediction and re-identification features for visual re-identification, thereby allowing tracking of many objects even when actual object detection data is sometimes weaker (e.g., due to occlusions or fast camera motion).

The resulting dynamic trajectories, represented as per-frame vehicle position and velocity sequences as feature sets, are aggregated into a temporal representation and fed to the LSTM–Attention network.

Here, an LSTM layer is used to model temporal dependencies in vehicle movement, while the Attention mechanism highlights time steps that are most informative, i.e., when behavioural changes such as acceleration/deceleration or a change in direction occur.

The predicted behaviour class for each tracked vehicle (static, moving or changing) is the output of a final softmax operation in the network. This design is an efficient hybrid system between spatial detection and temporal comprehension:

YOLOv8 for accurate per-frame localisation, DeepSORT to track identities consistently and LSTM–Attention for context-aware behaviour reasoning.

Taken together, these components form an end-to-end system that is capable of analysing complicated urban traffic scenes in aerial UAVs with high accuracy and real-time speed.

Figure 8: The final outputs generated by the proposed hybrid system

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

Dataset

The proposed framework was trained and tested on the UAVDT dataset, a challenging, large-scale benchmark for vehicle detection and tracking in crowded urban traffic scenes captured by UAVs.

The set encompasses four main vehicle classes -car, truck, bus, van- that are acquired in different operating conditions, including differences in altitude, luminosity, weather conditions and traffic.

To further enrich the analysis, three complementary categories -- static, moving, and dynamic -- were introduced based on motion patterns extracted from car tracks. A particularly tough issue is the extreme class imbalance during UAVDT testing: there are significantly more cars than buses and vans, which makes uniform classification performance challenging.

The database is composed by 35 video sequences for training and 11 sequences for validation. Due to the lack of an official test set, we created a personal test split by selecting two sequences from the training set and one from the validation set.

Overall, the UAVDT has more than 80k frames and around 2.7M annotated bounding boxes, which is one of the largest datasets for vehicle detection and tracking in UAV data (Du et al., 2018).

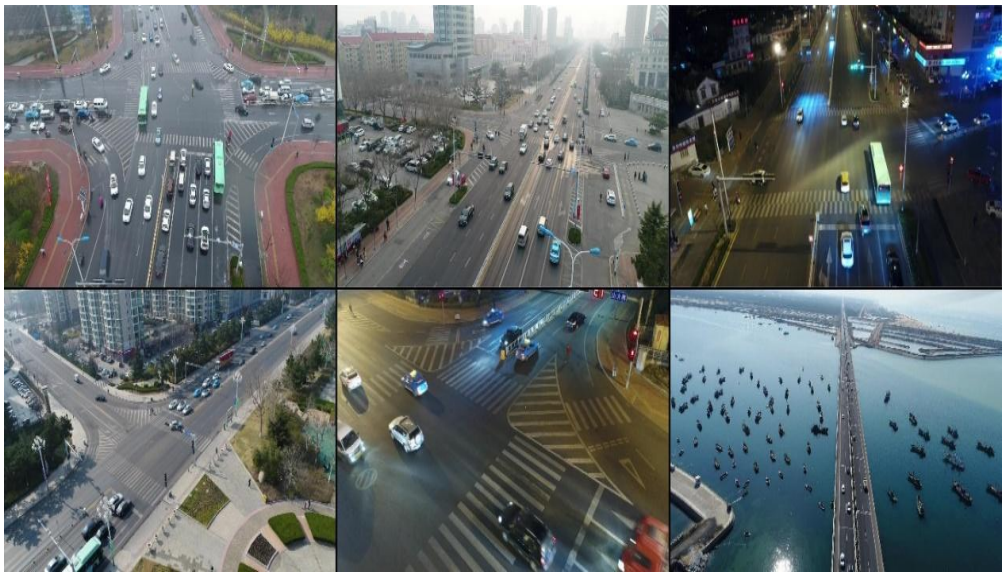

Plate 1: Sample images from different sequences in the UAVDT dataset

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

Data Preprocessing and Label Distribution Analysis

Before training, a structured preprocessing pipeline was applied to ensure consistency and improve model robustness. All images were resized to 640×640 pixels to match YOLOv8 input requirements, and all annotations were converted into the YOLO format *(class, x_center, y_center, width, height)*.

To increase generalisation capability, several augmentation techniques were applied, including Mosaic, horizontal flipping *(p = 0.5)*, HSV modifications *(hue = 0.015, saturation = 0.7, brightness = 0.4)*, and random translation and scaling to simulate changes in UAV altitude and viewpoint.

MixUp and CutMix were introduced in later phases to alleviate overfitting and improve performance on minority classes. To gain better insight into the dataset, we also performed a label distribution analysis. Figure Y show the distribution number of samples per class, it is clearly showing that there is an imbalance in the classes with a major dominance of car category. The correlogram for the bounding-box attributes is depicted in Figure Z (x, y, width, height), confirming broadly consistent patterns in the annotations.

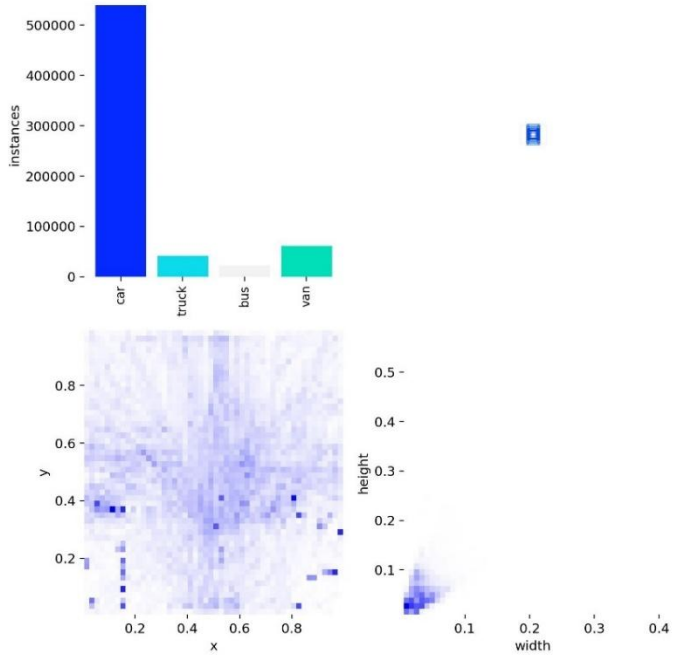

Figure 9: Class-wise label distribution in the UAVDT dataset

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

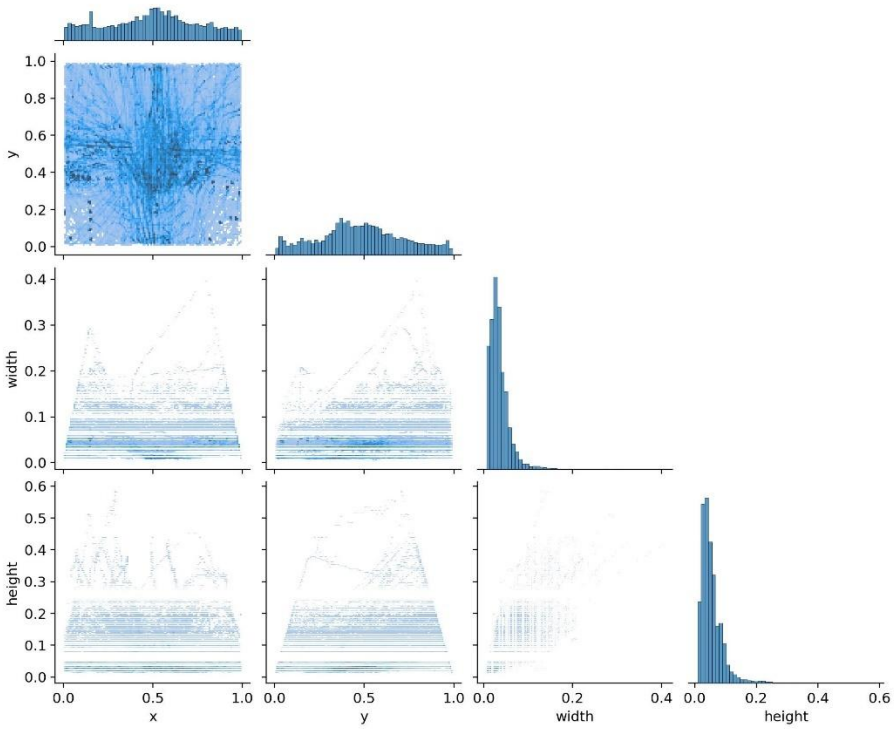

Figure 10: Bounding-box attribute correlogram (x, y, width, height)

Evaluation Metrics

For a quantitative evaluation of the proposed framework, we employ a set of comprehensive measurement metrics for each sub-task: detection, tracking , classification and behaviour analysis. These metrics are designed to be a consistent and interpretable measure of system performance, with attributes in the spatial, temporal, and behavioural domains.

1. *Detection Metrics (YOLOv8)*

Object detection performance is evaluated using Precision (P), Recall (R), and mean Average Precision (mAP) at different Intersection over Union (IoU) thresholds:

$$P = \frac{TP}{TP+FP} \ , \ R = \frac{TP}{TP+FP}$$

$$Confidence\ class = \text{p(obj)}\times\text{p(c|obj)}$$

(mAP@50) measures average precision at IoU = 0.5, while mAP@0.5–0.95 (mAP@50–95) provides a stricter evaluation by averaging precision across IoU thresholds from 0.5 to 0.95.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

These metrics are standard in YOLO-based object detection evaluation (Alsaihati et al., 2025).

2. *Tracking Metrics (DeepSORT)*
Multi-object tracking performance is assessed using MOTA (Multi-Object Tracking Accuracy), MOTP (Multi-Object Tracking Precision), IDF1, and the number of ID Switches (IDSW):

$$MOTA = 1 - \frac{FN+FP+IDSW}{GT}, \text{IDF1}=\frac{IDTP.2}{IDTP+IDFP+IDFN.2}$$

where GT is the number of ground-truth objects, and IDTP, IDFP, IDFN denote identity-level true positives, false positives, and false negatives. These metrics quantify both spatial tracking accuracy and identity preservation, and are widely used in multi-object tracking evaluation (Bernardin & Stiefelhagen, 2008).

3. *Classification Metrics (LSTM)*
For classification of vehicle types and behaviour categories, the framework reports Overall Accuracy (Acc), per-class Precision and Recall, as well as aggregated metrics:

- Macro Average: the unweighted mean across all classes, reflecting equal importance for each class
- Weighted Average: class-averaged metric weighted by the number of samples per class

$$F1 = 2\frac{P.R}{P+R}, ACC = \frac{TP+TN}{TP+TN+FP+FN}$$

The F1-score is also computed for each behavioural class (static, moving, changing) to provide a robust measure of recognition performance, particularly under class imbalance. Macro and weighted averaging allow a comprehensive evaluation of both rare and frequent classes. These evaluation measures are commonly used in machine learning evaluation classifiers(Estrada-Solano et al., n.d.).

The evaluation based on detection (mAP@50, mAP@50–95), tracking (MOTA, MOTP, IDF1, IDSW) and classification metrics (Macro Avg., Weighted Avg., F1-score) guarantees a comprehensive evaluation of spatial, temporal and behaviour performance and offers a solid benchmark for real-time UAV traffic monitoring systems.

Overall Pipeline
The entire process pipeline of the designed system consists of four stages that integrate into an end-to-end framework for real-time vehicle detection, tracking, and behaviour recognition using UAV-carried aerial videos.

1. *Detection and Classification:*
The vehicle detection results from this model are the class label (such as: car, truck, bus, van) and associated confidence score in each frame of video.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

2. *Tracking:*
   After detecting the objects, we implement the DeepSORT algorithm to maintain robust object identity across frames, including assigning unique IDs and motion tracks to each vehicle.
3. *Behaviour Recognition:*
   The obtained trajectories are then fed to an LSTM with Attention model, which models temporal dependencies and pays attention to key movement patterns, dividing each vehicle's behaviour into static/moving/changing.
4. *Output:*
   The final output is an annotated UAV video sequence containing bounding boxes, object IDs, class labels, and behaviour indicators for each detected and tracked vehicle, providing an interpretable visualisation of dynamic traffic behaviour in real time.

   This pipeline allows a consistent transmission and propagation of both spatial and temporal context information that further supports terrain understanding under aerial surveillance variations.

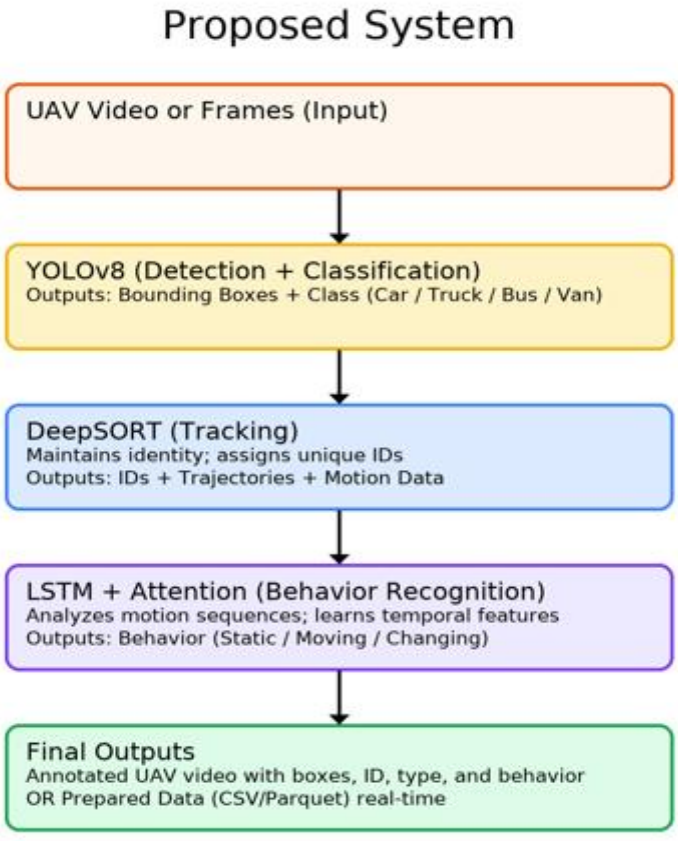

Figure 11: Overall Pipeline for the Proposed System

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

RESULTS AND EXPERIMENTS

Detection Performance

The detector YOLOv8x6 achieved high accuracy and excellent performance across the four vehicle categories (car, truck, bus, van) in terms of Precision, Recall-mAP@50, and mAP@50–95. However, YOLOv8x6 was computationally intensive despite being highly accurate. On the contrary, YOLOv8s achieved very good detection results and high real-time performance, so it was selected as our model in this paper.

Table 1: Comparison of the proposed YOLO model with state-of-the-art models on the UAVDT dataset

Detector	Precision	Recall	mAP@50	mAP@50–95
SRTSOD-YOLO-l (Xu et al., 2025)	-	-	47.2	28.7
RemDet-L† (C. Li et al., 2025)	-	-	34.5	20.6
LSOD-YOLOv8s (Liu et al., 2025b)	55.9	49.6	48.3	30.7
FB-YOLOv8 (Liu et al., 2022)	-	-	34.7	-
Our (yolov8s)	**89.8**	**75.3**	**82.2**	**52.6**
Our (yolov8x)	94.1	85.0	86.0	68.0

These findings contribute to the balanced trade-off between real-time and accurate classification, which has not been addressed simultaneously in several previous works.

Tracking Performance

The incorporation of DeepSORT greatly diminished identity swaps and confirmed that the object was traced through frames.

Table 2: Comparison of the proposed DeepSORT model with state-of-the-art models on the UAVDT dataset

Tracker	MOTA	MOTP	IDF1	IDSW
DeepSORT (Wojke et al., 2017c)	36.2	79.7	57.9	1626
UAVMOT (Ma et al., 2023)	46.4	72.7	67.3	456
AsyUAV (Ma et al., 2023)	48.0	-	67.5	349

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

BoT-SORT (Liu et al., 2022)	48.7	80.8	67.9	278
Our (DeepSORT)	**77.5**	**79.5**	**87.9**	**188**

These results show that the system can handle occlusions and traffic jams while preserving temporal consistency in real-world scenarios.

Behavioural Recognition Performance

The Attention mechanism-enhanced LSTM (LSTM with Attention) classified vehicle behaviour into three categories: static, moving, and changing.

Table 3: Evaluation of the proposed LSTM model on the UAVDT dataset

Class	Precision	Recall	F1-score	Support
Changing	0.92	0.98	0.95	308
Moving	0.97	0.86	0.91	125
Static	0.79	0.65	0.71	23
Accuracy	0.93	456		
Macro Avg	0.89	0.83	0.86	456
Weighted Avg	0.93	0.93	0.93	456

The recognition accuracy of the changing category (which is normally most difficult due to its dynamic variation and few samples) has been significantly improved by the Attention mechanism. Although many research works have been conducted to detect and classify vehicles with deep learning methods, most of them are based on a generic dataset or classifying the type of vehicle without considering vehicle behaviour states like Static, Moving, Changing, etc. As the UAVDT dataset is used in this study with a similar behaviour classification average precision, a comparison to existing works could not be conducted.

Real-Time Performance

The real-time performance of the framework was also tested by averaging the processing speeds in frames/second (FPS) under identical conditions to those applied for accuracy evaluation. We performed all experiments on NVIDIA A100 GPU (40 GB) with a batch size of 1 at the inference phase to simulate real-time deployment.

The system is found to carry out at an average processing speed of 32 FPS, which further ensures the real-time capability with respect to UAV traffic monitoring applications (video stream), where a minimum requirement is 30 FPS. This end-to-end demonstration covers the

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

entire pipeline: vehicle detection and classification with YOLOv8, multi-object tracking with DeepSORT, and temporal behaviour recognition via the proposed LSTM–Attention module.

A detailed analysis of the computation cost shows that the detector accounts for most of the processing load, while the tracker and behaviour recognition module do not introduce significant additional computational latency. The key point is that the temporal modelling module processes (asynchronously) the extracted trajectories, enabling effective parallel processing and causing little reduction in frame-level throughput.

Unlike current UAV-based traffic analysis solutions, our framework achieves a good trade-off between performance and efficiency. Although there are sometimes lightweight models that achieve high FPS at the cost of detection and tracking performance, or heavyweight models that achieve higher accuracy but cannot meet real-time requirements, our proposed model consistently performs well across all tasks with real-time processing.
These results indicate that our solution is applicable to real-world UAV-assisted ITS, emphasising timely and reliable traffic analysis.

Discussion

The experimental results demonstrate that the proposed framework achieves strong and consistent performance across detection, tracking, and behaviour recognition tasks, aligning with trends reported in recent literature.

Regarding object detection, the results indicate that the YOLOv8 architecture provides an effective balance between accuracy and computational efficiency. While the YOLOv8x variant achieved higher detection metrics (Precision: 94.1%, Recall: 85.0%, mAP@50: 86.0%), it imposed a significantly higher computational cost. In contrast, the YOLOv8s model achieved competitive performance (mAP@50: 82.2%, mAP@50–95: 52.6%) while maintaining real-time capability. This observation is consistent with previous studies such as SRTSOD-YOLO-l (Xu et al., 2025) and LSOD-YOLOv8s (Liu et al., 2025b), which show that lightweight detectors offer a better trade-off between speed and accuracy for UAV-based applications. The superior performance of the proposed model compared to existing methods highlights the effectiveness of the adopted training strategy and dataset-specific optimisation.

In terms of tracking performance, the integration of DeepSORT significantly improved identity preservation and temporal consistency. The obtained results (MOTA: 77.5, IDF1: 87.9, IDSW: 188) outperform several state-of-the-art approaches, including UAVMOT and BoT-SORT. This improvement can be attributed to the combination of motion and appearance features, which has been widely recognised as a key strength of DeepSORT in multi-object tracking tasks (Wojke et al., 2017). Moreover, the system demonstrated robustness in handling occlusions and dense traffic, although minor trajectory fragmentation may still occur during long-term occlusions.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

For behaviour recognition, the proposed LSTM enhanced with an attention mechanism achieved high classification performance, particularly for the "changing" class (F1-score: 0.95). This result is consistent with prior research emphasising the importance of temporal modelling in capturing motion dynamics. The attention mechanism further improved performance by focusing on the most relevant temporal features, aligning with findings from recent deep learning studies (Bond-Taylor et al., 2021). However, the relatively lower performance in the "static" class (F1-score: 0.71) can be attributed to class imbalance and the similarity between static and temporarily stationary vehicles, a challenge also reported in previous works.

It is important to note that direct comparison with existing studies in behaviour classification remains limited, as most previous research focuses primarily on vehicle detection or classification using generic datasets without incorporating temporal behaviour analysis. The use of the UAVDT dataset combined with a multi-stage pipeline (detection, tracking, and behaviour recognition) represents a notable contribution of this work.

Furthermore, the real-time evaluation confirms that the proposed system operates at an average speed of 32 FPS, satisfying real-time requirements for UAV-based traffic monitoring systems. This performance demonstrates a practical balance between computational efficiency and accuracy, which is often lacking in existing approaches that either prioritise accuracy at the expense of speed or vice versa.

Overall, the findings confirm that integrating detection, tracking, and temporal behaviour analysis within a unified framework improves system performance and robustness. The complementary strengths of YOLOv8 for detection, DeepSORT for tracking, and LSTM with attention for temporal modelling contribute to a reliable and efficient solution for intelligent traffic monitoring in UAV-based environments.

CONCLUSION

This paper introduced the design of an integrated real-time working system for vehicle detection, classification, tracking and behaviour recognition in UAV-based traffic monitoring. With YOLOv8s for precise detection and categorisation, DeepSORT for stable multi-object tracking, and LSTM+Attention to analyse temporal behaviour patterns, the system filled a gap in existing works that often modelled these problems independently. All these were verified through experimental results, indicating that the proposed framework strikes a good balance between accuracy and speed and can run at a rate above real-time, thereby enhancing performance on challenging classes/behaviours. Unlike the accuracy- or efficiency-focused methods, our method achieves both and can be applied to intelligent transportation systems and urban surveillance devices.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

Future work involves fine-tuning the framework for embedded UAV platforms, handling class imbalance with more advanced learning methods, and testing it on additional UAV datasets. In summary, the study indicates that combining the detection, classification, and tracking stages into a single pipeline offers an effective methodology for real-world UAV-based monitoring.

REFERENCES

Al Mudawi, N., Qureshi, A. M., Abdelhaq, M., Alshahrani, A., Alazeb, A., Alonazi, M., & Algarni, A. (2023). Vehicle detection and classification via YOLOv8 and deep belief network over aerial image sequences. *Sustainability*, *15*(19), 14597.

Alrayes, F. S., Ahmad, N., Alshuhail, A., Alshammeri, M., Alqazzaz, A., Alkhiri, H., Alqurni, J. S., & Said, Y. (2025a). Convolutional transform learning based fusion framework for scale invariant long term target detection and tracking in unmanned aerial vehicles. *Scientific Reports*, *15*(1), 28248.

Alrayes, F. S., Ahmad, N., Alshuhail, A., Alshammeri, M., Alqazzaz, A., Alkhiri, H., Alqurni, J. S., & Said, Y. (2025b). Convolutional transform learning based fusion framework for scale invariant long term target detection and tracking in unmanned aerial vehicles. *Scientific Reports*, *15*(1), 28248.

Alsaihati, F., Aldossary, H., Alzamil, R., Almadan, R., Al Mousa, Z., & Alahmadi, A. (2025). Waste Classification and Detection Model Using YOLOv8 for Waste Management. In *Integrating Big Data and IoT for Enhanced Decision-Making Systems in Business: Volume 2* (pp. 473–484). Springer.

Al-Selwi, S. M., Hassan, M. F., Abdulkadir, S. J., Muneer, A., Sumiea, E. H., Alqushaibi, A., & Ragab, M. G. (2024). RNN-LSTM: From applications to modeling techniques and beyond—Systematic review. *Journal of King Saud University-Computer and Information Sciences*, *36*(5), 102068.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *arXiv Preprint arXiv:1409.0473*.

Bartlett, B., Santos, M., Dorian, T., Moreno, M., Trslic, P., & Dooly, G. (2025). Real-time UAV surveys with the modular detection and targeting system: Balancing wide-area coverage and high-resolution precision in wildlife monitoring. *Remote Sensing*, *17*(5), 879.

Bernardin, K., & Stiefelhagen, R. (2008). Evaluating multiple object tracking performance: The clear mot metrics. *EURASIP Journal on Image and Video Processing*, *2008*(1), 246309.

Bochkovskiy, A., Wang, C.-Y., & Liao, H.-Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv Preprint arXiv:2004.10934*.

Bouguettaya, A., Zarzour, H., Kechida, A., & Taberkit, A. M. (2021). Vehicle detection from UAV imagery with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(11), 6047–6067.

Chang, H., & Wang, Z. (2024). *UAV vehicle detection system based on YOLOv8*. *2872*(1), 012019.

Du, D., Qi, Y., Yu, H., Yang, Y., Duan, K., Li, G., Zhang, W., Huang, Q., & Tian, Q. (2018). *The unmanned aerial vehicle benchmark: Object detection and tracking*. 370–386.

Du, Y., Zhao, Z., Song, Y., Zhao, Y., Su, F., Gong, T., & Meng, H. (2023). Strongsort: Make deepsort great again. *IEEE Transactions on Multimedia*, *25*, 8725–8737.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

Estrada-Solano, F., Caicedo, O. M., & S da Fonseca, N. L. (n.d.). An Approach Based on Incremental Deep Learning and Traffic-Flow Characteristics for Scheduling Elephant Flows in Software-Defined Data Center Networks. *Nelson L., An Approach Based on Incremental Deep Learning and Traffic-Flow Characteristics for Scheduling Elephant Flows in Software-Defined Data Center Networks*.

Financial Times, F. T. (2025). *China made millions of drones. Now it has to find uses for them*. https://www.ft.com/content/65e4bb30-b04c-4f0e-8686-effb96398999

Ghosh, K., Bellinger, C., Corizzo, R., Branco, P., Krawczyk, B., & Japkowicz, N. (2024). The class imbalance problem in deep learning. *Machine Learning*, *113*(7), 4845–4901.

Hanzla, M., Yusuf, M. O., Al Mudawi, N., Sadiq, T., Almujally, N. A., Rahman, H., Alazeb, A., & Algarni, A. (2024). Vehicle recognition pipeline via DeepSort on aerial image datasets. *Frontiers in Neurorobotics*, *18*, 1430155.

Hermens, F. (2024). Automatic object detection for behavioural research using YOLOv8. *Behavior Research Methods*, *56*(7), 7307–7330.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780.

Hung, I.-K., Unger, D., Kulhavy, D., & Zhang, Y. (2019). Positional precision analysis of orthomosaics derived from drone captured aerial imagery. *Drones*, *3*(2), 46.

IMARC Group, I. G. (2025). *Drones market size, share, trends and forecast by type, component, payload, point of sale, end-use industry, and region 2025–2033*. https://www.imarcgroup.com

Kalman, R. E. (1960). *A new approach to linear filtering and prediction problems*.

Krichen, M., & Mihoub, A. (2025). Long short-term memory networks: A comprehensive survey. *AI*, *6*(9), 215.

Kumar, N., Acharya, D., & Lohani, D. (2020). An IoT-based vehicle accident detection and classification system using sensor fusion. *IEEE Internet of Things Journal*, *8*(2), 869–880.

Kumar, S., Jain, A., Rani, S., Alshazly, H., Idris, S. A., & Bourouis, S. (2022). Deep Neural Network Based Vehicle Detection and Classification of Aerial Images. *Intelligent Automation & Soft Computing*, *34*(1).

Li, C., Zhao, R., Wang, Z., Xu, H., & Zhu, X. (2025). *Remdet: Rethinking efficient model design for uav object detection*. *39*(5), 4643–4651.

Li, X., Li, X., Li, Z., Xiong, X., Khyam, M. O., & Sun, C. (2021). Robust vehicle detection in high-resolution aerial images with imbalanced data. *IEEE Transactions on Artificial Intelligence*, *2*(3), 238–250.

Liu, S., Li, X., Lu, H., & He, Y. (2022). *Multi-object tracking meets moving UAV*. 8876–8885.

Liu, S., Shen, X., Xiao, S., Li, H., & Tao, H. (2025a). A Multi-Scale Feature-Fusion Multi-Object Tracking Algorithm for Scale-Variant Vehicle Tracking in UAV Videos. *Remote Sensing*, *17*(6), 1014.

Liu, S., Shen, X., Xiao, S., Li, H., & Tao, H. (2025b). A Multi-Scale Feature-Fusion Multi-Object Tracking Algorithm for Scale-Variant Vehicle Tracking in UAV Videos. *Remote Sensing*, *17*(6), 1014.

Loukinas, P. (2022). Drones for border surveillance: Multipurpose use, uncertainty and challenges at EU borders. *Geopolitics*, *27*(1), 89–112.

Ma, J., Liu, D., Qin, S., Jia, G., Zhang, J., & Xu, Z. (2023). An Asymmetric Feature Enhancement Network for Multiple Object Tracking of Unmanned Aerial Vehicle. *Remote Sensing*, *16*(1), 70.

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

Meimetis, D., Daramouskas, I., Perikos, I., & Hatzilygeroudis, I. (2023). Real-time multiple object tracking using deep learning methods. *Neural Computing and Applications*, *35*(1), 89–118.

Pemila, M., Pongiannan, R., Narayanamoorthi, R., Sweelem, E. A., Hendawi, E., & El-Sebah, M. I. A. (2024). Real-time classification of vehicles using machine learning algorithm on the extensive dataset. *IEEE Access*, *12*, 98338–98351.

Rishika, A. L., Aishwarya, C., Sahithi, A., & Premchender, M. (2023). Real-time vehicle detection and tracking using yolo-based deep sort model: A computer vision application for traffic surveillance. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *14*(1), 255–264.

Sekhar, G. B., Srilatha, M., Srinivasulu, J., & Chowdary, M. B. (2022). *Vehicle Tracking and Speed Estimation Using Deep Sort*. 146–151.

Sommer, L. W., Schuchert, T., & Beyerer, J. (2017). *Fast deep vehicle detection in aerial images*. 311–319.

Srivastava, A., & Prakash, J. (2023). Techniques, answers, and real-world UAV implementations for precision farming. *Wireless Personal Communications*, *131*(4), 2715–2746.

Teo, T.-A., Chang, M.-J., & Wen, T.-H. (2024). Automatic vehicle trajectory behavior classification based on unmanned aerial vehicle-derived trajectories using machine learning techniques. *ISPRS International Journal of Geo-Information*, *13*(8), 264.

Wang, J., Simeonova, S., & Shahbazi, M. (2019). Orientation-and scale-invariant multi-vehicle detection and tracking from unmanned aerial videos. *Remote Sensing*, *11*(18), 2155.

Wojke, N., Bewley, A., & Paulus, D. (2017a). *Simple online and realtime tracking with a deep association metric*. 3645–3649.

Wojke, N., Bewley, A., & Paulus, D. (2017b). *Simple online and realtime tracking with a deep association metric*. 3645–3649.

Wojke, N., Bewley, A., & Paulus, D. (2017c). *Simple online and realtime tracking with a deep association metric*. 3645–3649.

Wu, X., Li, W., Hong, D., Tao, R., & Du, Q. (2021). Deep learning for unmanned aerial vehicle-based object detection and tracking: A survey. *IEEE Geoscience and Remote Sensing Magazine*, *10*(1), 91–124.

Wu, Z., Suresh, K., Narayanan, P., Xu, H., Kwon, H., & Wang, Z. (2019). *Delving into robust object detection from unmanned aerial vehicles: A deep nuisance disentanglement approach*. 1201–1210.

Xia, G.-S., Bai, X., Ding, J., Zhu, Z., Belongie, S., Luo, J., Datcu, M., Pelillo, M., & Zhang, L. (2018). *DOTA: A large-scale dataset for object detection in aerial images*. 3974–3983.

Xu, Z., Zhao, H., Liu, P., Wang, L., Zhang, G., & Chai, Y. (2025). SRTSOD-YOLO: stronger real-time small object detection algorithm based on improved YOLO11 for UAV imageries. *Remote Sensing*, *17*(20), 3414.

Yang, W.-J., Liow, W.-J., Chen, S.-F., Yang, J.-F., Chung, P.-C., & Mao, S. (2022a). Improved vehicle detection systems with double-layer LSTM modules. *EURASIP Journal on Advances in Signal Processing*, *2022*(1), 7.

Yang, W.-J., Liow, W.-J., Chen, S.-F., Yang, J.-F., Chung, P.-C., & Mao, S. (2022b). Improved vehicle detection systems with double-layer LSTM modules. *EURASIP Journal on Advances in Signal Processing*, *2022*(1), 7.

You, L., Chen, Y., Xiao, C., Sun, C., & Li, R. (2024). Multi-object vehicle detection and tracking algorithm based on improved YOLOv8 and ByteTrack. *Electronics*, *13*(15), 3033.

ISSN: 2408-7920

African Journal of Applied Research
Vol. 12, No. 3 (2026), pp. 193-220
http://www.ajaronline.com
https://doi.org/10.26437/ajar.v12
Received May 3, 2025
Peer reviewed: December 12, 2025
Revised: February 19, 2026
Published: April 2026

Yu, H., Li, G., Zhang, W., Huang, Q., Du, D., Tian, Q., & Sebe, N. (2020). The unmanned aerial vehicle benchmark: Object detection, tracking and baseline. *International Journal of Computer Vision*, *128*(5), 1141–1159.

Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. *Neural Computation*, *31*(7), 1235–1270.

Yusuf, M. O., Hanzla, M., Al Mudawi, N., Sadiq, T., Alabdullah, B., Rahman, H., & Algarni, A. (2024). Target detection and classification via EfficientDet and CNN over unmanned aerial vehicles. *Frontiers in Neurorobotics*, *18*, 1448538.

Zhao, X., Xia, Y., Zhang, W., Zheng, C., & Zhang, Z. (2023). YOLO-ViT-based method for unmanned aerial vehicle infrared vehicle target detection. *Remote Sensing*, *15*(15), 3778.

Zheng, Z., Wang, P., Liu, W., Li, J., Ye, R., & Ren, D. (2020). *Distance-IoU loss: Faster and better learning for bounding box regression*. *34*(07), 12993–13000.

Zhu, P., Wen, L., Bian, X., Ling, H., & Hu, Q. (2018). Vision meets drones: A challenge. *arXiv Preprint arXiv:1804.07437*.

Zhu, P., Wen, L., Du, D., Bian, X., Ling, H., Hu, Q., Nie, Q., Cheng, H., Liu, C., & Liu, X. (2018). *Visdrone-det2018: The vision meets drone object detection in image challenge results*. 0–0.

Zhu, Y., Wang, Y., An, Y., Yang, H., & Pan, Y. (2024). Real-time vehicle detection and urban traffic behavior analysis based on uav traffic videos on mobile devices. *arXiv Preprint arXiv:2402.16246*.

Zou, C., Jeon, W.-S., & Rhee, S.-Y. (2024). Research on the multiple small target detection methodology in remote sensing. *Sensors*, *24*(10), 3211.

ISSN: 2408-7920