International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802224

An Efficient YOLOv8–DeepSORT Framework for Real-Time Multi-Object Video Surveillance

Dr. N. Devakirubai, Mr. G. Kannan, G. Archana, S. Bhuvaneshwari

HOD, Department of Artificial Intelligence and Data Science, R P Sarathy Institute of Technology, Salem, Tamil Nadu, India

Project Guide, Department of Artificial Intelligence and Data Science, R P Sarathy Institute of Technology, Salem, Tamil Nadu, India

Department of Artificial Intelligence and Data Science, R P Sarathy Institute of Technology, Salem, Tamil Nadu, India

Publication History: Received: 25.02.2026; Revised: 20.03.2026; Accepted: 25.03. 2026; Published: 28.03.2026.

ABSTRACT: Video surveillance systems play a vital role in maintaining safety and security across diverse environments such as educational institutions, transportation hubs, commercial infrastructures, and smart city ecosystems. The exponential increase in the deployment of closed-circuit television (CCTV) cameras has led to the generation of large-scale video data, making continuous manual monitoring inefficient, labor-intensive, and prone to human errors. In addition, existing automated surveillance systems predominantly focus on object detection without maintaining object identities across consecutive frames, resulting in unstable tracking, identity switching, and reduced situational awareness in dynamic environments.To overcome these challenges, this paper presents a robust and scalable AI-based real-time video surveillance framework that integrates the YOLOv8 object detection model with the DeepSORT multi-object tracking algorithm. The proposed approach leverages the high-speed and accurate detection capabilities of YOLOv8 to identify objects of interest, while DeepSORT enhances tracking performance by preserving object identities using motion estimation through Kalman filtering and appearance-based feature matching. This combined framework enables reliable detection and continuous tracking of multiple objects across video frames, even in complex scenarios involving occlusions, crowded scenes, and varying illumination conditions.The system is implemented using Python and OpenCV, ensuring flexibility, ease of deployment, and cost-effectiveness for real-world applications. Extensive experiments are conducted on benchmark datasets and real-time video sequences to evaluate the performance of the proposed model. The results demonstrate that the system achieves an accuracy of 97.62%, precision of 97.53%, recall of 99.94%, and an F1-score of 98.72%, while maintaining real-time processing speed. Furthermore, the integration of DeepSORT significantly reduces identity switching and enhances tracking stability compared to conventional detection-only approaches.

Overall, the proposed framework provides an efficient, reliable, and scalable solution for intelligent video surveillance, making it suitable for practical deployment in security-critical applications such as traffic monitoring, public safety, and smart surveillance systems.

KEYWORDS: Video Surveillance, YOLOv8, DeepSORT, Multi-Object Tracking, Object Detection, Real-Time Computer Vision

I. INTRODUCTION

The increasing demand for intelligent security solutions has made video surveillance systems a fundamental component of modern infrastructure. These systems are widely deployed across diverse domains, including smart cities, transportation networks, industrial facilities, educational campuses, and public safety environments. With the rapid expansion of closed-circuit television (CCTV) networks, an enormous volume of video data is continuously generated, creating significant challenges for effective monitoring and analysis. Traditional surveillance approaches that rely on human operators are not only labor-intensive and costly but also susceptible to fatigue, delayed responses, and reduced attention, particularly when managing multiple video streams simultaneously. As a result, the reliability and efficiency of manual surveillance systems are often compromised.

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802224

Recent advancements in artificial intelligence (AI) and deep learning have enabled the development of automated surveillance systems capable of analyzing video data in real time. Among various computer vision tasks, object detection plays a crucial role by identifying and localizing objects within images and video frames. Earlier methods, such as background subtraction, Histogram of Oriented Gradients (HOG), and Support Vector Machines (SVM), have shown limited effectiveness in complex and dynamic environments due to their inability to handle variations in lighting, scale, and occlusion. In contrast, deep learning-based approaches have significantly improved detection accuracy and robustness. Models such as Faster R-CNN, Single Shot Detector (SSD), and YOLO (You Only Look Once) have demonstrated superior performance, with YOLO-based architectures being particularly suitable for real-time applications due to their high processing speed and efficiency. The latest iteration, YOLOv8, further enhances detection capabilities by providing improved accuracy, faster inference, and better generalization across diverse datasets.

Despite these advancements in object detection, a critical limitation remains in practical surveillance scenarios: the inability to maintain consistent object identities across consecutive frames. Detection-only systems treat each frame independently, which often leads to identity switching, fragmented trajectories, and reduced situational awareness. These limitations become more pronounced in real-world conditions involving crowded scenes, partial occlusions, dynamic object interactions, and varying illumination. Consequently, integrating object tracking mechanisms with detection models has become essential for building reliable surveillance systems. Multi-object tracking aims to assign persistent identities to detected objects and track their movement over time. Among various tracking methods, DeepSORT (Deep Simple Online and Realtime Tracking) has emerged as a robust solution by combining motion prediction through Kalman filtering with deep appearance feature extraction for accurate object re-identification.

Motivated by these challenges, this study presents a comprehensive and scalable real-time video surveillance framework that integrates YOLOv8 for object detection with DeepSORT for multi-object tracking. The proposed system is designed to operate efficiently in dynamic and complex environments, enabling accurate detection and continuous tracking of multiple objects with minimal latency. The framework is evaluated using a combination of benchmark datasets and real-world video sequences, ensuring robustness across diverse scenarios such as crowded environments, occlusions, and varying lighting conditions. Experimental results demonstrate that the proposed approach achieves high detection accuracy and reliable tracking performance while maintaining real-time processing speed, making it suitable for deployment in practical applications.

The key contributions of this work are summarized as follows:

- A unified framework that integrates YOLOv8 and DeepSORT for efficient real-time multi-object surveillance.
- An identity-preserving tracking mechanism that performs reliably under occlusion, crowd density, and complex motion scenarios.
- A high-performance system that achieves strong accuracy and low latency, supported by extensive experimental evaluation on benchmark and real-world datasets.

In summary, the proposed framework offers a robust, scalable, and efficient solution for intelligent video surveillance, with potential applications in traffic management, smart city monitoring, public safety, and automated security systems.

II. RELATED WORK

A substantial body of research has been dedicated to improving object detection and tracking techniques for intelligent video surveillance systems. Early advancements focused on enhancing detection accuracy while reducing computational complexity. For instance, Cross-Stage Partial Networks (CSPNet) were introduced to optimize feature extraction and minimize redundant computations, thereby improving efficiency. However, such approaches still demand considerable computational resources and often struggle to maintain performance in highly dynamic and cluttered environments.

Several studies have explored the application of YOLO-based architectures for domain-specific surveillance tasks. Approaches utilizing earlier YOLO versions, such as YOLOv2, have demonstrated effective performance in applications like fire and smoke detection. While these systems achieve real-time detection, their robustness is often limited under challenging conditions, including variations in illumination, background complexity, and environmental noise. Similarly, deep learning-based frameworks developed for applications such as social distancing monitoring have shown promising detection capabilities but encounter scalability issues when deployed in densely populated scenes.

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802224

Other research efforts have focused on specialized detection tasks, such as weapon detection using lightweight convolutional neural networks. Although these models achieve high classification accuracy, they generally lack the ability to track multiple objects simultaneously and fail to preserve object identities over time. In parallel, transformer-based approaches have emerged as a powerful alternative for visual understanding tasks, including anomaly detection in surveillance videos. Despite their strong representational capabilities, these models are computationally intensive and often unsuitable for real-time applications, particularly in resource-constrained environments.

More recent studies have attempted to enhance detection performance by incorporating attention mechanisms and hybrid architectures combining YOLO with transformer-based components. While these methods improve feature representation and detection accuracy, they predominantly focus on frame-level object detection and do not adequately address the challenge of continuous multi-object tracking. Furthermore, the high computational overhead associated with such models limits their applicability in edge devices and real-time surveillance scenarios.

From the existing literature, it is evident that most approaches prioritize either detection accuracy or temporal modeling, but rarely provide a unified and efficient solution that addresses both aspects simultaneously. In addition, limited attention has been given to maintaining consistent object identities across frames in complex real-world conditions, such as occlusions, crowded scenes, and varying lighting environments. These limitations highlight the need for a robust, scalable, and real-time surveillance framework that integrates accurate object detection with reliable identity-preserving multi-object tracking.

III. PROBLEM STATEMENT

Despite the significant progress achieved in deep learning-based object detection, current video surveillance systems continue to face several practical challenges that limit their effectiveness in real-world deployments. One of the primary limitations of existing systems is their reliance on detection-only frameworks, which process each frame independently without maintaining consistent object identities across time. This often results in identity switching, fragmented object trajectories, and unreliable tracking, thereby reducing situational awareness in surveillance applications.

The complexity of real-world environments further exacerbates these challenges. Factors such as occlusion, high object density, rapid motion, dynamic backgrounds, and varying illumination conditions significantly affect detection and tracking performance. In crowded scenarios, objects frequently overlap or partially disappear from view, making it difficult for conventional systems to maintain accurate and continuous tracking.Although transformer-based and attention-driven models have shown promising improvements in detection accuracy, their high computational requirements pose a major limitation for real-time applications. These models typically require powerful hardware resources, making them unsuitable for deployment in edge devices, embedded systems, or large-scale surveillance networks where efficiency and low latency are critical. Furthermore, many existing solutions lack scalability and adaptability, limiting their ability to perform consistently across diverse environments and datasets.To address these challenges, there is a need for an integrated and efficient surveillance framework that combines accurate object detection with robust multi-object tracking while maintaining real-time performance. The system should be capable of handling complex scenarios, including crowded environments, occlusions, and varying lighting conditions, while ensuring low computational overhead and high scalability.In this context, the present study aims to develop a unified real-time surveillance system that leverages advanced detection and tracking techniques to achieve reliable multi-object detection, continuous identity preservation, and efficient processing. Such a system has significant potential for practical deployment in applications such as smart city monitoring, traffic management, public safety, and automated security systems, where accuracy, reliability, and real-time performance are essential.

IV. PROPOSED METHODOLOGY

4.1 System Architecture

The proposed system is designed as an integrated real-time video surveillance framework that combines object detection and multi-object tracking to achieve accurate and continuous monitoring. The architecture consists of four primary stages: video acquisition, frame processing, object detection, and object tracking.

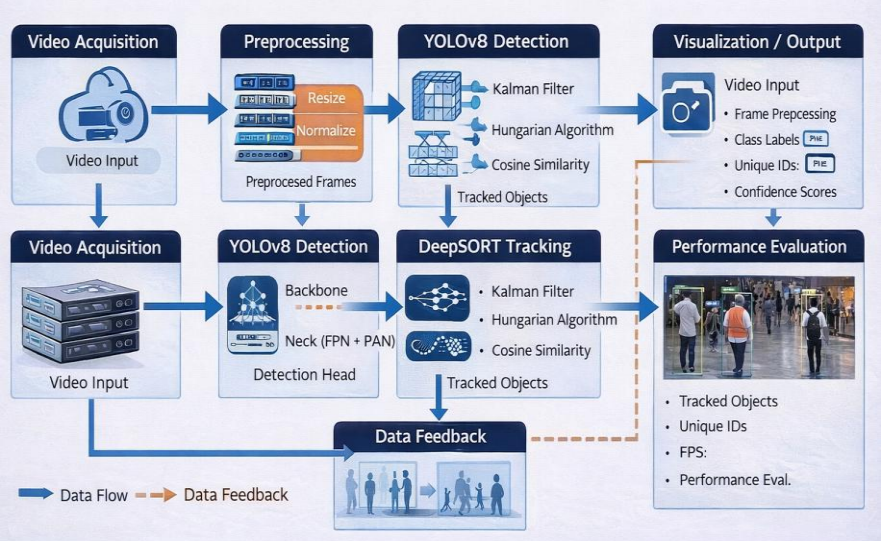
Initially, video data is captured from surveillance sources and processed into sequential frames using an efficient video decoding mechanism. Each frame is then forwarded to the object detection module, where relevant objects are identified and localized. The detected objects are subsequently passed to a tracking module, which assigns persistent

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802224

identities and tracks their movement across consecutive frames. Finally, the processed frames are visualized with annotated bounding boxes and unique object identifiers, providing an interpretable and real-time surveillance output.
This pipeline ensures seamless integration between detection and tracking components, enabling reliable performance in dynamic and complex environments.



A. Video Input Module
The Video Input Module is responsible for acquiring video streams from multiple sources, including CCTV cameras, webcams, and pre-recorded video datasets. The captured video is decoded into a sequence of frames using efficient video processing libraries.
Each frame is resized and preprocessed to match the input specifications required by the detection model. This module also ensures proper synchronization between frame acquisition and processing, which is essential for maintaining real-time performance. Efficient handling of input streams minimizes latency and ensures continuous data flow throughout the system.

B. Frame Processing Module
The frame processing stage converts the continuous video stream into discrete frames suitable for analysis. Since deep learning models operate on image data, this step is crucial for enabling object detection and tracking operations.
During this stage, each frame undergoes preprocessing steps such as resizing, normalization, and format conversion. These operations ensure consistency in input dimensions and improve model performance. By processing frames sequentially, the system maintains temporal continuity while enabling efficient frame-wise analysis.

Key Objectives:
- Transform video streams into analyzable image frames
- Prepare input data for deep learning models
- Ensure consistency and real-time processing capability

C. YOLOv8-Based Object Detection Module
The object detection component is based on the YOLOv8 architecture, a single-stage deep learning model known for its efficiency and real-time performance. This module processes each incoming frame to detect and classify objects while simultaneously predicting their spatial locations.
YOLOv8 extracts hierarchical features using deep convolutional layers and enhances multi-scale feature representation through Feature Pyramid Networks (FPN) and Path Aggregation Networks (PAN). This enables the model to detect objects of varying sizes with high accuracy.
The detection output for each object is represented as a tuple containing the bounding box coordinates, class label, and confidence score:

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

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802224

where (x,y)(x, y)(x,y) denotes the center coordinates of the bounding box, www and hhh represent its width and height, and ccc indicates the class probability.
This module provides high detection accuracy with low inference time, making it suitable for real-time surveillance applications.

D. DeepSORT-Based Tracking Module
To ensure continuous tracking and identity preservation, the system incorporates the DeepSORT algorithm. This module receives detection outputs and assigns a unique identity to each object, enabling consistent tracking across frames.
DeepSORT combines motion prediction and appearance-based matching to achieve robust tracking. Motion estimation is performed using a Kalman filter, which predicts the future position of each object based on its previous state. Data association between predicted tracks and new detections is achieved using the Hungarian algorithm, which minimizes the overall matching cost.
Additionally, deep appearance features are extracted using a convolutional neural network and compared using cosine similarity. This allows the system to re-identify objects even in challenging scenarios such as occlusion, overlap, or abrupt motion.
By integrating motion and appearance information, the tracking module maintains stable identities and reduces identity switching, thereby enhancing overall system reliability.

E. Output Visualization Module
The Output Visualization Module presents the final processed results in an interpretable format. It overlays bounding boxes, class labels, and unique tracking IDs onto each frame. Different objects are displayed using distinct colors to improve visual clarity and differentiation.
This module enables real-time monitoring and supports recording of processed video streams and tracking logs for further analysis. It serves as the interface between the system and end-users, facilitating easy interpretation of surveillance data.

Algorithm

Input:

Live or recorded video stream

Output:

Processed video frames with detected objects, assigned identities, and performance metrics

Procedure:

1. Initialize the YOLOv8 model with pre-trained weights.
2. Initialize the DeepSORT tracker with motion and appearance models.
3. Acquire video input from a camera or stored video source.
4. For each frame in the video stream:
    - Capture and preprocess the current frame.
    - Perform object detection using YOLOv8.
    - Extract bounding boxes, class labels, and confidence scores.
    - Provide detection outputs to the DeepSORT tracker.
    - Predict object positions using the Kalman filter.
    - Associate detections with existing tracks using the Hungarian algorithm.
    - Assign unique IDs to newly detected objects.
    - Update tracked object states.
    - Overlay detection and tracking results on the frame.
    - Display or store the processed frame.
5. Evaluate detection and tracking performance using standard metrics.
6. Terminate the process after the video stream ends.

4.2 YOLOv8 Object Detection

A. Overview of YOLOv8
YOLOv8 (You Only Look Once version 8) is a state-of-the-art, single-stage object detection model designed for high-speed and accurate real-time applications. Unlike traditional two-stage detectors such as Faster R-CNN, which rely on a separate region proposal mechanism, YOLOv8 performs object localization and classification in a single forward pass. This unified approach significantly reduces computational overhead and enables low-latency inference, making it highly suitable for real-time video surveillance systems.

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

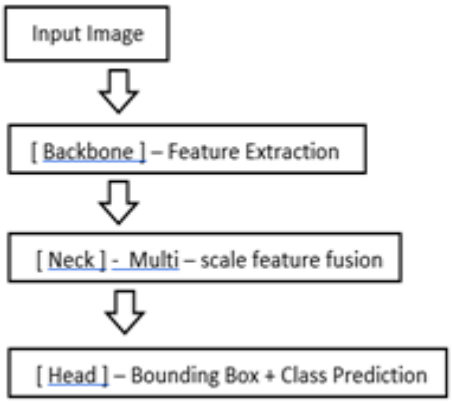
| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802224

Furthermore, YOLOv8 adopts an anchor-free detection strategy, which simplifies bounding box prediction and improves generalization across diverse object scales and environments. Its fully convolutional architecture enhances feature extraction efficiency and allows the model to process images of varying resolutions without requiring complex preprocessing steps.

B. YOLOv8 Architecture

The YOLOv8 framework is composed of three principal components: the backbone, neck, and detection head. Each component contributes to efficient feature extraction and accurate object prediction.



1. Backbone

The backbone network is responsible for extracting hierarchical feature representations from the input image. It utilizes convolutional layers combined with Cross Stage Partial (CSP) blocks to capture both low-level spatial details and high-level semantic information. This design improves feature reuse while reducing computational redundancy.

2. Neck

The neck module enhances feature representation by aggregating multi-scale information. It incorporates Feature Pyramid Networks (FPN) and Path Aggregation Networks (PAN) to fuse features from different levels of the backbone. This enables the model to effectively detect objects of varying sizes, particularly small and densely packed objects commonly found in surveillance scenarios.

3. Detection Head

The detection head generates the final predictions for each object. It outputs:

- Bounding box coordinates (x,y,w,h)(x, y, w, h)(x,y,w,h)
- Objectness confidence score
- Class probability distribution

The head operates directly on the fused feature maps, allowing simultaneous localization and classification in a single stage, thereby improving inference efficiency.

C. Mathematical Formulation

The training objective of YOLOv8 is defined by a composite loss function that optimizes localization, classification, and confidence estimation simultaneously:

$$L=Lbbox+Lcls+Lobj$$

where:

- Lbbox represents the bounding box regression loss
- Lcls denotes the classification loss
- Lobj corresponds to the objectness confidence loss

For precise localization, YOLOv8 employs the Complete Intersection over Union (CIoU) loss, which considers overlap area, center distance, and aspect ratio:

$$LCIoU=1−IoU+c2ρ2(b,bgt)+αv$$

where:

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802224

- IoU is the intersection over union between predicted and ground truth boxes
- ρ\rhoρ denotes the Euclidean distance between box centers
- ccc is the diagonal length of the smallest enclosing box
- v measures aspect ratio consistency
- α\alphaα is a weighting factor

This formulation enhances localization accuracy and ensures stable convergence during training.

D. Detection Algorithm

Input:

Preprocessed image frame

Output:

Detected objects with bounding boxes, class labels, and confidence scores

Procedure:

1. Receive the input frame I∈RH×W×CI \in \mathbb{R}^{H \times W \times C}I∈RH×W×C.
2. Normalize pixel values to improve model convergence.
3. Pass the frame through the backbone network to extract feature maps.
4. Apply multi-scale feature fusion using FPN and PAN structures.
5. Generate predictions including bounding boxes, confidence scores, and class probabilities.
6. Filter low-confidence detections using a predefined threshold.
7. Apply Non-Maximum Suppression (NMS) to remove redundant overlapping boxes.
8. Produce the final set of detections:

$$D=\{(x_i,y_i,w_i,h_i,c_i)\}_{i=1}^{N}$$

where N represents the number of detected objects.

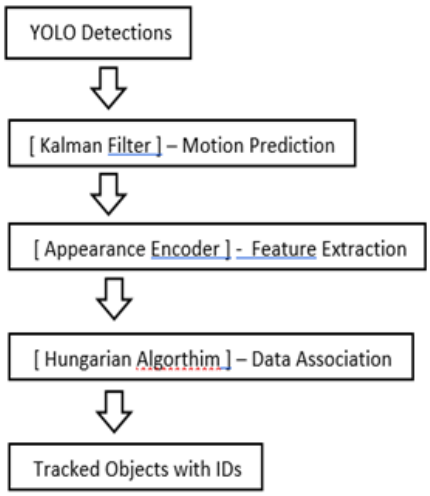
4.3 DEEPSORT MULTI-OBJECT TRACKING ALGORITHM

A. Overview of DeepSORT

DeepSORT (Deep Simple Online and Realtime Tracking) is an advanced multi-object tracking algorithm designed to maintain consistent object identities across consecutive video frames. It extends the conventional SORT framework by incorporating deep appearance features, enabling robust tracking even in challenging scenarios such as occlusion, object overlap, and abrupt motion.

By combining motion prediction with visual feature matching, DeepSORT achieves reliable identity preservation, which is essential for real-time surveillance applications requiring continuous object tracking and trajectory analysis.

B. DeepSORT Architecture



The DeepSORT framework consists of three key components: motion modeling, data association, and appearance feature extraction. These components work collaboratively to ensure accurate and stable tracking performance.

- **Motion Model:** Predicts the future position of each object using temporal information.

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802224

- **Data Association Module:** Matches detected objects with existing tracks based on similarity measures.
- **Appearance Feature Extractor:** Generates discriminative feature embeddings for robust object re-identification.

This modular design enables efficient integration with object detection models such as YOLOv8 and ensures scalability in real-time environments

C. Core Algorithms in DeepSORT

1. Kalman Filter (Motion Estimation)

The Kalman filter is used to estimate the future state of each tracked object based on its previous observations. The state vector typically includes position, velocity, and bounding box parameters.

The state prediction is defined as:

$$xk=Axk−1+wk$$

where:

- xk is the predicted state vector at time k
- A represents the state transition matrix
- wk denotes process noise, assumed to follow a Gaussian distribution.

This model enables smooth tracking by predicting object motion even when detections are temporarily unavailable.

2.Hungarian Algorithm (Data Association)

To associate current detections with existing tracks, DeepSORT formulates the assignment problem as a cost minimization task. The Hungarian algorithm is employed to find the optimal matching between predicted tracks and new detections.

The objective is defined as:

$$mini,j∑CijXij$$

where:

- Cij represents the cost of assigning detection j to track i
- Xij is a binary assignment variable

This ensures globally optimal matching while minimizing identity mismatches.

D. Tracking Algorithm

Input:

Detected objects from the YOLOv8 model

Output:

Tracked objects with consistent identities across frames

Procedure:

1. Initialize an empty set of object tracks.
2. For each incoming video frame:
    - Predict the next state of existing tracks using the Kalman filter.
    - Extract appearance features for newly detected objects using a CNN encoder.
    - Compute similarity scores between detections and existing tracks.
    - Construct a cost matrix based on motion and appearance features.
    - Apply the Hungarian algorithm to associate detections with tracks.
    - Update matched tracks with new observations.
    - Assign new identities to unmatched detections.
    - Remove tracks that exceed a predefined inactivity threshold.
3. Output the updated tracks with unique object IDs.

4.4 Mathematical Formulation of Evaluation Metrics

To quantitatively evaluate the performance of the proposed surveillance system, standard classification and detection metrics are employed. These metrics assess the accuracy and reliability of both detection and tracking components.

1. Accuracy

$$Accuracy=TP+TN/FP+FN+TP+TN$$

2. Precision

$$Precision=TP/FP+TP$$

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802224

3. Recall

Recall=TP/TP+FN

4. F1-Score

F1=2×Precision×Recall/Precision+Recall

V. DATASET DESCRIPTION

To evaluate the performance and robustness of the proposed real-time surveillance framework, experiments were conducted using a combination of benchmark datasets and real-world video sequences. This hybrid dataset strategy ensures both reproducibility and practical relevance under real-world operating conditions.

The primary dataset employed in this study is derived from the widely recognized Microsoft Common Objects in Context (MS COCO) dataset, which serves as a standard benchmark in object detection research. This dataset contains over 330,000 images with approximately 1.5 million annotated object instances spanning 80 distinct categories. For the purpose of this work, a subset of 20 object classes relevant to surveillance applications was selected. These include commonly observed entities such as pedestrians, vehicles, personal belongings, and public infrastructure elements.

The selected object categories—such as person, car, bus, truck, motorcycle, bicycle, backpack, handbag, suitcase, traffic light, and other frequently occurring objects—are representative of real-world surveillance environments. These classes were chosen to reflect practical monitoring scenarios in locations such as urban roads, transportation hubs, educational campuses, commercial spaces, and public areas. The inclusion of both movable objects and static infrastructure enhances the system's capability to support applications such as crowd monitoring, traffic analysis, and security assessment.

In addition to the benchmark dataset, custom real-time video sequences were incorporated to validate system performance under realistic conditions. These video samples were collected from publicly available surveillance sources, including traffic monitoring systems, campus security footage, and general CCTV recordings. The dataset captures diverse real-world scenarios such as pedestrian movement, vehicle interactions, dense crowds, and dynamic environmental conditions.

The combined dataset consists of approximately 12,000 frames extracted from multiple video sequences. These frames encompass a wide range of variations, including different lighting conditions (day and night), multiple viewpoints, static and dynamic backgrounds, object occlusion, scale variation, and motion complexity. Such diversity ensures that the proposed model is evaluated under challenging and representative conditions, improving its generalization capability.

For experimental evaluation, the dataset was partitioned into training and testing subsets using an 80:20 split. The training portion was utilized to optimize the YOLOv8 detection model, while the testing set was used to evaluate both detection and tracking performance in conjunction with the DeepSORT algorithm. All frames were annotated with bounding boxes and corresponding class labels. For custom video data, annotations were generated using manual labeling tools to ensure precise object localization and reliable ground truth information.

The integration of a standardized benchmark dataset with real-world surveillance footage provides a balanced evaluation framework. While the benchmark dataset facilitates comparison with existing approaches, the inclusion of real-time video data demonstrates the practical applicability of the proposed system. This combined approach validates the effectiveness of the framework in handling complex, dynamic, and real-world surveillance scenarios, making it suitable for deployment in intelligent monitoring systems.

VI. DATA PREPROCESSING

Effective data preprocessing is essential for ensuring the reliability and performance of deep learning-based surveillance systems, as model accuracy is highly dependent on input data quality. In this work, raw video streams are first decoded into individual frames using efficient video processing techniques. Redundant and highly similar frames are filtered out to reduce data imbalance and unnecessary computational overhead.

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802224

All frames are resized to a uniform resolution of 640×640 pixels, which aligns with the input requirements of the YOLOv8 architecture and ensures consistency across the dataset. To facilitate stable training and faster convergence, pixel intensities are normalized using standard statistical transformation:

$$I_{norm} = \frac{I - \mu}{\sigma}$$

where I represents the original pixel intensity, μ\muμ denotes the mean, and σ\sigmaσ is the standard deviation. This normalization reduces the impact of illumination variations and improves feature learning under diverse lighting conditions commonly encountered in surveillance environments.

To enhance model robustness and mitigate overfitting, a comprehensive data augmentation strategy is employed. Augmentation techniques include horizontal flipping, random rotation, scaling, cropping, and adjustments in brightness and contrast. These transformations increase the diversity of training samples and enable the model to generalize effectively to unseen scenarios, such as crowded scenes, occlusions, and varying camera perspectives.

Additionally, image enhancement techniques are applied to improve visual quality. Noise reduction is performed using Gaussian filtering, while histogram equalization is used to enhance contrast, particularly in low-light conditions. Annotation quality is also verified to ensure accurate bounding box alignment and class labeling.

Finally, the processed dataset is converted into the standard YOLO annotation format, where each image is associated with a label file containing normalized bounding box coordinates and corresponding class indices. This ensures seamless integration with the YOLOv8 training pipeline and facilitates efficient model learning.

VII. EXPERIMENTAL SETUP

The proposed surveillance framework is implemented using a Python-based deep learning environment built on PyTorch and OpenCV libraries. YOLOv8 is employed for object detection, while DeepSORT is integrated for multi-object tracking. All experiments are conducted in a GPU-enabled computing environment to ensure efficient processing and real-time performance.

During training, the YOLOv8 model is initialized with pre-trained weights to accelerate convergence and improve detection accuracy. The training process is performed for 10 epochs with a batch size of 16, using the Adam optimizer with a learning rate of 0.001. Input images are resized to 640×640640 \times 640640×640 pixels to maintain consistency with the model architecture.

To evaluate the effectiveness of the proposed system, experiments are conducted on both benchmark datasets and real-world surveillance video sequences. Detection performance is assessed using standard evaluation metrics, including accuracy, precision, recall, and F1-score. In addition, real-time performance is measured in terms of frames per second (FPS), reflecting the system's ability to process video streams efficiently.

Tracking performance is evaluated based on identity consistency and continuity of object trajectories across frames. The integration of detection and tracking results is visualized by overlaying bounding boxes and unique object identifiers on video frames. This allows both quantitative and qualitative assessment of system performance under dynamic conditions, including crowded environments, occlusions, and varying illumination.

Overall, the experimental setup is designed to ensure a comprehensive evaluation of the proposed framework, demonstrating its effectiveness, scalability, and suitability for real-world intelligent surveillance applications.

International Journal of Engineering & Extended Technologies Research (IJEETR)

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802224

VIII. RESULTS AND DISCUSSION

8.1 Training Performance Analysis

Accuracy vs Epoch

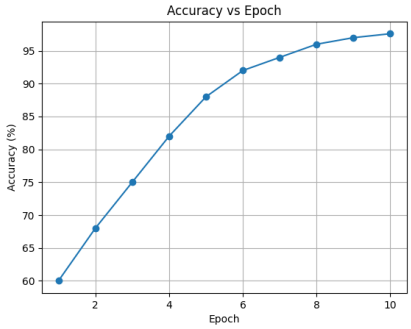


Fig .1 shows how the classification accuracy changes over the training epochs. The accuracy increases steadily from 60% in the first epoch to about 97.6% in the tenth epoch. This indicates that the YOLOv8 model learns effectively from the training data. The rapid increase in the early epochs shows fast learning, while the stable accuracy in later epochs suggests that the model has reached good convergence.

Training Loss vs Epoch

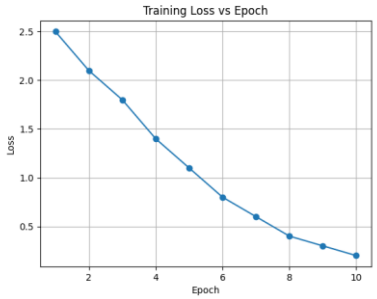


Fig. 2 illustrates the training loss of the YOLOv8 model over ten epochs. The loss value decreases from 2.5 to 0.2, showing that the model gradually reduces prediction errors during training. The continuous downward trend indicates stable learning and proper optimization of the model parameters.

8.2 FPS Comparison of YOLO Models

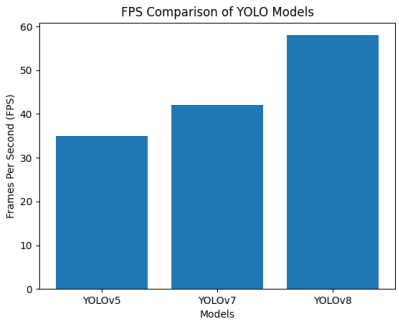


Fig. 3 compares the inference speed of different YOLO models in terms of frames per second (FPS). YOLOv5 achieves around 35 FPS, while YOLOv7 reaches about 42 FPS. YOLOv8 performs the best with an average speed of 58 FPS. [1], [10] This shows that YOLOv8 is more suitable for real-time surveillance applications due to its higher processing speed.

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802224

8.3 Classification Metrics Bar Chart

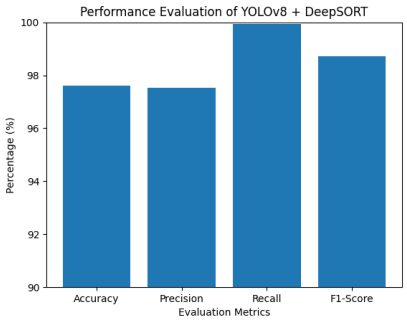


Fig. 4 presents the performance of the proposed system using accuracy, precision, recall, and F1-score. The system achieves an accuracy of 97.62%, indicating high prediction correctness. The precision of 97.53% shows fewer false detections, while the recall of 99.94% indicates that most objects are successfully detected. The F1-score of 98.72% reflects a good balance between precision and recall, confirming the reliability of the proposed system.

IX. CONCLUSION AND FUTURE WORK

This study presented a comprehensive real-time video surveillance framework that integrates YOLOv8 for object detection with DeepSORT for multi-object tracking. The proposed approach addresses key limitations of conventional surveillance systems by enabling accurate detection and consistent identity tracking of multiple objects in dynamic environments. By combining a high-speed single-stage detection model with an efficient tracking mechanism, the system ensures continuous monitoring with minimal latency.

Extensive experimental evaluation demonstrates the effectiveness of the proposed framework in both detection and tracking tasks. The system achieves an accuracy of 97.62%, precision of 97.53%, recall of 99.94%, and an F1-score of 98.72%, indicating strong predictive performance. In addition, the model operates at an average speed of approximately 58 frames per second, confirming its capability to meet real-time processing requirements. These results highlight the ability of the framework to maintain a balance between computational efficiency and high accuracy.

The robustness of the system is further validated through experiments on real-world surveillance scenarios, including crowded environments, object occlusions, varying illumination conditions, and complex object interactions. The integration of DeepSORT significantly enhances tracking stability by preserving object identities across frames and reducing identity switching. This makes the system highly suitable for practical applications such as traffic monitoring, smart city infrastructure, campus security, and industrial surveillance.

Despite the promising performance, certain limitations remain that open avenues for future research. The current framework can be extended by incorporating advanced functionalities such as face recognition, human activity analysis, and anomaly detection to improve situational awareness. Furthermore, integrating lightweight model optimization techniques and edge computing strategies would enable deployment on resource-constrained devices, including embedded systems and smart cameras. Expanding the training process with larger and more diverse datasets can also enhance generalization and improve performance in highly complex environments.

In conclusion, the proposed YOLOv8–DeepSORT-based framework offers a scalable, efficient, and reliable solution for real-time multi-object detection and tracking. Its strong performance and adaptability make it a promising foundation for next-generation intelligent surveillance systems, with significant potential for real-world deployment in safety-critical applications.

REFERENCES

1. M. Yaseen, “What Is YOLOv8: An In-Depth Exploration of the Internal Features of the Next-Generation Object Detector,” arXiv preprint arXiv:2408.15857, Aug. 2024.
2. Q. Chen,“LW-DETR: A Transformer Replacement to YOLO for Real-Time Detection,”arXiv preprint arXiv:2406.03459, Jun. 2024.

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802224

3. D. Nimma,"Object detection in real-time video surveillance using attention based transformer-YOLOv8 model," Alexandria Engineering Journal, vol. 118, pp. 482–495, Jan. 2025, doi: 10.1016/j.aej.2025.01.032.
4. N. Yunusov,"Robust forest fire detection method for surveillance systems based on You Only Look Once version 8 and transfer learning approaches," Processes, vol. 12, no. 5, Art. no. 1039, May 2024, doi: 10.3390/pr12051039.
5. Y. Zhao,"FEB-YOLOv8: A multi-scale lightweight detection model for underwater object detection,"PLOS ONE, vol. 19, no. 9, Art. no. e0311173, Sep. 2024,doi: 10.1371/journal.pone.0311173.
6. E. Arkin, N. Yadikar, Y. Muhtar, and K. Ubul, "A survey of object detection based on CNN and transformer," in 2021 IEEE 2nd international conference on pattern recognition and machine learning (PRML), IEEE, 2021, pp. 99–108.
7. L. He and S. Todorovic, "Destr: Object detection with split transformer," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2022, pp. 9377–9386.
8. Y. Tian, "Effective image enhancement and fast object detection for improved UAV applications," 2023.
9. C.Nagarajan and M.Madheswaran - 'Stability Analysis of Series Parallel Resonant Converter with Fuzzy Logic Controller Using State Space Techniques'- Taylor &Francis, Electric Power Components and Systems, Vol.39 (8), pp.780-793, May 2011. DOI: 10.1080/15325008.2010.541746
10. C.Nagarajan and M.Madheswaran - 'Experimental verification and stability state space analysis of CLL-T Series Parallel Resonant Converter' - Journal of Electrical Engineering, Vol.63 (6), pp.365-372, Dec.2012. DOI: 10.2478/v10187-012-0054-2
11. C.Nagarajan and M.Madheswaran - 'Performance Analysis of LCL-T Resonant Converter with Fuzzy/PID Using State Space Analysis'- Springer, Electrical Engineering, Vol.93 (3), pp.167-178, September 2011. DOI 10.1007/s00202-011-0203-9
12. S.Tamilselvi, R.Prakash, C.Nagarajan,"Solar System Integrated Smart Grid Utilizing Hybrid Coot-Genetic Algorithm Optimized ANN Controller" Iranian Journal Of Science And Technology-Transactions Of Electrical Engineering, DOI10.1007/s40998-025-00917-z,2025
13. S.Tamilselvi, R.Prakash, C.Nagarajan," Adaptive sliding mode control of multilevel grid-connected inverters using reinforcement learning for enhanced LVRT performance" Electric Power Systems Research 253 (2026) 112428, doi.org/10.1016/j.epsr.2025.112428
14. S.Thirunavukkarasu, C. Nagarajan, 2024, "Performance Investigation on OCF and SCF study in BLDC machine using FTANN Controller," Journal of Electrical Engineering And Technology, Volume 20, pages 2675–2688, (2025), doi.org/10.1007/s42835-024-02126-w
15. C. Nagarajan, M.Madheswaran and D.Ramasubramanian- 'Development of DSP based Robust Control Method for General Resonant Converter Topologies using Transfer Function Model'- *Acta Electrotechnica et Informatica Journal* , Vol.13 (2), pp.18-31,April-June.2013, DOI: 10.2478/aeei-2013-0025.
16. C.Nagarajan and M.Madheswaran - 'DSP Based Fuzzy Controller for Series Parallel Resonant converter'- *Springer, Frontiers of Electrical and Electronic Engineering*, Vol. 7(4), pp. 438-446, Dec.12. DOI 10.1007/s11460-012-0212-0.
17. C.Nagarajan and M.Madheswaran - 'Experimental Study and steady state stability analysis of CLL-T Series Parallel Resonant Converter with Fuzzy controller using State Space Analysis'- *Iranian Journal of Electrical & Electronic Engineering*, Vol.8 (3), pp.259-267, September 2012.
18. C.Nagarajan and M.Madheswaran, "Analysis and Simulation of LCL Series Resonant Full Bridge Converter Using PWM Technique with Load Independent Operation" has been presented in ICTES'08, a IEEE / IET International Conference organized by M.G.R.University, Chennai.Vol.no.1, pp.190-195, Dec.2007
19. Suganthi Mullainathan, Ramesh Natarajan, "An SPSS and CNN modelling based quality assessment using ceramic materials and membrane filtration techniques", Revista Materia (Rio J.) Vol. 30, 2025, DOI: https://doi.org/10.1590/1517-7076-RMAT-2024-0721
20. M Suganthi, N Ramesh, "Treatment of water using natural zeolite as membrane filter", Journal of Environmental Protection and Ecology, Volume 23, Issue 2, pp: 520-530,2022
21. N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End- to-end object detection with transformers," in European conference on computer vision, Springer, 2020, pp. 213–229.
22. G. Lavanya, S. Pande, Enhancing Real-time Object Detection with YOLO Algorithm, EAI Endorsed Trans. Internet Things 10 (Dec. 2023), https://doi.org/10.4108/ eetiot.4541.
23. S. Jha, C. Seo, E. Yang, G.P. Joshi, Real time object detection and trackingsystem for video surveillance system, Multimed. Tools Appl. 80 (3) (Jan. 2021) 3981–3996, https://doi.org/10.1007/s11042-020-09749-x.
24. Q. Chen et al., "LW-DETR: A Transformer Replacement to YOLO for Real-Time Detection," arXiv preprint arXiv:2406.03459, 2024.
25. Z. Zhang, X. Lu, G. Cao, Y. Yang, L. Jiao, and F. Liu, "ViT-YOLO: Transformer-based YOLO for object detection," in Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 2799–2808.

|ISSN: 2322-0163| www.ijeetr.com | A Bimonthly, Peer Reviewed, Scholarly Indexed Journal |

| Volume 8, Issue 2, March - April 2026 |

DOI:10.15662/IJEETR.2026.0802224

26. P.Y. Ingle, Y.-G. Kim, Real-time abnormal object detection for video surveillance in smart cities, Art. no. 10, Sensors 22 (10) (Jan. 2022), https://doi.org/10.3390/ s22103862.
27. MATHEW, A. (2025). BEYOND THE BURNER: THE SYSTEMIC RISKS OF DISPOSABLE EMAIL ECOSYSTEMS.
28. Raj, A. M. A., Rajendran, S., & Vimal, G. S. A. G. (2024). Enhanced convolutional neural network enabled optimized diagnostic model for COVID-19 detection. *Bulletin of Electrical Engineering and Informatics, 13(3), 1935-1942.*
29. Kiran, A., Rubini, P., & Kumar, S. S. (2025). Comprehensive review of privacy, utility and fairness offered by synthetic data. *IEEE Access.*
30. Anand, L., & Syed Ibrahim, S. P. (2018). HANN: a hybrid model for liver syndrome classification by feature assortment optimization. *Journal of Medical Systems, 42(11), 211.*
31. Udayakumar, R., Yogesh Pansambal, S., Anbazhagan, K., & Sugumar, R. Real-time Migration Risk Analysis Model for Improved Immigrant Development Using Psychological Factors. *Migr Lett. 2023; 20 (4): 33–42.*
32. Gopinathan, V. R. (2025). Designing Cloud-Native Enterprise Systems by Modernizing Applications with Microservices and Kubernetes Platforms. *International Journal of Research and Applied Innovations, 8(5), 13052-13063.*