International Journal of Computer Applications (0975 - 8887)
Volume 187 - No.4, May 2025

Deep Learning-based Person Tracking: A Smart Approach to Security and Civic Monitoring

Shailendra Singh Kathait
Co-Founder and Chief Data Scientist
Valiance Solutions
Noida, India

Ashish Kumar
Principal Data Scientist
Valiance Solutions
Noida, India

Samay Sawal
Intern Data Scientist
Valiance Solutions
Noida, India

Ram Patidar
Data Scientist
Valiance Solutions
Noida, India

Khushi Agrawal
Intern Data Scientist
Valiance Solutions
Noida, India

ABSTRACT

Restricted-area violations, such as entering into vehicle zones, create serious security issues in a variety of monitoring applications. This research presents a deep learning-based framework for real-time detection and surveillance of individuals who violate designated restricted zones. The proposed system uses advanced object detection algorithms, specifically YOLOv8, for head detection and spatial reasoning to track individuals who enter restricted areas. The framework uses centroid-based tracking to accurately detect and count violations, ensuring that each individual is flagged once within a frame only once. The method improves detection accuracy further by modifying bounding boxes and using region-specific polygonal filtering, allowing for more exact violation detection. Visual feedback is provided by overlaying boundary boxes and labels on the detected individuals, while cumulative violation counts are recorded. This method is highly effective, providing stable performance in changing conditions, and can be used for crowd management, security, and surveillance. The system's architecture is flexible, with the ability to add capabilities like movement direction and speed analysis for more context-aware violations.

Keywords

Computer Vision, Traffic Surveillance, YOLO, Vehicle Speed Detection, Direction Detection, Helmet Detection, Lane Violation, Non-ANPR Cameras

1. INTRODUCTION

As urban highways become more congested, municipal planners and police have a crucial problem in managing human and automobile safety. People entering vehicle zones, such as roads, highways, or areas reserved exclusively for automobiles, pose significant risks to both the individuals involved and the overall flow of traffic. Traditional monitoring approaches, such as manual inspections or static surveillance cameras, have proven insufficient for real-time identification and regulation, especially in unpredictable traffic situations where violations might occur unexpectedly.

Current surveillance systems heavily depend on human oversight or specific technology such as motion sensors, which cannot frequently detect and track violations automatically and on a large scale. While modern technologies such as automatic number plate recognition (ANPR) cameras can track automobiles, there is an important technological gap in tracking people who accidentally enter vehicle zones. Furthermore, these old systems are frequently expensive, require complicated infrastructure, and necessitate extensive maintenance, restricting their use in many areas.

Recent developments in computer vision and deep learning [11] offer a chance to overcome these difficulties by utilizing existing traffic cameras and surveillance systems. These technologies, combined with strong object identification algorithms like YOLO (You Only Look Once), allow for real-time detection and tracking of people entering vehicle zones. These systems can detect violations automatically by analyzing video feeds from numerous general-purpose security cameras, avoiding the need for expensive or specialized equipment.

This paper presents a deep learning-based method for detecting and tracking people who violate vehicle zones on the road. The technology uses powerful object identification techniques to identify people in real-time and follow their movements as they enter restricted vehicle areas. The system detects violations precisely by utilizing centroid-based tracking and spatial reasoning, flags those identified, and provides visual feedback. This automated technology reduces the need for staff involvement, improves the effectiveness of road safety management, and offers a flexible, cost-effective method for metropolitan locations to utilize current surveillance infrastructure.

1.1 Motivation

This research is motivated by the need to improve road safety, ensure pedestrian-vehicle separation in traffic zones, and use ex-

isting infrastructure for intelligent monitoring. Cities can upgrade general-purpose security cameras to advanced systems capable of detecting human violations in vehicle-only zones.

(1) Provides real-time monitoring of restricted traffic areas without the need for expensive, specific technology.
(2) Enhance police consistency, resulting in greater obedience to traffic safety rules.
(3) Incorporate easily with existing smart city frameworks and data analytics platforms to enhance urban traffic and safety management.

1.2 Contributions

The primary contributions of this work are:

(1) **Utilizing Public Cameras**: Demonstrating the feasibility of utilizing general-purpose surveillance cameras, which are widely used for security purposes, to monitor human invasions in car zones.
(2) **Violation Detection Framework**: Creating a full structure that uses YOLO-based object identification algorithms to identify and track people entering restricted traffic areas in real-time.
(3) **Centroid-Based Tracking**: Using a centroid-based technique allows for pedestrian identity consistency between frames, which improves detection accuracy as well as reliability.
(4) **Zone-Specific Analysis**: Uses spatial reasoning and polygon-based filtering to detect violations within designated vehicle zones, ensuring accurate monitoring of restricted regions.
(5) **Adaptable and Scalable System**: Creating a flexible framework that adjusts to different camera angles and urban settings, as well as the capacity to integrate with the cloud, allowing for large-scale application in smart cities.

2. RELATED WORK

Several studies have explored person tracking using overhead views to address challenges faced in traditional front-view datasets [10]. One approach combines Faster-RCNN for object detection with the GOTURN architecture for tracking, enabling robust person tracking in various indoor and outdoor environments [1]. Another work focuses on long-term identity-aware multi-person tracking for multi-camera surveillance, leveraging spatial and appearance manifolds to propagate identity information across frames [2]. The SiamMask framework adopts fully convolutional Siamese networks for fast online object tracking and segmentation, integrating pixel-wise binary mask prediction with bounding box coordinates for improved tracking accuracy and efficiency [3]. Similarly, multi-camera tracking algorithms have been compared using techniques such as particle filtering, face detection, and edge alignment to track individuals in indoor environments with overlapping camera views [4]. In the context of 3D multi-object tracking, SimpleTrack introduces improvements like aggressive non-maximum suppression (NMS) and enhanced association techniques to reduce redundancy and improve localization and identification in complex scenarios [5].

These studies highlight advancements in tracking technologies through innovative deep learning and algorithmic approaches.

2.1 Data Acquisition

The dataset for this project was generated using real-world CCTV footage gathered from safety cameras monitoring public areas. To guarantee accurate human detection and tracking, pre-trained models, like YOLO, were refined using a carefully selected collection of overhead video streams. This fine-tuning procedure adjusts pre-trained models to the particular requirements of overhead view person tracking while using their broad feature extraction abilities.

2.2 YOLO Model Architecture

The YOLO [9] (You Only Look Once) model is a real-time object detection framework known for its speed and accuracy. Unlike traditional methods that involve region proposal followed by classification, YOLO treats object detection as a single regression problem. It divides the input image into a grid and predicts bounding boxes, class probabilities, and confidence scores for each cell.

Key Components of the YOLO Architecture:

(1) **Input Layer:** Resizes the image to a fixed dimension and normalizes pixel values.
(2) **Feature Extraction:** A series of convolutional layers extract features from the image, with batch normalization and pooling layers to improve generalization and reduce overfitting.
(3) **Grid-based Prediction:** The image is divided into an S×S grid. Each cell predicts bounding boxes, objectness scores, and class probabilities.
(4) **Bounding Box Regression:** Each bounding box is defined by its center coordinates, width, height, and confidence score, indicating the probability of the box containing an object.
(5) **Non-Maximum Suppression (NMS):** Filters overlapping boxes to retain the most probable detection for each object.

YOLO's unified architecture makes it exceptionally fast, making it suitable for real-time applications like traffic monitoring. Its trade-off between speed and accuracy allows scalable deployment on edge devices or cloud platforms.

For further details, refer to the article "Computer Vision and Deep Learning-based Approach for Traffic Violations due to Overspeeding and Wrong Direction Detection"

2.3 Model Architecture and Discussion

To detect and track persons in real-time, the system uses a customizable deep-learning framework. It efficiently detects humans using a YOLO-based architecture that detects objects instantly. Advanced geographical analysis and rule-based algorithms are then used to evaluate activities like accessing restricted areas or breaking from designated lines. .

Step-wise YOLO Model Process:
Input Image → Grid Based Detection → Bounding Box Prediction → Class Prediction → Non Maximum Suppression

Output: Detection of heads with bounding boxes, class labels, and confidence scores.

This paper utilized the YOLO Model for object identification and localization and integrated a tracking mechanism to maintain consistent object identities across frames. The following subsections elaborate on key components of the system:

(1) **Object Detection and Object Tracking :** A pre-trained YOLO model is used for object detection and tracking.

(a) **Object Detection :**The YOLO model identifies pedestrians in real time, dividing the input frame into an S×S grid. Each grid cell predicts bounding boxes and their confidence scores.

$$\text{Confidence} = P(\text{Object}) \cdot \text{IOU}_{\text{Pred,True}}$$

where $P(\text{Object})$ indicates the probability of an object in the cell, and $\text{IOU}_{\text{Pred,True}}$ is the intersection-over-union of the predicted and true bounding boxes.

(b) **Object Tracking :** A tracking mechanism assigns unique IDs to the detected person and maintains consistency across frames. Centroids and bounding box dimensions are used to match detections frame-to-frame.

(2) **Centroid Calculation and Categorization:** Each detected person is assigned a centroid [8] for simplified spatial representation:

$$\text{centroid}_x = \frac{x_1 + x_2}{2}, \quad \text{centroid}_y = \frac{y_1 + y_2}{2}.$$

Centroid positions are then checked against a pre-defined polygon representing the restricted vehicle zone. Persons whose centroids fall within this region are flagged as violators.

(3) **Restricted Zone Violation Detection:** The system checks whether a person is entering a vehicle zone:
•A polygonal restricted zone is pre-defined using its coordinates.
•If the person's centroid is within this zone, the system records the violation, flags the individual, and updates the violation count.

2.4 Visualization and Output:

The proposed system provides a real-time visualization framework to identify and track pedestrian violations in restricted vehicle zones. Each processed frame includes:

(1) Detected pedestrians in restricted vehicle zones are highlighted with unique bounding boxes.
(2) Each detected pedestrian is labeled with their tracking ID for consistent identification across frames.
(3) The restricted vehicle zone is outlined in the frame, providing a clear visual context for determining violations.
(4) Pedestrians entering the restricted zone are flagged with labels such as "Violation Detected."

This system ensures an effective real-time solution for detecting unauthorized pedestrian entry into vehicle zones.

3. CONCLUSION AND FUTURE WORK

The developed system successfully detects and tracks pedestrians entering restricted vehicle zones in real time. Using YOLO-based object detection and centroid tracking, it accurately identifies violations and highlights them on screen. The system also keeps track of violations and presents this data clearly, making it easier for authorities to act on the information.

This solution reduces the need for manual monitoring and contributes to safer roads by enforcing pedestrian-vehicle boundaries. Its flexible and scalable design makes it suitable for a wide range of urban traffic scenarios, while also being capable of integration into larger smart city systems.

Fig. 1. Turquoise Box around Pedestrian Invading Vehicle Area

Although the system works well in its current form, there are several ways to improve it further:

(1) Adding features to detect patterns like jaywalking or sudden intrusions could give more context to reported violations.
(2) Enhancing the logic to distinguish between accidental and intentional violations could help authorities take more precise actions.
(3) Making the system run efficiently on smaller devices placed at the cameras themselves could reduce delays and make it scalable for large deployments.
(4) Improving the system to adapt to different lighting, weather, or camera angles would make it even more reliable in real-world conditions.

4. REFERENCES

[1] Ahmad, Misbah & Ahmed, Imran & Khan, Fakhri & Qayum, Fawad & Aljuaid, Hanan. (2020). Convolutional neural network–based person tracking using overhead views. International Journal of Distributed Sensor Networks. 16. 155014772093473. 10.1177/1550147720934738.

[2] Yu, S., Yang, Y., Li, X., & Hauptmann, A. G. (2016). Long-Term Identity-Aware Multi-Person Tracking for Surveillance Video Summarization. ArXiv. https://arxiv.org/abs/1604.07468

[3] Wang, Q., Zhang, L., Bertinetto, L., Hu, W., & Torr, P. H. (2018). Fast Online Object Tracking and Segmentation: A Unifying Approach. ArXiv. https://arxiv.org/abs/1812.05050

[4] Liem, Martijn & Gavrila, Dariu. (2013). A Comparative Study on Multi-person Tracking Using Overlapping Cameras. 203-212. 10.1007/978-3-642-39402-7_21.

[5] Pang, Z., Li, Z., & Wang, N. (2021). SimpleTrack: Understanding and Rethinking 3D Multi-object Tracking. ArXiv. https://arxiv.org/abs/2111.09621

[6] Shailendra Singh Kathait, Ashish Kumar, Ram Patidar, Khushi Agrawal, Samay Sawal (2024). Computer Vision and Deep Learning based Approach for Traffic Violations due to Over-speeding and Wrong Direction Detection. International Journal of Computer Applications, paper-id: 6e503f15-f6c9-4ee2-9212-4db588484729, DOI: 10.5120/ijca2025924477

[7] Shailendra Singh Kathait, Ashish Kumar, Ram Patidar, Khushi Agrawal, Samay Sawal (2024). Computer Vision and Deep Learning based Approach for Violations due to Illegal Parking Detection. International Journal of Computer Applications, DOI: 10.5120/ijca2025924506

[8] Rupesh Parthe, The Importance of Centroid in Image Processing. INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT, DOI: 10.55041/IJSREM30775

[9] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, You Only Look Once: Unified, Real-Time Object Detection. DOI: https://doi.org/10.48550/arXiv.1506.02640

[10] Shailendra Singh Kathait, Ashish Kumar, Ram Patidar, Khushi Agrawal, Samay Sawal (2024). Deep Learning-based Approach for Detecting Traffic Violations Involving No Helmet Use and Wrong Cycle Lane Usage. International Journal of Computer Applications, DOI: 10.5120/ijca2025924714

[11] Shailendra Singh Kathait, Shubhrita Tiwari, Application of Image Processing and Convolution Networks in Intelligent Character Recognition for Digitized Forms Processing, DOI: 10.5120/ijca2018915460