Dynamic Emerging Pathways in Entrance and Exit Detection: Integrating Deep Learning and Mathematical Modeling

LOAI ABDALLAH, MUTLAQ HIJAZI, MURAD MUSTAFA BADARNA
Department of Information Systems,
The Max Stern Yezreel Valley Academic College,
D.N. Emek Yezreel, 1930600,
ISRAEL

Abstract: - Entrance and exit event detection in dynamic environments has a lot of real-world applications in security, crowd management, and retail analytics. Traditional methods used for this problem, namely Line Partition and Bounding Box Diameter methods often struggle in complex scenarios that contain less predictable movement patterns of individuals. This paper proposes a model that integrates deep learning-based object detection and tracking techniques with linear regression to enhance the overall performance of enter and exit detection in static and dynamic environments. This approach captures the movement patterns using advanced object detection and tracking algorithms, enabling the extraction of y-coordinate variations from bounding box centers which are used to calculate the tangent of the linear regression equation and determine if the event is entrance or exit. Experimentations were conducted on 132 video sequences and show the superiority of our approach over the traditional methods achieving an overall accuracy of 86.36% and an F1-score of 0.86. These results demonstrate the high efficiency of this approach to accurately detect entrance and exit events, making it highly reliable and applicable to this problem. This research contributes to computer vision by integrating object detection and tracking algorithms with linear regression offering a solution for enhancing entrance and exit events detection in dynamic environments.

Key-Words: - Deep Learning, Mathematical Modeling, Linear Regression, Object Detection, Behavioral Analysis, Computer Vision.

Received: April 11, 2024. Revised: October 13, 2024. Accepted: November 15, 2024. Published: December 16, 2024.

1 Introduction

Detection of entrance and exit events can be useful and even critical in domains such as security, crowd management, and retail analytics. For example, it can be very useful to detect the entrance of people to a certain area for security or safety-related causes. This type of application is helpful in decision-making, and resource allocation, and therefore it is crucial to make accurate detections. This research proposes an activity detection model that focuses on detecting the entrance and exit events of individuals by integrating two main components: deep learning-based object detection and mathematical modeling using linear regression. Object Detection is a task that can be solved using deep learning, YOLO is such a framework that proved its efficiency and accuracy in localizing objects with high accuracy and in real time, [1]. When it comes to mathematical modeling, Linear regression has long been employed to predict patterns and trends within datasets, [2]. This paper contributes new knowledge to the field of computer vision by integrating deep learning with mathematical modeling to tackle the difficulties in detecting entrance and exit events. Through this combination, we seek to achieve two main objectives:

1. Deep learning-based object detection techniques perform more precise localization and tracking of individuals within complex scenes, using a reliable algorithm to track individuals leads to better detection of entrance and exit events.
2. Traditional methods often struggle in dynamic real-world environments with occlusions, varying lighting conditions, and intricate movement patterns, [3]. We aim to develop a method that can robustly operate in such environments.

To the best of our knowledge, most of the research that was conducted on dynamic environments did not consider the activity detection of individuals' entry and exit. Instead, most of the research focused on other real-world problems such as traffic direction detection.

While there have been efforts such as [4], [5] that focus on automating entrance and exit surveillance in areas where cameras are restricted for privacy reasons, like washrooms and changing rooms, their approach uses video frames from entrance areas, analyzing RGB color histogram variations to detect entrance and exit events. This method encounters difficulties in environments with significant lighting changes or where individuals wear similarly colored clothing, which can result in misclassifications or false detections. By selecting specific grids in the camera view and employing temporal analysis, the research aims to confirm and classify events such as Entry, Exit, or Miscellaneous. The method relies on continuous learning for grid selection in dynamic environments limits its ability to adapt to changing scenes or crowded spaces.

Studies akin to our work can be observed within the employment of deep learning for detecting moving violations in vehicular scenarios, [6], [7]. Intelligent Transportation System (ITS) proposed in [8] detects vehicles that travel in the wrong-way direction and notifies the monitoring center. Although this system achieves high accuracy, it might struggle in dynamic traffic situations and in environments with complex road layouts affecting its real-time performance. Another study proposed in [9] combines YOLO for object detection and Gradient Boosting Machine to prioritize emergency vehicles in a faster and more accurate manner by addressing issues related to shadow problems, and real-time execution of existing models. Although it achieves high accuracy, it might struggle to accurately identify vehicles, especially emergency vehicles in complex environments with more difficult conditions like challenging weather patterns with limited visibility. The system proposed in [10] detects vehicles driving against the traffic flow by using the YOLO framework for vehicle detection, and centroid tracking within a defined region of interest. This system's drawbacks might be in highly congested traffic scenarios or where vehicles frequently change lanes which pose challenges for the system and potentially leads to false positives or missed detections of wrong-way vehicles. Our research tackles the problem of detecting the entrance and exit events of individual which differs significantly from detecting vehicles that are traveling in wrong directions, and that's due to the different environments of each task where the movement patterns and its predictability are different. Individual entrance and exit events detection focuses on identifying and monitoring movements within specific zones, such as entryways or restricted areas, and analyzing the flow of individuals entering or exiting these spaces. This task requires tracking and differentiating between various directional movements within a controlled environment. On the other hand, detection of wrong-way travel typically involves identifying vehicles or pedestrians moving against the designated traffic flow on roads or pathways.

When comparing our method to others, a high accuracy of 91.98% in wrong-way drivers' detection was achieved in [8], which works great for structured environments like roads where the movement of the tracked objects is predictable. However, our method is built for the movement of people, which are often less predictable environments. While the approach in [11] also did well with vehicles using PTZ cameras, these approaches don't quite handle the complexity of human movement like ours does. The method in [4] relied on color variations, but that falls short when lighting changes or people wear similar colors. On the other hand, the approaches in [6] and [12] are strong in vehicle tracking, but they don't adapt as well to the intricate, varied human activity that our method excels in.

The key difference lies in the context and nature of the movement being observed. While both involve tracking movement patterns, dynamic entrance and exit detection focuses on environments where movement patterns of tracked objects are less predictable with typically involve lower risk if an event is misclassified, whereas wrong-way detection deals with more predictable movement patterns, like those on roads and pathways, where misclassification can pose significant risks.

Finally, our study distinguishes itself from previous research by introducing its innovative approach that focuses on activity detection by combining object detection and linear regression techniques to tackle the challenges of dynamic entrance and exit detection challenges. In contrast to traditional methods that often struggle in dynamic environments by applying linear regression to the detected individual's coordinates. Our method achieves higher accuracy in differentiating entrance and exit events, making it particularly well-suited for dynamic environments.

In this paper, we first outline the methodology used in our study, then, we present the results and analysis of our findings. Subsequently, we discuss the implications of our results and their significance in the broader context. Finally, we conclude with suggestions for future research directions.

2 Methodology Enhanced Entrance-Exit Detection using Object Detection and Linear Regression

The methodology of our approach integrates deep learning-based object detection and tracking and linear regression techniques to enhance entrance and exit detection accuracy within dynamic environments. Our approach includes data collection, cleanup, analysis, and decision-making, with each step helping to enhance our entrance and exit detection.

2.1 Data Acquisition and Preprocessing

The initial step of our approach involves acquiring spatial information from video frames using the YOLOv5 object detection model, individuals are localized within the scene at each frame producing a series of bounding boxes of individual locations.

In the next step, the StrongSORT tracking algorithm [6] is applied to keep consistent object tracking across frames, resulting in a set of trajectories, each characterized by a sequence of bounding box coordinates over time which is fed to a windowing approach to extract movement patterns effectively. In the windowing approach, sequences of n-consecutive frames are selected, and for each frame, the corresponding y-coordinates of the tracked object centers are recorded. This windowed dataset captures localized movement tendencies, enhancing the granularity of our subsequent analysis.

2.2 Feature Extraction and Linear Regression

From the acquired y-coordinate data, pairwise distances are computed between all points within each window. The selection of points with minimum pairwise distances aims to identify coherent movement trajectories while filtering out potential noise and outliers. With the selected y-coordinate points, a linear regression model is constructed. This model captures the underlying linear relationship between time and vertical displacement. Specifically, we express the direction of movements as a function of time $(t_1, t_2, \dots, t_n)$, $y-$coordinates $(y_1, y_2, \dots, y_n)$ using the equation:

$$a(t, y, n) = \frac{3}{n(n+1)(2n+1)} \left[\sum_{i}^{n} 2y_i t_i - (1+n)y_i\right] \quad (1)$$

describes the tangent of the linear regression line which is the rate of vertical movement. We compute the distance between the coordinates and eliminate closes values to make the approach more stable and to identify coherent trajectories. We achieved that using a function $\eta(x)$ that defined the threshold:

$$\eta(x) = \begin{cases} 0, & |x| < 1 \\ 1, & |x| \geq 1 \end{cases} \quad (2)$$

$$a(t, y, n, h) = \frac{3}{n(n+1)(2n+1)} \left[2y_1 - (n+1)y_1 + \sum_{i=2}^{n} 2y_i t_i - (n+1)y_i\right] = \frac{3}{n(n+1)(2n+1)} \left[(1-n)y_1 + \sum_{i=2}^{n} \left((2y_i t_i - (n+1)y_i) \cdot \eta\left(\left\|\frac{y_i - y_{i-1}}{h}\right\|^2\right)\right)\right] \quad (3)$$

where h denotes the threshold.

Our algorithm steps are described as follows:

1. YOLO5 is used for object detection and localization.
2. A StrongSORT tracker used to track objects across video stream frames.
3. Sequential frames are grouped in windows of frames each for localized data acquisition.
4. Calculate pairwise distances between y-coordinates to identify coherent trajectories.
5. Build linear regression models for selected y-coordinate data in order to capture linear relationships.
6. Extract the tangent 'a' from the linear regression equation.
7. Apply threshold h to 'a' for precise entrance/exit identification.

3 Experiments and Results

To rigorously evaluate the effectiveness of our proposed approach for entrance and exit detection, we conducted a series of comprehensive experiments encompassing both quantitative assessments and qualitative analysis. These experiments aimed to ascertain the accuracy, robustness, and real-world applicability of our methodology, showcasing its potential to address challenges such as occlusions, camera locations, filming angles, tracked objects' distance from the camera, and complex movement patterns.

3.1 Experiments Setup

We employed a diverse dataset comprising video sequences captured in dynamic environments (i.e., various camera locations, altitude, and the angle at which the video sequences were captured). In addition, the video sequences include corridors, entrances, and exits. The dataset encompassed scenarios characterized by camera view restrictions and multi-directional movement patterns, thereby simulating real-world complexities.

For the experiment purposes, we captured video sequences in an unoccupied space that we segmented into three rows: front, center, and back in relation to the entrance/exit location (Figure 1). We further subdivided each row into 3 locations: left, center, and right in relation to the camera's location. In addition, the video sequences captured in each location were characterized by three altitudes: high, middle, and low in relation to the ground. We also captured video sequences in which the camera's capturing angle was subtly adjusted at each location at the altitude which was categorized as high. For simplicity, we divided the various camera locations into a 3x3 grid, where each cell has three video sequences, and the camera is either located at a high, middle, or low altitude, in addition to the video sequence that the camera's capturing angle was subtly adjusted, the angle adjustments occur only on the right and left columns of the grid. We have 27 locations, 3 videos were captured at each location, in addition to 6 videos with capturing angle adjustment. To simulate real-world scenarios, we applied two entrance, and two exit scenarios that simulate real-world scenarios: straight movement of the tracked individual towards or against the location of the camera which makes in a total of 66 videos of entrances, and 66 videos of exits, 132 in total that was tested.

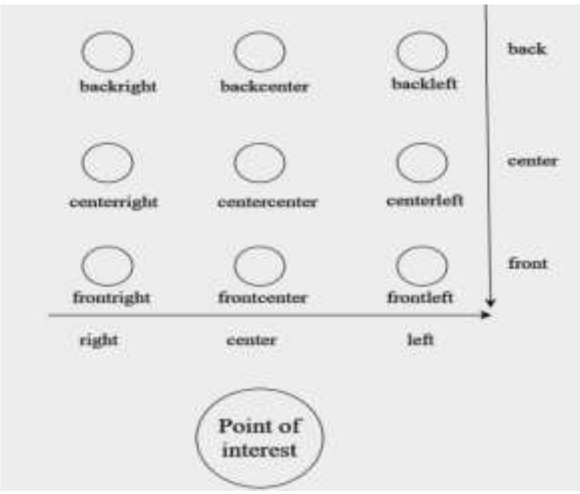

Fig. 1: Camera Location Grid: we captured at every camera location the entrance and exit video at 3 different altitude levels: high, middle, and low, in addition to angle-adjusted videos at left and right locations at level high

3.1.1 Video Sequence Characteristics

Our research included 133 videos where each video is an entrance scenario or an exit scenario, our dataset includes the following statistics:

1. The videos were captured with a constant frame rate of 20 frames per second (fps).
2. The videos range from 40-frame videos to longer ones consisting of 180 frames, which makes it range from 2 seconds videos up to 9 seconds videos, while the mean duration is approximately 4.84 seconds.
3. The videos were captured with a frame width of 704 pixels and a height of 576 pixels.

3.1.2 Compared Methods

Line Partition Method: This approach is based on a scene division strategy using a designated line, the line was horizontally at 60% of the width. If tracked individuals are found above the line at the end of a video sequence, then the scenario is interpreted as an entrance, while their presence below the line signifies an exit. While this method is simple and easy to implement, it heavily depends on the exact placement of the partition line and fails to consider the specific movement patterns on either side of the line. In contrast, our method leverages tangent values for nuanced direction detection, allowing for greater accuracy in complex scenarios.

Bounding Box Diameter Method: In this method, we captured the dimensions of the bounding box obtained from YOLOv5 object detection over each frame in the video sequence. Next, a window approach is applied to calculate the mean of these bounding box diameter values over 20 consecutive frames. At each frame, the mean value of the bounding box diameters in the frame window is updated. If the new mean is smaller than the previous one, the method suggests that the event is an exit, whereas if it is larger, it indicates an entrance. The scenario is classified based on the last two means, if the last one is larger than its predecessor then the scenario is classified as exit, otherwise, it is classified as entrance.

3.2 Experiments

We have evaluated the performance of our proposed approach by comparing it with two alternative methods mentioned in the previous section: the Line Partition, and the Bounding Box Diameter methods.

To evaluate our method, we formulated performance metrics to measure the accuracy of our entrance and exit detection method. We use metrics such as F1-score and accuracy to gauge the method's ability to accurately detect entrances and exits in comparison to ground truth annotations. We calculate these metrics across a range of dynamic environment scenarios to assess the method's robustness and consistency. Each method's ability was evaluated to detect entrances against its ability to detect exits, in addition to the performance of each method at each camera location.

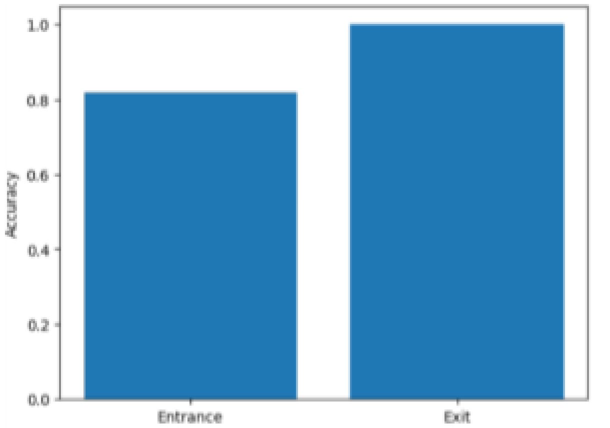

Fig. 2: Linear Regression-based method's accuracy in detecting entrances vs. exits showing accuracy comparison in a straight motion

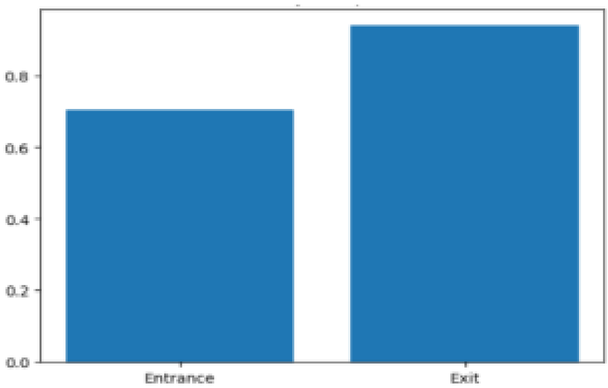

Fig. 3: Linear Regression-based method's accuracy in detecting entrances vs. exits showing accuracy inside motion

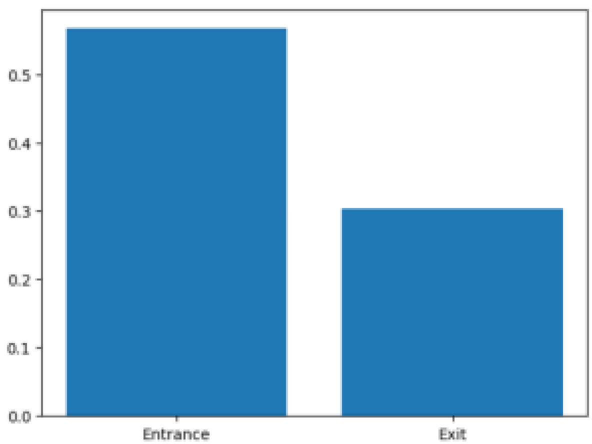

Fig. 4: Line Partition method's accuracy in detecting entrances vs. exits showing overall accuracy scores for our method

As part of our experiments, the performance of each method was compared to detect entrance and exit while the tracked individual's movement is vertical, meaning it is moving towards the location of the camera in the entrance or against the location of the camera in the exit. As can be seen in Figure 2, Figure 3 and Figure 4. The accuracy score achieved by our method is 90.9%, and the f1-score is 0.9 in a straight motion, while in side-motion it achieved 81.8% accuracy and f1-score of 0.81. The overall accuracy score of our method has achieved 86.36%, and an f1-score of 0.86.

Figure 5, Figure.6 and Figure 7 show the results of the Bounding Box Diameter method. The accuracy score achieved is 71.21%, and the f1-score is 0.68 in a straight motion, while in side-motion it achieved 56.72% accuracy and an f1-score of 0.5. The overall accuracy score that this method has achieved is 63.91%, and the f1-score of 0.59.

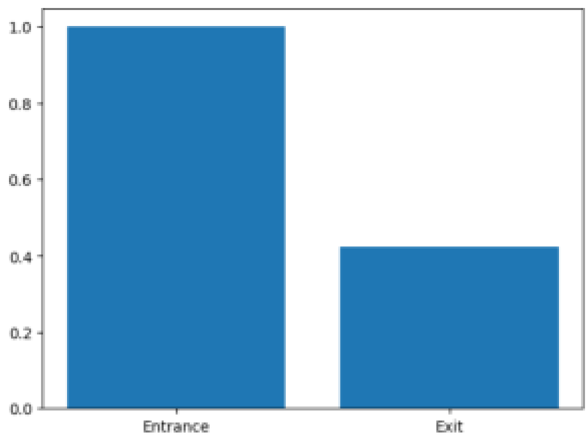

Fig. 5: Bounding Box Diameter method's accuracy in detecting entrances vs. exits showing accuracy comparison in straight motion

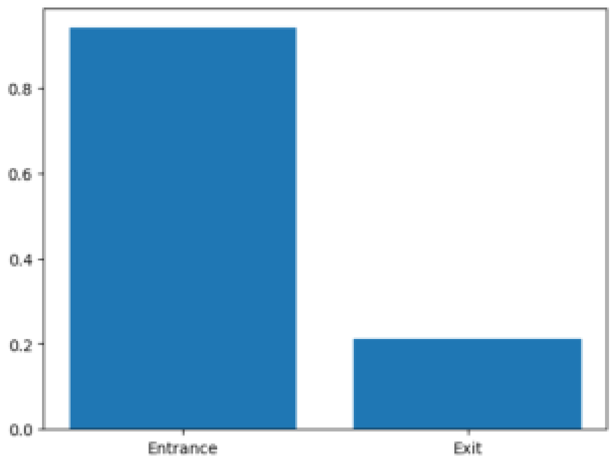

Fig. 6: Bounding Box Diameter method's accuracy in detecting entrances vs. exits showing accuracy on side motion

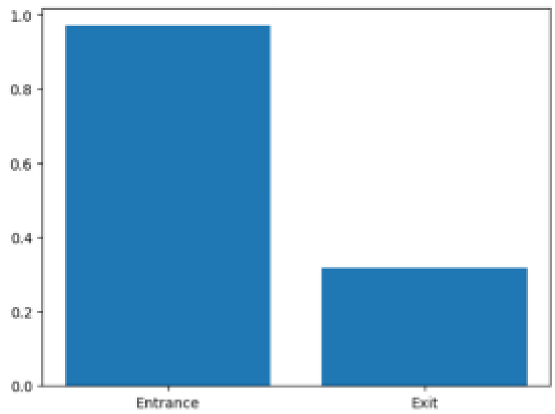

Fig. 7: Bounding Box Diameter method's accuracy in detecting entrances vs. exits showing overall accuracy scores for our method

Next, a comparison between all the methods was conducted to evaluate each method's exit detection, entrance detection, and overall performance against each other. As can be seen in Figure 8 and Figure 9, the linear regression-based algorithm outperforms the other two methods in both metrics F1-score and Accuracy. This is to be expected because the linear regression-based algorithm considers better the dynamic nature of scenarios captured by the tested videos.

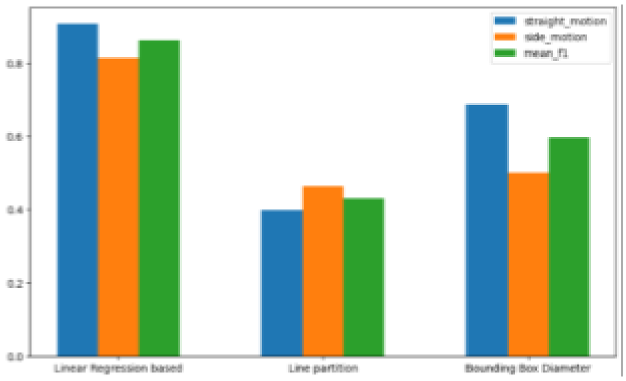

Fig. 8: F1-Score comparison between methods

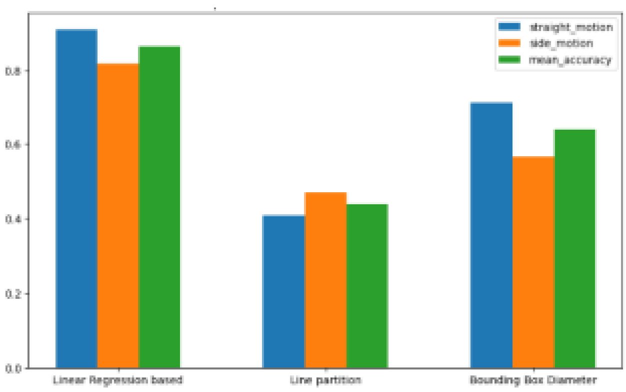

Fig. 9: Accuracy Comparison for different methods

3.3 Qualitative Analysis

In addition to quantitative measures, we conducted a qualitative analysis by visualizing the trajectory predictions generated by our linear regression algorithm. Through these visualizations, we discerned that our approach successfully captured intricate movement patterns, effectively distinguishing between entrance and exit directions. Figure 10, Figure 11 and Figure 12 visually depict how each of the discussed methods has been applied to scenarios involving entrances and exits.

This observation was particularly evident in scenarios featuring multiple individuals and occlusions, where traditional methods often faltered. We observe that each method is better at one of the scenarios examined. As expected, our method has achieved higher accuracy scores in detecting exit events than in entrance events, while the other two methods had higher accuracy scores in detecting entrance events than exit. This is because the duration time of the exit activity is shorter than the entering activity in the tested videos, making the detection difficult for the other methods to detect the individual in all required video frames. As a result, other methods are not able to capture the exit activity. Our method tracks the individual in the video frames and based on his movement direction it detects the correct activity.

In all metrics, and across all scenarios, our method performs best, followed by the Bounding Box Diameter method, while the Line Partition method falls behind. The goal of the experiments that we conducted was to examine the ability of different methods to detect entrances and exits with minimum human intervention by applying the same configuration of each method on different camera locations, altitudes, and angles. This constraint is a main cause that makes Line Partition method struggle in many different locations, altitudes, and angles.


Fig. 10: On the left Line Partition method on straight motion entrance scenario in a location with low altitude. On the right Line Partition method exit scenario in a location with high altitude

Fig. 11: On the left Bounding Box Diameter method on side motion entrance scenario in a location with middle altitude. On the right Bounding Box Diameter method side motion exit scenario in a location with low altitude

Fig. 12: On the left Linear Regression based method on straight motion entrance scenario in a location with high altitude and adjusted angle. On the right Linear Regression method side motion exit scenario in a location with high altitude and adjusted angle

4 Discussion

In terms of the Line-Partition approach, although it is a widely used method for activity detection, it exhibits notable limitations. Firstly, it relies on manual human intervention to define entrance and exit zones within the monitored area, where if the tracked individuals are present in an exit zone the event is considered to be an exit, otherwise it's the entrance. This logic may not seamlessly integrate into dynamic monitoring environments and might necessitate the intervention of humans at each change of environment settings. Without this human intervention, and similar to the findings in [5], the method struggles to accurately distinguish true entrances and exits in environments with variability in movements, potentially leading to false detections. Secondly, the division of the monitored area into entrance and exit areas leads to disregarding important information which might be crucial for activity detection and that's because it does not track the exact movement of the individuals inside the predefined area. A better approach to tackle this problem is the Bounding Box Diameter approach, which is not dependent on human intervention and achieves better results in dynamic environments. The main disadvantage of this approach is that it is not directly motion-based, instead, it relies on the bounding box dimensions resulting by object detection which can sometimes exhibit sudden and unexpected changes that do not accurately represent the motion of the tracked objects making it not highly stable and is likely to face challenges when the movement patterns of individuals are less predictable. The results we conducted in this research demonstrate the effectiveness and the dynamism of our approach by achieving extraordinary results in automatic entrance and exit events detection in dynamic environments where the movement patterns of individuals are less predictable and higher adaptability to dynamic environments than prior studies such as [8].

5 Conclusions

In this paper, we introduce a novel approach to entrance and exit detection in dynamic environments by integrating deep learning-based object detection with mathematical linear regression. Commonly used methods that address this problem, namely the Line-Based method and the Bounding Box Diameter method, struggle in dynamic environments where the movement patterns of the monitored individuals are less predictable. Our approach demonstrates high adaptability to various environment settings with no manual human intervention while achieving superior results over commonly used methods with an overall accuracy of 86.36%, and an F1-score of 0.86, making it a suitable option for complex real-world environments that introduce additional factors such as occlusions, varying lighting conditions, and complex movement patterns which challenge both existing methods and systems proposed by other studies which are more suitable for structured environments such as roadways where the movement patterns of the tracked objects are more predictable than the movement patterns of human individuals. In addition, the adaptability of our approach while maintaining good performance makes our approach well-suited for applications in domains such as security, crowd management, and retail analytics, where precise monitoring of human movement is critical. Future research could explore the adaptation of this method to other areas, including behavior analysis, vehicular movement analysis, or real-time monitoring in public spaces, to further validate its effectiveness and versatility. We also show that the integration of deep learning and mathematical modeling holds significant potential for advancing the field of computer vision and developing more robust and accurate detection systems. This work lays the foundation for further innovations in intelligent monitoring technologies, paving the way for enhanced applications in increasingly complex environments.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work the authors used ChatGPT developed by OpenAI in order to perform minor revisions to enhance readability if certain phrases. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References:

[1] Ultralytics, *YOLOv5: "A state-of-the-art real-time object detection system"*, 2021.

[2] Hocking, R. R. "Developments in Linear Regression Methodology: 1959-1982." *Technometrics*, vol. 25, no. 3, 1983, pp. 219–230. DOI: https://doi.org/10.2307/1268603. Accessed 21 Dec. 2023.

[3] Majumder, M., & Wilmot, C. "Automated Vehicle Counting from Pre-Recorded Video Using You Only Look Once (YOLO) Object Detection Model." *J. Imaging*, vol. 9, no. 7, 2023, p. 131. DOI: https://doi.org/10.3390/jimaging9070131.

[4] Vinay Kumar V. & P. Nagabhushan. *"Monitoring of people entering and exiting private areas using Computer Vision."* arXiv, 2019. DOI: https://doi.org/10.48550/arXiv.1908.00716.

[5] Vinay Kumar V., & P. Nagabhushan. "Entry-Exit event detection from video frames." *International Journal of Computer Sciences and Engineering*, vol. 6, no. 2, 2018, pp. 112-118. http://dx.doi.org/10.26438/ijcse/v6i2.112118.

[6] Yunhao Du, Zhicheng Zhao, Yang Song, Yanyun Zhao, Fei Su, Tao Gong, & Hongying Meng. *"StrongSORT: Make DeepSORT Great Again."* arXiv, 2023. DOI: https://doi.org/10.48550/arXiv.2202.13514.

[7] Manasa, A., & Renuka Devi, S. M. "An Enhanced Real-Time System for Wrong-Way and Over Speed Violation Detection Using Deep Learning." *Fourth International Conference on Image Processing and Capsule Networks*, 2023, pp. 309-322. DOI: https://doi.org/10.1007/978-981-99-7093-3_21.

[8] Usmankhujaev, Saidasul, Shokhrukh Baydadaev, & Kwon Jang Woo. "Real-time, deep learning based wrong direction detection." *Applied Sciences*, 2020. DOI: https://doi.org/10.3390/app10072453.

[9] Vikruthi, S., Archana, M., & Tanguturi, R. C. "A Novel Framework for Vehicle Detection and Classification Using Enhanced YOLO-v7 and GBM to Prioritize Emergency Vehicle." *International Journal of Intelligent Systems and Applications in Engineering*, vol. 12, no. 1S, 2023, pp. 302-312.

[10] Bharadwaj, R., Billade, A., Chenna, S., Chandrapatle, A., & Chinchalpalle, G. "Wrong Way Vehicle Detection in Single and Double Lane." *IJRITCC*, vol. 11, no. 6S, 2023, pp. 457-462. DOI: https://doi.org/10.17762/ijritcc.v11i6s.6953.

[11] Haghighat, A., & Sharma, A. "A computer vision-based deep learning model to detect wrong-way driving using pan–tilt–zoom traffic cameras." *Computer-Aided Civil and Infrastructure Engineering*, vol. 38, no. 1, 2023, pp. 119–132. DOI: https://doi.org/10.1111/mice.12819.

[12] Suttiponpisarn, P., Charnsripinyo, C., Usanavasin, S., & Nakahara, H. "An Autonomous Framework for Real-Time Wrong-Way Driving Vehicle Detection from Closed-Circuit Televisions." *Sustainability*, vol. 14, no. 16, 2022. DOI: https://doi.org/10.3390/su141610232.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed in the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

No funding was received for conducting this study.

Conflict of Interest

The authors have no conflicts of interest to declare.