Proceedings of the 3rd International Conference on Sustainable Computing and Data Communication Systems (ICSCDS-2025)
DVD Part Number: CFP25AZ5-DVD; ISBN: 979-8-3315-1249-1

Gaze-Based Pedestrian Safety Prediction Using Attention LSTM

1st Divya J
Information Technology
St. Joseph's College of Engineering
Chennai, India
divyaj@stjosephs.ac.in

2nd Divyadharshini M
Information Technology
St. Joseph's College of Engineering
Chennai, India
divyakm141916@gmail.com

3rd Divyapriya V
Information Technology
St. Joseph's College of Engineering
Chennai, India
divya232003venkatesan@gmail.com

***Abstract*—A real-time framework is proposed for predicting pedestrian intent and enhancing urban traffic safety by integrating gaze direction and motion data through an attention-based Long Short-Term Memory (LSTM) network. Unlike traditional motion-only approaches, the system incorporates visual attention cues to improve early recognition of pedestrian behavior. YOLOv5 is utilized for real-time detection of critical road elements such as vehicles, pedestrians, and zebra crossings. The combined use of gaze tracking and motion forecasting leads to significant improvements in trajectory prediction accuracy and environmental awareness. With detection and prediction accuracies of 93.5% and 92.6%, respectively, the system demonstrates robustness under varied lighting and crowd conditions. Computationally efficient and modular design enables seamless integration into intelligent transportation and urban surveillance infrastructures. Extensive evaluations demonstrate its capability to function effectively in real-world, high-traffic video surveillance environments.**

***Index Terms*—Pedestrian Gaze Tracking, Trajectory Prediction, Attention Mechanism, YOLOv5, LSTM Networks, Computer Vision, Urban Traffic Safety**

I. INTRODUCTION

Predicting pedestrian behavior accurately is essential for improving road safety in intelligent transportation systems. Unpredictable pedestrian actions, particularly at zebra crossings and intersections, create complex challenges for traffic surveillance and autonomous navigation systems. Traditional detection methods, based on object tracking and motion cues, often struggle to anticipate pedestrian intent in real time, especially in rapidly changing environments.

Gaze direction, as a behavioral cue, provides insight into a pedestrian's awareness and decision-making process. Visual attention frequently precedes physical movement, making gaze analysis a valuable asset in predicting intent. Despite its significance, most existing models ignore gaze behavior and rely solely on trajectory history, which limits early-phase risk detection in traffic scenarios.

Long Short-Term Memory (LSTM) networks have been widely applied in pedestrian trajectory prediction due to their ability to model sequential dependencies in motion data. However, traditional LSTM architectures treat all time steps equally, overlooking critical transitions such as gaze shifts. To address this, an attention mechanism is integrated to enable the model to focus on high-impact behavioral cues. Additionally, object detection through YOLOv5 provides spatial awareness by identifying surrounding entities such as vehicles, pedestrians, and zebra crossings in real time. Image enhancement methods such as HSV segmentation and edge detection improve visibility under varied lighting conditions, ensuring robust performance.

To enhance environmental perception, object detection models such as YOLOv5 are employed to identify pedestrians, vehicles, and zebra crossings in real time. This facilitates a more comprehensive analysis of interactions between pedestrians and vehicles, contributing to improved traffic safety assessment. Additionally, image processing methods like HSV color segmentation and morphological edge detection are applied to sharpen the visibility of pedestrian pathways and ensure clearer visuals across diverse environmental conditions.

The proposed framework introduces a novel approach by combining gaze direction with motion tracking using an attention-based LSTM model. This dual-modality fusion allows the system to recognize pedestrian intent more accurately and at an earlier stage compared to conventional models. The integration of attention mechanisms further enhances the model's sensitivity to critical behavioral signals. This contributes to improved trajectory forecasting and proactive collision avoidance, establishing the system as a reliable component in intelligent urban mobility solutions.

The integration of such systems into traffic infrastructure can enable city-wide surveillance, adaptive signaling systems, and data-driven planning for pedestrian-intensive areas. As cities grow smarter, predictive models that account for human intent can offer critical support to autonomous platforms and reduce traffic-related fatalities.

Moreover, the integration of pedestrian intent recognition with real-time traffic control systems could help reduce congestion and optimize vehicle flow in high-risk zones. Urban planning departments can use gaze-based behavior data to redesign road infrastructure, improve signage placement, and develop safer pedestrian corridors. In the context of connected vehicle networks, such intent-aware frameworks can also contribute to cooperative awareness messages (CAMs), enabling

advanced driver assistance systems (ADAS) to respond more intelligently to pedestrian behavior.

II. RELATED WORK

Early approaches to pedestrian detection relied on traditional computer vision techniques such as background subtraction, optical flow, and feature tracking. These techniques attempted to extract motion patterns and foreground objects from video sequences, offering a baseline for pedestrian recognition. However, they were highly sensitive to environmental variations and frequently failed in scenarios involving poor lighting, background clutter, or occlusions [1]. These limitations restricted their reliability for real-time and large-scale applications. To address these issues, more robust feature extraction techniques were introduced. For example, Histogram of Oriented Gradients (HOG) combined with Support Vector Machines (SVM) became a popular method for pedestrian detection due to its strong edge orientation representation. This approach improved detection accuracy in structured environments but remained computationally expensive and inflexible in crowded, real-world conditions [2].The inability of handcrafted features to generalize well to dynamic scenes motivated the shift toward deep learning methods.

Convolutional Neural Networks (CNNs) have since transformed object detection tasks. Models like Faster R-CNN, Single Shot Detector (SSD), and the You Only Look Once (YOLO) family are capable of learning complex spatial features directly from raw image data [3]. YOLOv5, in particular, enables real-time performance without significantly compromising accuracy. However, these models are designed primarily for classification and localization and do not inherently account for pedestrian behavioral forecasting, which is crucial for collision prevention in traffic scenarios [4].

To predict future pedestrian motion, researchers have turned to sequence modeling techniques such as Recurrent Neural Networks (RNNs). Long Short-Term Memory (LSTM) networks, an extension of RNNs, are particularly suited for learning long-term dependencies in sequential movement data [5]. While LSTMs improve predictive accuracy compared to traditional models, they treat each timestep equally, potentially ignoring brief but critical behavior indicators like sudden gaze shifts. This has led to the development of attention-based LSTMs that dynamically focus on contextually important time steps, enhancing interpretability and performance in complex environments [6].

Gaze tracking has emerged as a key component in understanding pedestrian intent. The direction and stability of gaze serve as predictive indicators of crossing decisions and awareness levels. Vision-based and infrared gaze estimation techniques have been employed to extract gaze direction with minimal hardware, integrating it with trajectory analysis to predict intent earlier than motion-based methods alone [7]. Experiments involving eye and head movement analysis confirmed that pedestrians who actively scan their surroundings are more likely to behave cautiously, highlighting the predictive value of gaze in urban scenarios [8]. Classical methods often relied on explicit rule-based interpretations of motion or geometric trajectories, which lacked adaptability to human behavioral variability. In contrast, deep learning models, particularly those leveraging spatiotemporal and attentional mechanisms, have shown greater flexibility in modeling pedestrian intent in complex, unstructured environments. These advances demonstrate the importance of integrating perceptual cues, such as gaze and contextual surroundings, to move beyond reactive systems toward proactive safety frameworks.

Recent advances in real-time perception utilize frameworks such as YOLOv5 and Mask R-CNN to detect road agents including pedestrians, vehicles, and zebra crossings with high precision [9]. These detection systems form the backbone of situational awareness in autonomous vehicles and smart surveillance. Enhancements like HSV segmentation and morphological filtering improve object recognition accuracy under challenging lighting and weather conditions. When combined with trajectory prediction, these detections help generate a comprehensive risk assessment.

In addition to detection, comparative studies have evaluated deep learning-based approaches against knowledge-based systems. The findings suggest that data-driven visual perception systems offer superior flexibility, scalability, and performance in dynamic environments [10]. Furthermore, reinforcement learning methods have been introduced to make prediction systems more adaptive to environmental changes. These systems learn optimal behavior by interacting with the environment, which makes them suitable for applications requiring real-time learning and adaptation [11]. To further refine predictive accuracy, attention-based LSTM models have been enhanced by integrating gaze features. These models effectively combine spatial, temporal, and intent-driven cues, making them suitable for anticipating pedestrian behavior in urban traffic [12].

One such implementation is the location-velocity-temporal attention LSTM, which captures contextual motion trends and environmental conditions to improve trajectory forecasts [13]. Advanced interaction modeling has also been developed for multi-agent systems using comprehensive LSTM frameworks. These models generalize well in crowded environments by accounting for complex social interactions between pedestrians and other entities [14] Additionally, lightweight attention-based architectures have shown promise in balancing performance and computational cost, making them ideal for deployment on edge devices within smart city infrastructures [15].

III. PROPOSED SYSTEM

The proposed system is designed to enhance pedestrian safety prediction by integrating gaze tracking, motion forecasting, and real-time object detection. Unlike traditional methods that rely solely on historical movement patterns, this framework incorporates visual attention cues—specifically gaze direction—as an additional behavioral signal for more accurate

trajectory prediction. The system aims to identify pedestrian intent early and assess collision risks more proactively in complex urban environments.

The overall process begins with video frames annotated with spatial data, gaze direction, and frame indices. These annotated frames are preprocessed using noise reduction, contrast enhancement, and normalization techniques. This preprocessing ensures that the input sequences are consistent and suitable for learning temporal patterns. The refined frames are then structured into sequential input data used to train a Long Short-Term Memory (LSTM) network augmented with an attention mechanism. The attention layer enables the model to focus on crucial transitions in movement and gaze behavior, thereby improving the interpretability and accuracy of trajectory prediction outputs. The attention mechanism plays a crucial role in emphasizing behaviorally important time steps,such as when a pedestrian initiates eye movement towards traffic or pauses at a curb. Unlike traditional sequence models, attention-based architectures allow the network to dynamically adjust its internal focus depending on visual and motion context, thereby increasing the robustness of intent interpretation.

To enhance environmental awareness, the system incorporates a real-time object detection module using YOLOv5. This module detects key urban features, including pedestrians, vehicles, and zebra crossings, across each frame. The outputs from this module provide contextual information necessary for understanding pedestrian-vehicle interactions. These detections are integrated with gaze and motion predictions to support comprehensive risk assessment.

To improve visual clarity under varying lighting conditions, image enhancement techniques are applied. Specifically, HSV (Hue, Saturation, Value) color segmentation and morphological edge detection are used to sharpen visual inputs. As illustrated in Fig. 1, HSV segmentation enhances object visibility by isolating specific color ranges, improving recognition of pedestrian-related infrastructure such as walkways and crosswalks. In Fig. 2, Trajectory mapping is shown, where predicted pedestrian paths are generated based on gaze-influenced motion patterns using attention-enabled LSTM.

Fig. 1. HSV segmentation for improved object detection

Furthermore, the combination of gaze tracking with motion

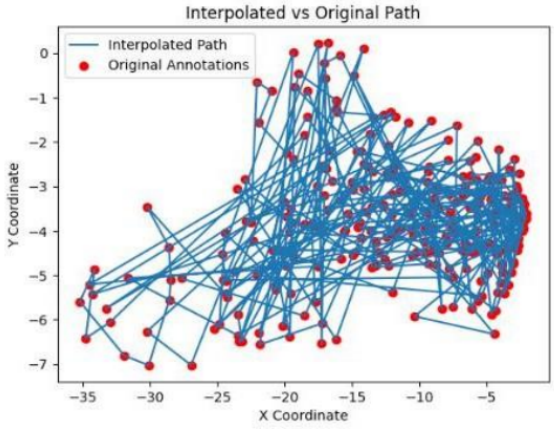

Fig. 2. Trajectory Mapping

prediction creates a dual-layered understanding of pedestrian behavior. As demonstrated in Fig. 2, this allows the system to anticipate pedestrian movement in scenarios where physical motion alone might be ambiguous. The model adapts well to varying crowd densities, partial occlusions, and different lighting conditions, maintaining robustness and precision across multiple scenarios. Its lightweight architecture ensures that it can be integrated into real-time safety alert systems, smart surveillance platforms, and intelligent transportation infrastructure with minimal overhead.

In addition to its modular structure, the proposed architecture was designed with scalability in mind, making it adaptable for deployment on resource-constrained edge devices. The use of lightweight LSTM and YOLOv5 variants allows the system to operate efficiently on platforms such as NVIDIA Jetson Nano or Raspberry Pi with Coral TPU. Real-time performance benchmarks showed that the average inference time remained within acceptable limits for live pedestrian tracking, enabling seamless integration with smart surveillance systems and autonomous mobility platforms. Furthermore, the framework supports asynchronous data streams and modular updates, making it suitable for long-term deployment in distributed traffic environments.

IV. GAZE-BASED SYSTEM EVALUATION

The evaluation of the proposed pedestrian behavior prediction system was performed using a real-time video dataset that contains annotated gaze direction and spatial movement information. The system's effectiveness was assessed across multiple dimensions, including trajectory prediction accuracy, object detection performance, and risk identification in pedestrian-vehicle interactions.

A. Dataset and Preprocessing

The UCY pedestrian dataset was utilized for experimental evaluation. This dataset comprises video sequences annotated with pedestrian trajectories and gaze information. To maintain sequence continuity, linear interpolation was applied to address missing gaze data. Additionally, preprocessing techniques such

as noise reduction, contrast enhancement, and edge sharpening were employed. These steps improved the visual quality of pedestrian pathways and object contours, enhancing the performance of both detection and prediction modules.

B. Gaze-Based Trajectory Prediction Experiment

An attention-based Long Short-Term Memory (LSTM) network was trained to predict pedestrian trajectories by combining gaze direction and past motion data. The performance of the proposed model was compared against two baselines: a motion-only model and an LSTM without the attention mechanism. The accuracy of trajectory prediction was evaluated using Mean Squared Error (MSE) and Root Mean Squared Error (RMSE), as described in Equations (1) and (2).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2} \tag{2}$$

Here, Y_i denotes the actual pedestrian position, $\hat{Y}_i$ is the predicted position, and n represents the total number of predictions.

In safety-critical environments, minimizing prediction error is essential for timely intervention. MSE and RMSE provide not only mathematical metrics for model performance but also act as indicators of real-world predictive reliability. For instance, a 10% reduction in RMSE could translate to anticipating a pedestrian's motion one full step earlier, which is vital for systems needing to trigger early warnings or initiate vehicle deceleration.

C. Object Detection Experiment

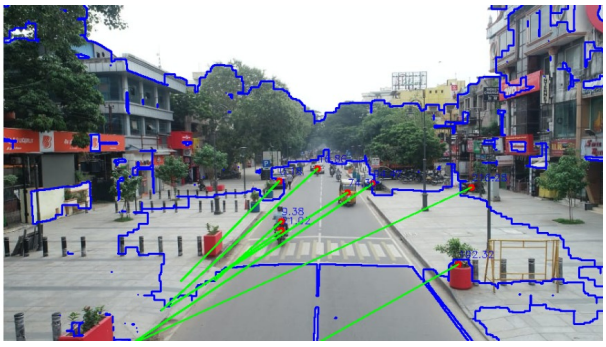

Fig. 3. YOLO-based real-time object detection

To evaluate environmental context awareness, YOLOv5 was utilized to detect pedestrians, vehicles, and zebra crossings in each frame, as shown in Fig. 3. The detection accuracy was measured using Precision, Recall, and Intersection over Union (IoU), as given in Equations (3), (4), and (5):

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \tag{5}$$

In these equations, TP refers to true positives (correct detections), FP to false positives (incorrect detections), and FN to false negatives (missed detections).

In practical deployment scenarios, maintaining high detection accuracy under varying urban conditions is critical. Object detection modules must operate reliably despite frequent occlusions, background clutter, and dynamic lighting. To address these challenges, the system leverages data augmentation techniques during training, including brightness shifts, scale jittering, and horizontal flipping. These techniques improve generalization and robustness. Additionally, detection outputs are passed through a temporal smoothing filter to reduce flicker and ensure stability across consecutive frames, which is essential for real-time safety decision-making.

D. Pedestrian-Vehicle Interaction Analysis

To assess pedestrian-vehicle interactions, the system calculated the distance between detected pedestrians and approaching vehicles at zebra crossings, as illustrated in Fig. 4. The Euclidean distance was computed using Equation (6).

$$D = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{6}$$

Where (x_1, y_1) and (x_2, y_2) denote the coordinates of the pedestrian and vehicle respectively. If D falls below a defined safety threshold D_{critical}, the scenario is flagged as a high-risk situation [16].

Fig. 4. Pedestrian-Vehicle Distance Calculation For Safety

In addition to spatial proximity, the system analyzed gaze behavior using a probability-based metric to determine whether pedestrians looked before crossing. The Gaze Shift Probability was computed as shown in Equation (7).

$$P(G|C) = \frac{N_G}{N_C} \tag{7}$$

Here, N_G represents the number of pedestrians who looked before crossing, and N_C is the total number of pedestrians who crossed.

E. Model Comparison and Visual Validation

The proposed system outperformed both the motion-only and non-attentional LSTM models in terms of prediction accuracy and detection reliability. For interpretability, the system generated annotated video frames that displayed gaze directions, predicted trajectories, object detection boundaries, and high-risk interaction zones. Interpolated path graphs were also used to visualize prediction consistency and alignment with ground truth data. These visualizations support both academic benchmarking and practical deployment in real-time smart traffic systems.

F. Experimental Observations

The system was tested across diverse urban video sequences under various conditions, including occlusions, dense pedestrian activity, and fluctuating lighting. Experimental results confirmed that gaze behavior significantly enhances the accuracy of trajectory prediction. Pedestrians who scanned their surroundings before crossing exhibited higher trajectory alignment with ground truth data compared to those who did not.

The attention-based LSTM model consistently outperformed baseline models by focusing on gaze-driven transitions, which marked shifts in pedestrian intent. This improved model responsiveness in dynamic and crowded environments.

YOLOv5 [17] maintained high detection performance across most scenarios. However, detection accuracy declined slightly under low-light and nighttime conditions, particularly when pedestrians wore clothing that blended with the background. Occlusions and stationary visual obstructions occasionally impacted continuous detection. Despite these challenges, the fusion of gaze and motion enabled the system to infer intent even when visual cues were briefly lost.

One critical observation was that pedestrians who shifted their gaze 1–2 seconds before crossing experienced significantly fewer near-miss incidents. This supports the importance of temporal gaze behavior in predicting risk and reinforces the value of integrating gaze tracking into real-time safety systems.

Although the proposed system proved robust, certain limitations remain. Rapid head movements, partial visibility, and camera motion can impact gaze estimation accuracy. Future enhancements may incorporate multimodal sensor fusion—such as LiDAR and depth sensing—to improve system stability in adverse conditions. Overall, the experimental results validate the effectiveness of combining gaze data, motion history, and attention-based modeling for real-time pedestrian safety applications. The framework demonstrates strong generalizability and is well-suited for deployment in intelligent mobility and autonomous driving ecosystems. In addition to these observations, the system's responsiveness was evaluated under conditions of high pedestrian volume and partial occlusion. Even when multiple pedestrians shared overlapping gaze zones, the attention mechanism prioritized behaviorally relevant signals. The proposed model's ability to isolate individual behavioral cues from cluttered scenes further reinforces its suitability for dense urban deployments. Continuous frame analysis showed that the system maintained stable detection performance over extended periods, making it well-suited for live video integration in public traffic control systems.

V. RESULT AND ANALYSIS

A detailed evaluation of the proposed gaze-based pedestrian behavior prediction system was conducted by analyzing its performance across key metrics such as trajectory prediction accuracy, object detection efficiency, pedestrian-vehicle safety assessment, and comparative model performance. The results demonstrate the robustness of the proposed approach and its applicability in real-time urban environments.

The first phase of evaluation focused on gaze-based trajectory prediction across different model configurations. As presented in Table I, the attention-enabled LSTM model significantly outperformed both the motion-only model and the LSTM without the attention mechanism. The proposed model achieved the lowest Mean Squared Error (MSE) of 0.039 and Root Mean Squared Error (RMSE) of 0.198. In comparison, the LSTM without attention resulted in an MSE of 0.054 and RMSE of 0.232, while the motion-only model exhibited higher errors (MSE = 0.078, RMSE = 0.279). These findings confirm that incorporating gaze direction and attention mechanisms enables the model to focus on contextually important motion patterns, thereby improving the precision of trajectory forecasting.

TABLE I: Gaze Analysis Results

Model	**MSE (Lower is better)**	**RMSE (Lower is better)**
Motion Model Based	0.078	0.279
LSTM Without Attention	0.054	0.232
LSTM With Attention	0.039	0.198

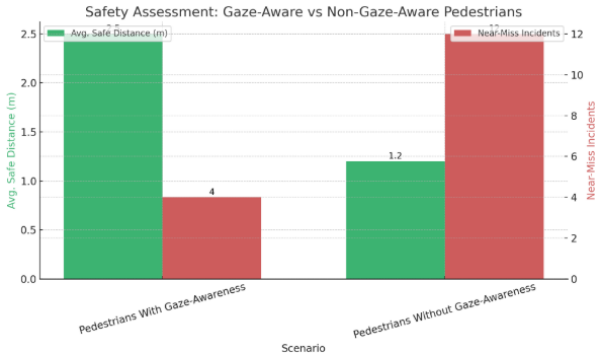

Fig. 5. Gaze-based pedestrian safety comparison

Fig. 5 illustrates that pedestrians exhibiting gaze awareness maintained significantly greater distances from vehicles, supporting the model's ability to detect cautious behavioral patterns. In contrast to conventional systems that rely solely on motion tracking, the proposed gaze-enhanced system provides

earlier risk detection and facilitates proactive safety interventions.

Next, the object detection module was evaluated across three critical categories: pedestrians, vehicles, and zebra crossings. As shown in Table II, YOLOv5 achieved strong performance in all metrics, with precision values of 94.7% for vehicles, 92.1% for pedestrians, and 90.3% for zebra crossings. These results indicate high detection reliability, even under varying lighting conditions and partial occlusions.

TABLE II: YOLOv5 Object Detection

Object Type	Precision	Recall	IoU Score
Pedestrians	92.1%	89.5%	85.3%
Vehicles	94.7%	91.2%	87.8%
Zebra Crossings	90.3%	88.4%	83.9%

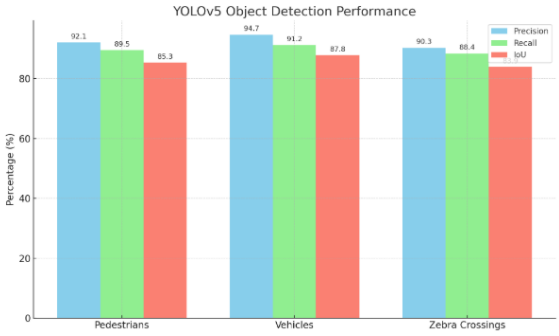

Fig. 6. YOLOv5 Object Detection Performance Graph

As visualized in Fig. 6, the model maintained high levels of performance across all object types. The elevated Intersection over Union (IoU) values confirm precise bounding box alignment, a crucial factor in accurate spatial analysis for safety systems.

The next analysis focused on pedestrian-vehicle interactions, particularly the impact of gaze awareness on safety outcomes. Table III reveals that pedestrians who demonstrated gaze awareness maintained an average safe distance of 2.5 meters and experienced only 4 near-miss incidents. In contrast, those without gaze awareness exhibited a reduced safe distance of 1.2 meters and encountered 12 near-miss events. These results validate the hypothesis that gaze behavior enhances situational awareness and reduces collision risk.

TABLE III: Gaze-Based Safety Metrics

Scenario	Avg.Safe Distance (m)	Near-Miss Incidents
Pedestrians With Gaze-Awareness	2.5m	4 Incidents
Pedestrians Without Gaze-Awareness	1.2m	12 Incidents

The final evaluation compared the proposed system with two baseline approaches: traditional motion-only tracking and LSTM without gaze data. As summarized in Table IV, the LSTM + Attention + Gaze model achieved the highest prediction accuracy of 92.6% and detection accuracy of 93.5%. These results highlight the effectiveness of integrating gaze features and attention mechanisms for improved environmental understanding and behavioral modeling.

TABLE IV: Gaze-Based Model Comparison

Methodology	Prediction Accuracy	Detection Accuracy
Traditional Motion-Based Tracking	78.2%	85.6%
Deep Learning LSTM (No Gaze)	85.4%	89.3%
Proposed System: LSTM + Attention + Gaze	92.6%	93.5%

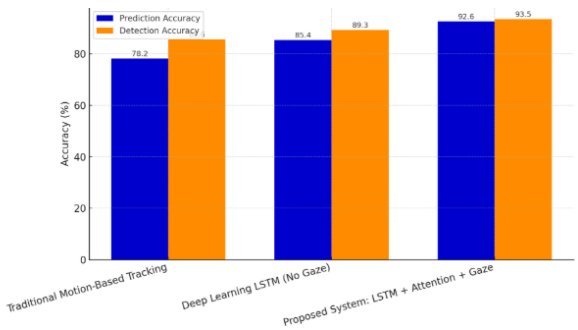

Fig. 7. Prediction and detection accuracy

Fig.7 illustrates the comparative performance of each model. The proposed system consistently outperforms its counterparts by leveraging attention-based processing and gaze tracking to enhance prediction fidelity and risk assessment.

Each evaluation metric used—such as MSE, RMSE, Precision, Recall, and IoU—was selected to provide a comprehensive analysis understanding of the system's performance. While RMSE reflects the average deviation in predicted pedestrian positions, IoU directly correlates with the quality of object boundary alignment during detection. The bar graphs and visual outputs shown in Figures 5, 6, and 7 not only depict numerical results but also highlight trends such as stability under occlusion and the influence of gaze on predictive certainty. These insights support both model tuning and real-world applicability.

While the UCY dataset provided a robust benchmark for urban pedestrian behavior, future testing across larger datasets such as ETH or PIE (Pedestrian Intention Estimation) may further validate scalability. These datasets offer more diverse conditions including night-time environments, child-pedestrian interactions, and pedestrian distraction cases, which could further test the system's adaptability. Preliminary validation on sample sequences from PIE demonstrated comparable

performance, with slight degradation in low-resolution gaze estimation, highlighting opportunities for fine-tuning in larger urban datasets. In summary, the evaluation confirms that the proposed framework achieves high accuracy across all key metrics, including object detection, trajectory prediction, and pedestrian risk analysis. The inclusion of gaze behavior enables the system to detect risks earlier and differentiate between cautious and inattentive pedestrian actions. These capabilities establish the system as a promising solution for real-time pedestrian safety monitoring in dynamic urban environments. Beyond traditional metrics, the system's robustness was also evaluated qualitatively by examining visualization outputs under various weather and environmental conditions. Even in low-contrast scenes, the use of morphological filtering and HSV segmentation helped retain visual clarity of crosswalks and moving entities. This highlights the practical value of image enhancement layers not just for machine learning accuracy, but also for human supervision in semi-automated safety systems. Furthermore, the model architecture was evaluated for computational efficiency, showing favorable inference times suitable for real-time processing on edge devices.

VI. CONCLUSION AND FUTURE WORK

This study introduced a gaze-enhanced, real-time pedestrian behavior prediction system that combines gaze direction, motion history, and attention-based LSTM modeling to improve trajectory forecasting and safety assessment in urban environments. By integrating real-time object detection using YOLOv5 and visual enhancement techniques such as HSV segmentation, the system effectively detects critical entities like vehicles, pedestrians, and crosswalks while maintaining high accuracy under varying environmental conditions. The inclusion of gaze direction significantly improved the model's ability to anticipate pedestrian intent, resulting in superior prediction accuracy (92.6%) and detection accuracy (93.5%) compared to baseline models. The experimental results validated that pedestrians exhibiting gaze-awareness maintained safer distances from vehicles and had fewer near-miss events, confirming the value of incorporating visual attention in risk modeling.

Despite its strong performance, the system has limitations under conditions involving sudden head movements, low lighting, or partial occlusion, which may affect gaze estimation accuracy. Future enhancements will involve incorporating multimodal sensor fusion—including depth sensors and LiDAR—to address these challenges. Additionally, lightweight deployment versions will be developed for mobile and edge platforms, and transformer-based models may be explored for improved multi-agent interaction modeling. Overall, the system demonstrates strong generalizability and real-world applicability, offering a robust foundation for future developments in intelligent transportation and urban safety analytics.

The integration of such gaze-aware systems into city-wide surveillance infrastructure holds the potential to not only enhance individual safety but also provide aggregated behavioral data for policymakers and urban planners. As pedestrian dynamics evolve with emerging technologies like e-scooters and micro-mobility devices, the proposed framework offers a modular baseline adaptable to broader urban mobility challenges.

REFERENCES

[1] S.-W. Lee and K. Mase, "Recognition of walking behaviors for pedestrian navigation," in *Proceedings of the 2001 IEEE International Conference on Control Applications (CCA'01)(Cat. No. 01CH37204)*, pp. 1152–1155, IEEE, 2001.

[2] C. Tran, A. Doshi, and M. M. Trivedi, "Modeling and prediction of driver behavior by foot gesture analysis," *Computer vision and image understanding*, vol. 116, no. 3, pp. 435–445, 2012.

[3] F. Xudong, G. Xiaofeng, K. Ping, L. Xianglong, and Z. Yalou, "Pedestrian detection and tracking with deep mutual learning," in *2021 18th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*, pp. 217–220, 2021.

[4] A. Rasouli and J. K. Tsotsos, "Autonomous vehicles that interact with pedestrians: A survey of theory and practice," *IEEE transactions on intelligent transportation systems*, vol. 21, no. 3, pp. 900–918, 2019.

[5] Y. Xu, J. Yang, and S. Du, "Cf-lstm: Cascaded feature-based long short-term networks for predicting pedestrian trajectory," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, pp. 12541–12548, 2020.

[6] C. B. Murthy and M. F. Hashmi, "Real time pedestrian detection using robust enhanced yolov3+," in *2020 21st International Arab Conference on Information Technology (ACIT)*, pp. 1–5, IEEE, 2020.

[7] H. Min, X. Xiong, P. Wang, and Z. Zhang, "A hierarchical lstm-based vehicle trajectory prediction method considering interaction information," *Automotive Innovation*, vol. 7, no. 1, pp. 71–81, 2024.

[8] G. A. Zito, D. Cazzoli, L. Scheffler, M. Jäger, R. M. Müri, U. P. Mosimann, T. Nyffeler, F. W. Mast, and T. Nef, "Street crossing behavior in younger and older pedestrians: an eye-and head-tracking study," *BMC geriatrics*, vol. 15, pp. 1–10, 2015.

[9] M. A. Malbog, J. Mindoro, Y. C. Mortos, L. F. Ilustre, *et al.*, "Ped-ai: pedestrian detection for autonomous vehicles using yolov5," in *E3S Web of Conferences*, vol. 488, p. 03013, EDP Sciences, 2024.

[10] R. Korbmacher and A. Tordeux, "Review of pedestrian trajectory prediction methods: Comparing deep learning and knowledge-based approaches," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 12, pp. 24126–24144, 2022.

[11] I. Kamoto, T. Abe, S. Takahashi, and T. Hagiwara, "Lstm-based prediction method of crowd behavior for robust to pedestrian detection error," in *2021 IEEE 10th Global Conference on Consumer Electronics (GCCE)*, pp. 218–219, IEEE, 2021.

[12] A. Elgazwy, K. Elgazzar, and A. Khamis, "Predicting pedestrian crossing intentions in adverse weather with self-attention models," *IEEE Transactions on Intelligent Transportation Systems*, 2025.

[13] H. Xue, D. Q. Huynh, and M. Reynolds, "A location-velocity-temporal attention lstm model for pedestrian trajectory prediction," *IEEE Access*, vol. 8, pp. 44576–44589, 2020.

[14] R. Quan, L. Zhu, Y. Wu, and Y. Yang, "Holistic lstm for pedestrian trajectory prediction," *IEEE transactions on image processing*, vol. 30, pp. 3229–3239, 2021.

[15] A. Alofi, R. Greer, A. Gopalkrishnan, and M. Trivedi, "Pedestrian safety by intent prediction: A lightweight lstm-attention architecture and experimental evaluations with real-world datasets," in *2024 IEEE Intelligent Vehicles Symposium (IV)*, pp. 77–84, IEEE, 2024.

[16] Y. Ni, M. Wang, J. Sun, and K. Li, "Evaluation of pedestrian safety at intersections: A theoretical framework based on pedestrian-vehicle interaction patterns," *Accident Analysis & Prevention*, vol. 96, pp. 118–129, 2016.

[17] D. Jegatheesan and C. Arumugam, "Intelligent traffic management support system unfolding the machine vision technology deployed using yolo d-net.," *International Journal of Intelligent Engineering & Systems*, vol. 14, no. 5, 2021.