An Enhanced Time-Attention-Based Social GAN with Traffic Light Semantic Integration for Real-Time Pedestrian Risk Alert on Edge Devices

Jing-Tai Chang*
Department of Information Engineer and Computer Science
Feng Chia University
Taichung, Taiwan
m1307396@o365.fcu.edu.tw

Feng-Cheng Lin
Department of Information Engineer and Computer Science
Feng Chia University
Taichung, Taiwan
fclin@fcu.edu.tw

***Abstract*—This study introduces an enhanced Social GAN augmented by a temporal attention mechanism and integrated with traffic light semantic vectors for real-time pedestrian trajectory prediction. Deployed on the NVIDIA Jetson AGX Xavier, the system leverages YOLOv10 and ByteTrack to facilitate low-latency inference (15 FPS via TensorRT). By encoding real-time signal phases into the model's latent space, the proposed TA-SGAN effectively captures environmental constraints at intersections. Experimental results and field tests on public buses in Taichung City demonstrate a 20% reduction in ADE/FDE compared to traditional models, confirming its practical feasibility for proactive urban risk-alert services.**

Keywords—Pedestrian trajectory prediction, Temporal Attention, Edge computing, Safety warning system

I. INTRODUCTION

Rapid urbanization demands real-time pedestrian trajectory prediction to enhance traffic safety and reduce accident risks. This study introduces a proactive risk warning system that integrates YOLOv10-based detection, ByteTrack tracking, and an enhanced Temporal-Attention Social GAN (TA-SGAN). A key innovation is the integration of traffic light semantic vectors; by encoding real-time signal phases (Red, Yellow, Green) into the model's latent space, the system dynamically captures environmental constraints—such as deceleration at intersections—improving adaptability in rule-governed urban settings.Powered by the NVIDIA Jetson AGX Xavier, this edge-computing approach enables low-latency inference independent of cloud services. Experimental results demonstrate significant reductions in Average Displacement Error (ADE) and Final Displacement Error (FDE) compared to Social LSTM and Vanilla GAN. With an optimized inference speed of 15 FPS via TensorRT, the system confirms its real-time feasibility and robust performance through field tests on public buses in Taichung City, providing a practical solution for intelligent transportation safety.

II. METHODS

A. Trajectory Pedestrian on Dataset

To develop and evaluate pedestrian movement forecasting in diverse real-world environments, this study utilizes five trajectory datasets (Fig. 1): ETH, Hotel, Univ, Zara1, and Zara2. These datasets provide sequential x, y coordinate annotations, enabling the model to learn realistic motion patterns influenced by social interactions and spatial constraints.

- **ETH** and **Hotel**: Feature structured, low-density urban settings where pedestrian movements follow more predictable paths with minimal interactions.
- **Univ**: Represents a high-density university campus with diverse movement patterns and multi-directional pedestrian flows, requiring an adaptive forecasting approach.
- **Zara1** and **Zara2**: Capture dynamic pedestrian behaviors in commercial districts, where movement is less structured and influenced by unpredictable human interactions.

ETH-dataset Hotel-dataset Zara1/Zara2-dataset Univ-dataset

Fig. 1. Pedestrian Trajectory Prediction Datasets

B. YOLOv10 and ByteTrack for Pedestrian Detection and Tracking

YOLOv10 represents the latest advancement in the You Only Look Once (YOLO) series, introducing an enhanced backbone network, adaptive anchor-free detection, and an optimized object detection pipeline. These improvements significantly boost both detection speed and accuracy, making YOLOv10 highly suitable for real-time pedestrian detection in dynamic environments.

To improve pedestrian tracking stability, the ByteTrack multi-object tracking (MOT) algorithm utilizes both highand low-confidence detections. This approach effectively mitigates challenges such as occlusions and motion blur, leading to robust and consistent tracking over time. When integrated with YOLOv10, ByteTrack delivers precise pedestrian localization and continuous trajectories, establishing a solid foundation for real-time trajectory prediction and risk assessment in urban environments.

C. Temporal-Attention Enhanced Trajectory Prediction with Semantic Integration

To improve the accuracy of pedestrian trajectory forecasting, this study incorporates a time-attention mechanism integrated with environmental semantics into the trajectory prediction model. The detailed network architecture is illustrated in Fig. 2. Our proposed generator adopts an Encoder-Pooling-Decoder structure based on LSTMs, where the Temporal Attention module dynamically weights past motion sequences at both the encoding and decoding stages, allowing the model to focus on critical temporal dependencies in pedestrian movement.

Furthermore, we specifically introduce Traffic Light Semantic Vectors as environmental constraints. As shown in Fig.3, a specialised YOLOv10s-based detector is employed to identify signal states in real-time.

To quantitatively integrate these constraints, the contextual representation Ct is modulated by the traffic light weight $W_{tsignal}$ as follows:

$$Ct = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \bullet V \times W_{tsignal} \quad (1)$$

where $W_{tsignal}$ is a learnable scaling factor derived from the discrete signal states (Red, Green, Yellow). This allows the attention mechanism to prioritize or suppress specific trajectory candidates based on the current traffic rules. By integrating Social Attention and this semantic embedding, the model effectively captures both spatial interactions and environmental context, resulting in more precise trajectory predictions suited for real-time safety-critical applications on edge devices.

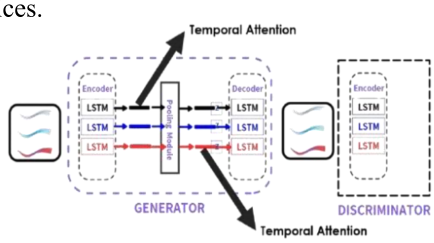

Fig. 2. Enhancing Trajectory Prediction with Temporal Attention

Fig. 3. Real-time Traffic Light Detection using YOLOv10sn

D. Temporal-Attention Enhanced Trajectory Prediction

The proposed system is deployed on the NVIDIA Jetson AGX Xavier to enable real-time, low-latency inference independent of cloud services. As illustrated in Fig. 4, the pipeline integrates YOLOv10 and ByteTrack for robust detection and tracking, followed by the TA-SGAN for trajectory forecasting. By leveraging GPU acceleration and TensorRT optimization, the architecture effectively captures spatiotemporal dependencies, ensuring responsive and proactive decision-making in dynamic urban environments.

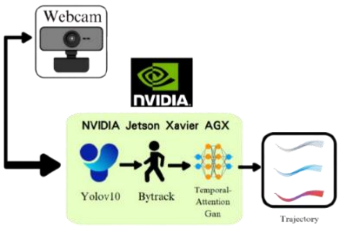

Fig. 4. Integrated Edge Deployment Pipeline on Jetson AGX Xavier

III. Experiment Result

A. Performance of Deep Learning Models

The proposed model was evaluated across five datasets (ETH, Hotel, Univ, Zara1, and Zara2) representing diverse urban complexities. As shown in Table 1, our TA-SGAN significantly reduces ADE and FDE compared to S-LSTM , S-GAN , and STGCNN. This performance gain highlights the model's superior efficiency in capturing complex spatiotemporal patterns and its robustness across varied pedestrian densities and social interactions.

Table 1. Evaluation of Model Performance

	My-Design Model	**S-LSTM Model** [1]	**S-GAN Model** [2]	**STGCNN Model** [3]
ETH	0.42/0.91	1.09/2.35	0.87/1.62	0.64/1.11
Hotel	0.41/0.65	0.79/1.76	0.67/1.37	0.49/0.85
Univ	0.34/0.63	0.67/1.40	0.76/1.52	0.44/0.79
Zara1	0.29/0.39	0.47/1.00	0.35/0.68	0.34/0.53
Zara2	0.31/0.39	0.56/1.17	0.42/0.84	0.30/0.48

B. Inference Performance and Real-Time Feasibility on Edge Devices

To evaluate the model's real-time applicability, we compared inference performance across multiple deep learning models using both desktop and embedded platforms. While baseline models such as Social-LSTM [1], Social-GAN [2], and STGCNN [3] were tested on a high-performance GPU (RTX 4080), the proposed TA-SGAN model was deployed on an embedded Jetson AGX Xavier module optimized with TensorRT for low-latency inference.

Table2. **Evaluation of Model Performance**

Model	**My-Design Model**	**S-LSTM Model [1]**	**S-GAN Model [2]**	**STGCNN Model [3]**
Inference Time (ms)	85	210	185	150
FPS	11.8	4.7	5.4	6.2
Platform	Jetson AGX Xavier (TensorRT)	GPU (RTX4080)	GPU (RTX4080)	GPU (RTX4080)

IV. CONCLUSION

A pedestrian safety enhancement system was developed, integrating deep learning trajectory prediction with edge computing. The core is a Time-Attention-based Social GAN deployed on the NVIDIA Jetson AGX Xavier, enabling efficient real-time detection, tracking, and prediction. Evaluation on public datasets showed a 20% reduction in ADE and FDE. Real-world tests on public buses in Taichung City achieved an average latency of 66 ms (15 FPS), confirming the system's feasibility for proactive risk-alert services in smart city environments.

REFERENCE

[1] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social LSTM: Human Trajectory Prediction in Crowded Spaces," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[2] A. Gupta, J. Johnson, F.-F. Li, S. Savarese, and A. Alahi, "Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[3] A. Mohamed, K. Qian, M. Elhoseiny, and C. Claudel, "Social- STGCNN: A Social Spatio-Temporal Graph Convolutional Neural Network for Human Trajectory Prediction," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.