Proactive Prediction of Child Pedestrian Trajectories Using Trajectory Unified Transformer (TUTR) Models

Sungmin Yoo*[1], Hanuel Park[2], Chiwoo Roh[2], Sungeun Cho[2], Jaehyun(Jason) So[2]

[1] Master's Student, Department of Data/Network/AI Convergence, Ajou University, Republic of Korea

[2] Ph.D. Student, Ajou University, Republic of Korea

[2] Master's Student, Ajou University, Republic of Korea

[2] Undergraduate Student, Ajou University, Republic of Korea

[2] Associate Professor, Ajou University, Republic of Korea

SHORT SUMMARY

Enhancing child pedestrian safety is crucial due to their unpredictable behaviors, which pose significant risks. Existing post-accident measures, such as traffic regulations and infrastructure improvements, face limitations in real-time applicability and cost-efficiency. This study proposes a deep learning approach using the Trajectory Unified Transformer (TUTR) model to predict the future trajectories of child pedestrians, enabling proactive safety interventions. The AI-HUB dataset, featuring hazardous child behaviors like sudden road appearances and abrupt direction changes, was used for training. Methodologies include video data preprocessing, trajectory extraction, and data augmentation to improve model performance. Evaluation metrics such as Average Displacement Error (ADE) and Final Displacement Error (FDE) demonstrate high prediction accuracy, with a 0.24 difference between ADE and FDE values. This research highlights the potential of integrating TUTR models into autonomous vehicles and infrastructure, providing real-time safety measures and reducing child pedestrian accidents.

Keywords: Child-Pedestrian, trajectory, Transformer, Forecasting, Deep Learning

1. INTRODUCTION

To improve the traffic safety of child pedestrians, numerous studies have explored both proactive approaches (Moradi-Pari et al., 2022; Shen et al., 2018) that address hazardous situations in real-time, and post-accident management (Gitelman et al., 2019; Morrongiello & Barton, 2009; Schwebel et al., 2012) focusing on analyzing accident severity (Khan et al., 2024), accident counts (Khanum et al., 2023), and implementing safety facilities like fences. Reactive measures, such as improving crosswalk visibility, implementing traffic calming strategies, and offering traffic safety education, face challenges in real-time application. They incur high maintenance costs and must align with existing road infrastructure, reducing their effectiveness (Namatovu et al., 2022). These limitations have led researchers to focus on proactive systems leveraging advanced technologies. One example is the Smart Walk Assistant (SWA) (Khosravi et al., 2018), which integrates a smartphone application with roadside units (RSUs) to provide real-time intersection information. However, despite its benefits, this system lacks predictive capabilities to anticipate hazardous situations like a child unexpectedly running into the road (Saleh, 2022; Schwebel et al., 2012). Addressing this gap requires predictive models capable of anticipating child pedestrians' trajectories in pre-accident scenarios.

Recent advancements in deep learning have enabled the development of trajectory prediction models for pedestrians. Early studies using RNNs like LSTM and GRU achieved moderate success (Alahi et al., n.d.; Liu et al., 2021) but faced limitations, including vanishing gradients, slow processing, and difficulty capturing complex dependencies (Sato et al., 2023; Shi et al., 2021). Transformer-based models, such as the Trajectory Unified Transformer (TUTR), have emerged as superior alternatives, offering improved accuracy and efficiency through multi-attention mechanisms (Shi et al., n.d.; Vaswani et al., 2017).

TUTR distinguishes itself by eliminating post-processing, enhancing real-time applicability (Xu et al., 2022; Yuan et al., 2021). It employs a mode-level encoder and social-level decoder to predict motion patterns and interactions among pedestrians, making it particularly suitable for addressing unpredictable child pedestrian behaviors (Juzdani et al., 2021). Despite its strengths, current research largely focuses on general pedestrian trajectories, leaving a gap in predicting child pedestrian behaviors (Swedler et al., 2024). Children often exhibit more erratic patterns, making them especially vulnerable in traffic environments (Kendi & Johnston, 2023).

Integrating TUTR with autonomous vehicle systems can provide real-time insights, improving driver stability and reducing reaction times during critical situations (Classen et al., 2023; Saleh, 2022). This study aims to bridge the gap by employing TUTR to predict child pedestrian trajectories in real-time. By adapting TUTR's capabilities to unique child behavioral patterns, this research seeks to enhance safety and reduce risks. Furthermore, integrating TUTR with autonomous vehicle path planning and collision avoidance systems offers a promising approach to improving traffic safety (Classen et al., 2023).

Trajectory prediction can be categorized into vehicle and pedestrian prediction. Vehicle models like GRU and LSTM predict paths with high accuracy, enhancing overall safety (Ip et al., 2021; Liu et al., 2021; Messaoud et al., 2020). For pedestrians, methodologies have evolved from physics-based models (Helbing & Molnar, 1995) to deep learning approaches. While physics-based models simulate movement through interactions (Karamouzas et al., 2014), their high computational costs and limited accuracy drove the adoption of data-driven models like LSTM, GAN, and SGCN (Fang et al., 2022; Sadeghian et al., 2018; Shi et al., 2021). These models offer improved efficiency but struggle with long-term dependencies and transparency issues (Shi et al., 2021).

Transformer models (Vaswani et al., 2017) address these limitations by enabling parallel data processing. TUTR builds on this foundation with innovative features, including Global Prediction, Mode-Level Transformer Encoder, and Social-Level Transformer Decoder, which normalize trajectories and cluster motion modes. This dual-stage approach improves accuracy and reliability (Juzdani et al., 2021).

Although TUTR and similar models excel at general pedestrian trajectory prediction, datasets like ETH, UCY, and SSD lack subgroup-specific distinctions (Lerner et al., 2007; Pellegrini et al., n.d.). This study focuses on predicting child pedestrian trajectories, addressing a critical gap in transportation safety. By analyzing and predicting children's risk behaviors in real-time, the research aims to prevent accidents and improve outcomes. Leveraging TUTR's capabilities provides a robust framework for addressing the unique challenges posed by child pedestrians, contributing valuable insights to traffic safety. The model's structure is illustrated in Figure 1.

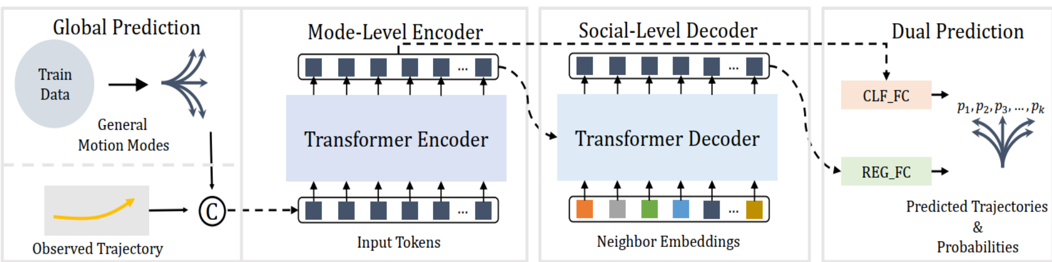

Figure 1TUTR Architecture (Shi et al., n.d.)

2. METHODOLOGY

Data Collection

This study utilized the AI-HUB dataset "Child Pedestrian Road Risk Behavior Videos in School Zones," a comprehensive resource designed to train artificial intelligence systems to recognize hazardous behaviors in school zones. The dataset includes 1,124 CCTV and 443 black box recordings, categorized into ten types of hazardous behaviors such as jaywalking, sudden appearances, and fighting. For this study, 519 videos focusing on six specific behaviors—driveway walking, fighting, jaywalking, using umbrellas, sudden appearances, and walking with dogs—were selected. Each video contains 300 image frames annotated in Pascal VOC format, providing detailed object class information and bounding box coordinates. The dataset was obtained from AI-HUB (2020).

Data Annotation and Preprocessing

Since the Pascal VOC format does not support pedestrian ID tracking, the DarkLabel tool was used for preprocessing, enabling the extraction of pedestrian coordinates for each frame. This preprocessing step allowed the conversion of data into a format suitable for trajectory prediction models, similar to ETH and UCY datasets. A total of 1,351 child pedestrian trajectories were extracted, and data augmentation techniques such as random rotation, axis inversion, and reflection were applied to expand the dataset, resulting in 4,053 augmented trajectories.

Model Training and Testing

The model implementation was based on the open-source GitHub repository, TUTR: Trajectory Unified Transformer for Pedestrian Trajectory Prediction (https://github.com/lssiair/TUTR). This repository serves as the official implementation of TUTR, providing a comprehensive framework for pedestrian trajectory prediction.
The Trajectory Unified Transformer (TUTR) model was trained and tested using the leave-one-out method, where the dataset was divided into five subsets: four for training and one for testing. Key model parameters included an observation radius of 2, 8 observation frames, 12 prediction frames, a hidden dimension size of 64, and 100 clusters. Training parameters were set with a learning rate of 0.001, a batch size of 71, a total of 31 epochs, and 30 batches per epoch. During the testing phase, 30 samples were generated to ensure optimal prediction scenarios. For evaluation, a world scale value of 1 was applied to align predictions with real-world settings.

3. RESULTS AND DISCUSSION

To evaluate the difference between actual and predicted trajectories for each hazardous behavior type and assess prediction reliability, metrics such as Average Displacement Error (ADE), Final Displacement Error (FDE), Brier-ADE, and Brier-FDE (Wang et al., 2022) were used. ADE calculates the average distance between predicted and actual trajectories over time (Equation 1):

$$\mathrm{ADE} = \frac{1}{\mathrm{T_{pred}}} \sum_{\mathrm{t=T_{obs}}}^{\mathrm{T}} \sqrt{(\mathrm{x_t} - \hat{\mathrm{x}}_\mathrm{t})^2 - (\mathrm{y_t} - \hat{\mathrm{y}}_\mathrm{t})^2} \tag{1}$$

FDE measures the distance between the final predicted and actual points (Equation 2):

$$\mathrm{FDE} = \sqrt{(x_T - \hat{x}_T)^2 - (y_T - \hat{y}_T)^2} \quad (2)$$

Brier-ADE and Brier-FDE extend these metrics by incorporating prediction probability , as defined in Equations 3 and 4:

$$\text{Brier-Ade} = \mathrm{ADE} + (1 - p)^2 \quad (3)$$

$$\text{Brier-Fde} = \mathrm{FDE} + (1 - p)^2 \quad (4)$$

Among the six behavior types, "suddenly_appear" showed the largest discrepancy between ADE/FDE and Brier-ADE/Brier-FDE, at 0.6, indicating lower prediction probability and higher difficulty (Table 1).

Table 1: Evaluation Metrics

Types of risky behaviors	**driveway_walk**	**fighting**	**jay_walk**	**putup_ umbrella**	**Suddenly appear**	**with_dog**	**AVG**
Evaluation metrics							
ADE	3.53	3.60	3.56	3.60	3.61	3.54	3.57
FDE	6.39	6.50	6.42	6.50	6.52	6.41	6.45
Brier-ADE	3.57	3.85	3.57	3.65	4.21	3.94	3.8
Brier FDE	6.43	6,85	6.43	6,55	7.12	6.81	6.7

The model's loss decreased from 5.62 to 3.75 during training, demonstrating successful convergence (Figure 2).

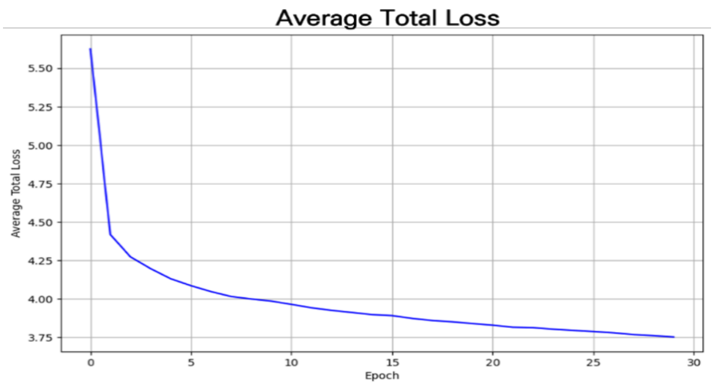

Figure 2:Average Total Loss

Visualization results demonstrated that the predicted trajectories effectively captured various behavioral patterns exhibited by child pedestrians (Figure 3).

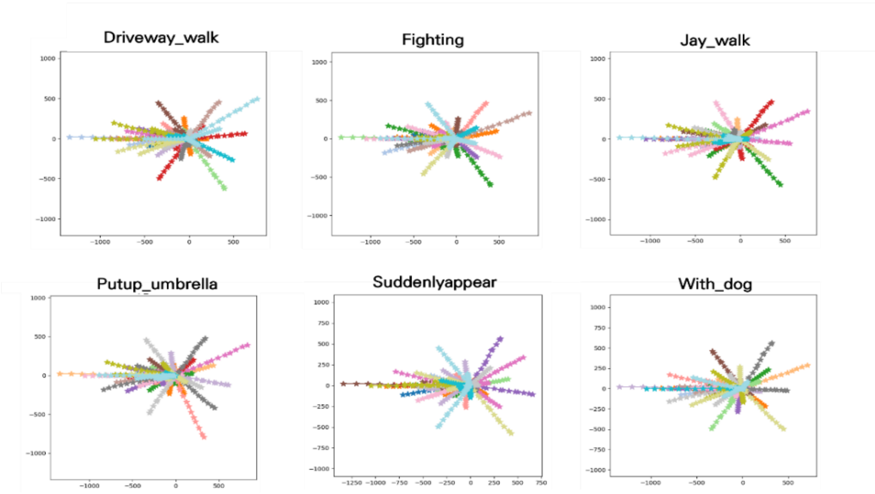

Figure 3:General Motion Mode

For instance, in the "suddenly_appear" scenario, the model successfully reflected the complex actions of a child running onto the road, engaging in hazardous behavior such as jaywalking, and returning to the sidewalk. While the predicted trajectories generally aligned with the observed paths, slight deviations were noted due to the unpredictable and erratic nature of this behavior (Figure 4).

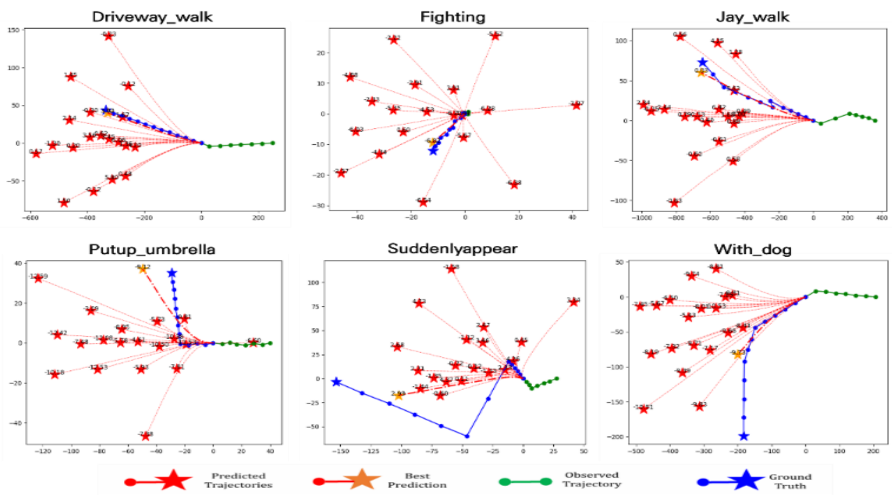

Figure 4:Predicted Trajectories and Corresponding Probabilities

In Figure 3, the general motion modes are visualized, showing the normalized trajectories of child pedestrians moving from right to left. This provides an intuitive overview of how the model processes and predicts standard movement patterns. Figure 4 highlights individual trajectories with detailed visual representation: the observed trajectory in green, the ground truth in blue, and the predicted trajectory in red. For trajectories with the highest accuracy, the final position is marked with an orange star, emphasizing the model's capability to predict future positions with reasonable precision despite the variability in behavior types.

The model achieved an average ADE of 3.57 and FDE of 6.45, with Brier-ADE and Brier-FDE averaging 3.8 and 6.7. Behaviors like "driveway_walk" and "putup_umbrella" demonstrated high

reliability due to consistent patterns. Conversely, "suddenly_appear" was more challenging to predict due to its erratic nature, evidenced by a higher discrepancy in Brier metrics.
These results highlight TUTR's suitability for predicting child pedestrian behaviors, particularly in real-time applications. However, reliance on reenacted data limits generalizability. Future research should utilize real-world datasets and integrate advanced sensors, such as radar, to improve accuracy and reliability, ultimately enhancing autonomous system safety in school zones.

4. CONCLUSIONS

This study aimed to improve traffic safety by predicting the trajectories of child pedestrians using the AI-Hub dataset "Child Pedestrian Road Risk Behavior Videos in School Zones" and the TUTR model. Unlike prior studies focusing on adult pedestrian behaviors or post-accident measures, this research addressed the unique and unpredictable movement patterns of children, emphasizing real-time risk prediction. By analyzing hazardous behaviors such as jaywalking and sudden appearances, the study demonstrated the potential to reduce accidents involving child pedestrians. The TUTR model achieved strong predictive reliability, with overall ADE/FDE values of 3.57/6.45 and Brier-ADE/Brier-FDE values of 3.8/6.7, reflecting minimal discrepancies and high accuracy. For behaviors like driveway_walk and jay_walk, the differences between ADE and Brier-ADE metrics were as low as 0.04 and 0.01, respectively, further confirming the model's effectiveness. Despite these promising results, the reliance on reenacted video data presents limitations. Future research should employ real-world datasets to capture more dynamic behaviors and enhance model applicability. Integrating such predictive systems with autonomous vehicle path planning and collision avoidance could significantly enhance traffic safety, ensuring both the safety of child pedestrians and improved driver response in critical situations.

ACKNOWLEDGEMENTS

This work was supported by the Korea Institute of Police Technology (KIPoT) grant funded by the Korea government (KNPA) (No. RS-2024-00403630, Development of autonomous driving patrol service for active prevention and response to traffic accidents).

REFERENCES

Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., & Savarese, S. (n.d.). *Social LSTM: Human Trajectory Prediction in Crowded Spaces*.

Classen, S., Sisiopiku, V. P., Mason, J. R., Yang, W., Hwangbo, S. W., McKinney, B., & Li, Y. (2023). Experience of drivers of all age groups in accepting autonomous vehicle technology. *Journal of Intelligent Transportation Systems: Technology, Planning, and Operations*. https://doi.org/10.1080/15472450.2023.2197115

Fang, F., Zhang, P., Zhou, B., Qian, K., & Gan, Y. (2022). Atten-GAN: Pedestrian Trajectory Prediction with GAN Based on Attention Mechanism. *Cognitive Computation*, *14*(6), 2296–2305. https://doi.org/10.1007/s12559-022-10029-z

Gitelman, V., Levi, S., Carmel, R., Korchatov, A., & Hakkert, S. (2019). Exploring patterns of child pedestrian behaviors at urban intersections. *Accident Analysis and Prevention*, *122*, 36–47. https://doi.org/10.1016/j.aap.2018.09.031

Helbing, D., & Molnar, P. (1995). *Social force model for pedestrian dynamics* (Vol. 51, Issue 5).

Ip, A., Irio, L., & Oliveira, R. (2021). Vehicle Trajectory Prediction based on LSTM Recurrent Neural Networks. *IEEE Vehicular Technology Conference*, *2021-April*. https://doi.org/10.1109/VTC2021-Spring51267.2021.9449038

Juzdani, M. H., Morgan, C. H., Schwebel, D. C., & Tabibi, Z. (2021). Children's road-crossing behavior: Emotional decision making and emotion-based temperamental fear and anger. *Journal of Pediatric Psychology*, *45*(10), 1188–1198. https://doi.org/10.1093/JPEPSY/JSAA076

Karamouzas, I., Skinner, B., & Guy, S. J. (2014). *Universal power law governing pedestrian interactions*. https://doi.org/10.1103/PhysRevLett.113.238701

Kendi, S., & Johnston, B. D. (2023). Epidemiology and Prevention of Child Pedestrian Injury. *Pediatrics*, *152*(1). https://doi.org/10.1542/peds.2023-062508

Khan, M. N., Das, S., & Liu, J. (2024). Predicting pedestrian-involved crash severity using inception-v3 deep learning model. *Accident Analysis and Prevention*, *197*. https://doi.org/10.1016/j.aap.2024.107457

Khanum, H., Garg, A., & Faheem, M. I. (2023). Accident severity prediction modeling for road safety using random forest algorithm: an analysis of Indian highways. *F1000Research*, *12*, 494. https://doi.org/10.12688/f1000research.133594.1

Khosravi, S., Beak, B., Head, K. L., & Saleem, F. (2018). Assistive system to improve pedestrians' safety and mobility in a connected vehicle technology environment. *Transportation Research Record*, *2672*(19), 145–156. https://doi.org/10.1177/0361198118783598

Lerner, A., Chrysanthou, Y., & Lischinski, D. (2007). Crowds by example. *Computer Graphics Forum*, *26*(3), 655–664. https://doi.org/10.1111/j.1467-8659.2007.01089.x

Liu, J., Mao, X., Fang, Y., Zhu, D., & Meng, M. Q.-H. (2021). *A Survey on Deep-Learning Approaches for Vehicle Trajectory Prediction in Autonomous Driving*. http://arxiv.org/abs/2110.10436

Messaoud, K., Deo, N., Trivedi, M. M., & Nashashibi, F. (2020). *Trajectory Prediction for Autonomous Driving based on Multi-Head Attention with Joint Agent-Map Representation*. http://arxiv.org/abs/2005.02545

Moradi-Pari, E., Tian, D., Mahjoub, H. N., & Bai, S. (2022). The Smart Intersection: A Solution to Early-Stage Vehicle-to-Everything Deployment. *IEEE Intelligent Transportation Systems Magazine*, *14*(5), 88–102. https://doi.org/10.1109/MITS.2021.3093241

Morrongiello, B. A., & Barton, B. K. (2009). Child pedestrian safety: Parental supervision, modeling behaviors, and beliefs about child pedestrian competence. *Accident Analysis and Prevention*, *41*(5), 1040–1046. https://doi.org/10.1016/j.aap.2009.06.017

Namatovu, S., Balugaba, B. E., Muni, K., Ningwa, A., Nsabagwa, L., Oporia, F., Kiconco, A., Kyamanywa, P., Mutto, M., Osuret, J., Rehfuess, E. A., Burns, J., & Kobusingye, O. (2022). Interventions to reduce pedestrian road traffic injuries: A systematic review of randomized controlled trials, cluster randomized controlled trials, interrupted time-series, and controlled before-after studies. *PLoS ONE*, *17*(1 1). https://doi.org/10.1371/journal.pone.0262681

Pellegrini, S., Ess, A., Schindler, K., & Van Gool, L. (n.d.). *You'll Never Walk Alone: Modeling Social Behavior for Multi-target Tracking*.

Sadeghian, A., Kosaraju, V., Sadeghian, A., Hirose, N., Rezatofighi, S. H., & Savarese, S. (2018). *SoPhie: An Attentive GAN for Predicting Paths Compliant to Social and Physical Constraints*. http://arxiv.org/abs/1806.01482

Saleh, K. (2022). Pedestrian Trajectory Prediction for Real-Time Autonomous Systems via Context-Augmented Transformer Networks. *Sensors*, *22*(19). https://doi.org/10.3390/s22197495

Sato, Y., Sasaki, Y., & Takemura, H. (2023). STP4: spatio temporal path planning based on pedestrian trajectory prediction in dense crowds. *PeerJ Computer Science*, *9*. https://doi.org/10.7717/PEERJ-CS.1641

Schwebel, D. C., Davis, A. L., & O'Neal, E. E. (2012). Child Pedestrian Injury: A Review of Behavioral Risks and Preventive Strategies. *American Journal of Lifestyle Medicine*, *6*(4), 292–302. https://doi.org/10.1177/0885066611404876

Shen, B., Zhang, Z., Liu, H., Li, S., & Zhao, L. (2018). Research on a Conflict Early Warning System Based on the Active Safety Concept. *Journal of Advanced Transportation*, *2018*. https://doi.org/10.1155/2018/8372108

Shi, L., Wang, L., Long, C., Zhou, S., Zhou, M., Niu, Z., & Hua, G. (2021). *SGCN:Sparse Graph Convolution Network for Pedestrian Trajectory Prediction*. http://arxiv.org/abs/2104.01528

Shi, L., Wang, L., Zhou, S., & Hua, G. (n.d.). *Trajectory Unified Transformer for Pedestrian Trajectory Prediction*.

Swedler, D. I., Ali, B., Hoffman, R., Leonardo, J., Romano, E., & Miller, T. R. (2024). Injury and fatality risks for child pedestrians and cyclists on public roads. *Injury Epidemiology*, *11*(1). https://doi.org/10.1186/s40621-024-00497-2

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). *Attention Is All You Need*. http://arxiv.org/abs/1706.03762

Xu, P., Hayet, J.-B., & Karamouzas, I. (2022). *SocialVAE: Human Trajectory Prediction using Timewise Latents*. https://doi.org/10.1007/978-3-031-19772-7_30

Yuan, Y., Weng, X., Ou, Y., & Kitani, K. (2021). *AgentFormer: Agent-Aware Transformers for Socio-Temporal Multi-Agent Forecasting*. http://arxiv.org/abs/2103.14023

Yang, H. (2020). Child Pedestrian Road Risk Behavior Videos in School Zones. AI-Hub. https://www.aihub.or.kr/aihubdata/data/view.do?currMenu=&topMenu=&aihub-DataSe=data&dataSetSn=169

Lssiair. (n.d.). *Trajectory Unified Transformer (TUTR) for pedestrian trajectory prediction* [Computer software]. GitHub. Retrieved January 16, 2025, from https://github.com/lssiair/TUTR