13th International Conference on Applied Informatics
Eszterházy Károly Catholic University
Eger, Hungary, February 19–21, 2026

IP-Transformer: Latent Intent Projection for Trajectory Prediction in Autonomous Vehicles

Emin Bayramov[a], **Zoltán Istenes**[b]

[a]Academic-Industrial Cooperation Center, Faculty of Informatics, Eötvös Loránd University, Budapest, Hungary
ORCID: 0000-0002-0428-8908

[b]Academic-Industrial Cooperation Center, Faculty of Informatics, Eötvös Loránd University, Budapest, Hungary
ORCID: 0000-0002-0169-4791

Abstract. The capability of ADAS systems to forecast the future states of surrounding agents is hinged. It requires a Theory of Mind, the ability to speculate on the unobservable intentions of nearby agents. The existing Transformer-based architectures predominantly rely on deterministic history, failing to account for the stochastic nature of human decision-making. To address this, the IP-Transformer (Intent-Projected Transformer) is introduced, a novel three-stage architecture that explicitly injects hallucinated social intent into the forecasting pipeline. This enables the Interaction Transformer to reason about "what might happen" (intent) rather than just "what happened" (history).

The framework is validated through a dual-domain evaluation strategy. First, quantitative benchmarking on the nuScenes mini dataset demonstrates high-precision forecasting with an Average Displacement Error (ADE) of 0.36m and Final Displacement Error (FDE) of 0.68m. Second, through a novel Adversarial Stress Test Protocol, which evaluates the model against safety-critical edge cases, including aggressive cut-ins, emergency braking, and occluded pop-outs. Results from these adversarial scenarios indicate that the IP-Transformer exhibits robust risk-aware intelligence.

Keywords: Autonomous Driving, Trajectory Forecasting, Transformer Networks, Latent Variable Models, Social Interaction Modeling, Multimodal Prediction, Safety-Critical Perception

1. Introduction

The modern urban intersection portrays one of the most complex dynamic systems in the physical world [5]. It is a theatre of disorganized cooperation where heavy machines, fragile pedestrians, and agile cyclists encounter in a high-stakes dance of negotiation [6]. For the experienced human driver, this navigation is effortless, moderated by a cognitive machinery matured over millennia to predict social outcomes [3]. Meanwhile for the Autonomous Vehicle (AV), nevertheless, this environment remains a profound computational challenge [4].

1.1. Contributions

In light of these phenomenological complexities, the IP-Transformer (Intent-Projected Transformer) is proposed. Unlike standard architectures that map observed trajectories directly to future states, the proposed approach explicitly models the *"ghost in the machine"* by hallucinating latent intent vectors before trajectory generation. The main contributions of this research work are summarized as follows: **Latent Intent Projection:** A novel three-stage forecasting pipeline allowing the model to reason about "what might happen" rather than just "what happened."

Social Projector Mechanism: A fusion mechanism is proposed that account for both kinematic constraints and aggressive/conservative behavioral profiles.

2. Background and related works

Evolution of Interaction Modeling: From Social-LSTM to Attention Early data-driven approaches utilized Recurrent Neural Networks (RNNs), specifically Long Short-Term Memory (LSTM) networks, to model the temporal evolution of pedestrian and vehicle states [1], [7]. While adequate for sparse environments where agent-to-agent influence is negligible, this *"independent"* paradigm collapsed in crowded spaces.

Social Pooling: The Introduction of Joint Modeling Social-LSTM proposed a *"Social Pooling"* layer, a differentiable operation inserted between the LSTM steps. Social-LSTM was significant both theoretically and practically because it formalized the idea that the social environment of an agent must be represented to forecast their future. However, the architecture suffered from inherent structural limitations.

3. Methodology

The methodology presented herein introduces the **IP-Transformer (Intent Projection Transformer)**, a novel deep learning framework designed to address the stochastic and interactive nature of multi-agent trajectory prediction in autonomous driving.

The IP-Transformer framework is formed as a multi-stage pipeline, precisely separating perception, reasoning, and decoding, while maintaining end-to-end differentiability. As illustrated in the architectural diagram Fig. 1

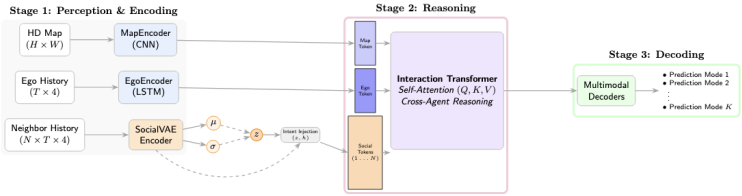

Figure 1. Architecture of the proposed IP-Transformer (Intent Projection Transformer)

4. Experimental setup and Results

Consequently, this experimental framework is bifurcated into two primary domains. First, **Standardized Benchmarking** utilizes the massive scale of the nuScenes dataset [2] and a novel **Adversarial Stress Test Protocol** is presented to evaluate the behavioral robustness of the model.

References

[1] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, S. Savarese: *Social lstm: Human trajectory prediction in crowded spaces*, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 961–971.

[2] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, O. Beijbom: *nuscenes: A multimodal dataset for autonomous driving*, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 11621–11631.

[3] B. Fruchart, B. Le Blanc: *Cognitive machinery and behaviours*, in: International Conference on Artificial General Intelligence, Springer, 2020, pp. 121–130.

[4] R. George, J. Clancy, T. Brophy, G. Sistu, W. O'Grady, S. Chandra, F. Collins, D. Mullins, E. Jones, B. Deegan, et al.: *Infrastructure Assisted Autonomous Driving: Research, Challenges, and Opportunities*, IEEE Open Journal of Vehicular Technology (2025).

[5] S. Gupta, M. Vasardani, S. Winter: *Negotiation between vehicles and pedestrians for the right of way at intersections*, IEEE Transactions on Intelligent Transportation Systems 20.3 (2018), pp. 888–899.

[6] A. Rasouli, J. K. Tsotsos: *Autonomous vehicles that interact with pedestrians: A survey of theory and practice*, IEEE transactions on intelligent transportation systems 21.3 (2019), pp. 900–918.

[7] P. Zhang, J. Xue, P. Zhang, N. Zheng, W. Ouyang: *Social-aware pedestrian trajectory prediction via states refinement LSTM*, IEEE transactions on pattern analysis and machine intelligence 44.5 (2020), pp. 2742–2759.