Supplementary Material for
ViTE: Virtual Graph Trajectory Expert Router for Pedestrian Trajectory Prediction

Ruochen Li[1], **Zhanxing Zhu**[2], **Tanqiu Qiao**[1], **Hubert P. H. Shum**[1*]

[1]Department of Computer Science, Durham University, UK
[2]School of Electrical and Computer Science (ECS), University of Southampton, UK

This supplementary document offers additional technical and experimental details that support the findings reported in the main paper. We begin with an overview of the baseline methods used for comparison. We then describe our implementation details. Finally, we present supplementary experiments and qualitative results to further illustrate the effectiveness of our proposed framework.

Details of Baselines

- **STAR** (Yu et al. 2020): This model introduces a spatial-temporal graph transformer framework focusing on trajectory prediction.
- **PECNet** (Mangalam et al. 2020): This model proposes a predicted endpoint-conditioned network that captures stochastic goals and introduces a novel non-local social pooling layer for human trajectory prediction.
- **GroupNet** (Xu et al. 2022a): This method proposes a multiscale hypergraph neural network for multi-agent trajectory prediction.
- **MemoNet** (Xu et al. 2022b): MemoNet is an instance-based trajectory prediction framework that mimics human memory by retrieving similar past scenarios from a memory bank to improve prediction performance.
- **MID** (Gu et al. 2022): This method introduces a parameterized Markov chain and a transformer-based diffusion model for trajectory prediction
- **NPSN** (Bae, Park, and Jeon 2022): This method tackles trajectory variability using Quasi-Monte Carlo sampling and introduces the Non-Probability Sampling Network (NPSN) for improved prediction.
- **DynGroupNet** (Xu et al. 2024): This paper introduces a dynamic-group-aware model improves interaction modeling for trajectory prediction.
- **EigenTraj** (Bae, Oh, and Jeon 2023): The paper presents EigenTrajectory (EigenTraj), a method for enhancing pedestrian trajectory prediction by constructing a compact ET space to represent movement patterns.
- **EqMotion** (Xu et al. 2023): EqMotion predicts trajectories with Euclidean equivariance and models interactions in a transformation-invariant manner.
- **LED** (Mao et al. 2023): LED is a diffusion-based trajectory predictor that generates diverse outputs efficiently by using a leapfrog initializer to skip denoising steps.
- **SingularTraj** (Bae, Park, and Jeon 2024): The paper presents SingularTrajectory (SingularTraj), a diffusion-based framework for pedestrian trajectory prediction.
- **MART** (Lee et al. 2024): This paper introduces multi-scale relational transformer for trajectory prediction.
- **PCHGCN** (Chen, Sang, and Zhao 2025): This paper introduces a high-order spatial-temporal framework for pedestrian trajectory prediction.

Implementation Details

As shown in Table 1, we summarize the key model components and training hyperparameters used in our ViTE framework. In the one-hop interaction expert, we adopt a two-layer Graph Convolutional Network (GCN) (Kipf and Welling 2017) to model local interactions. For the virtual graph module, the first stage employs a single-layer Graph Attention Network (GAT) (Veličković et al. 2018) for feature aggregation, while the second stage uses a single-layer GCN for feature distribution. We use the MultiStepLR scheduler with a decay factor of 0.5 to adjust the learning rate during training. For ETH/UCY datasets, we apply data augmentation following (Mohamed et al. 2022). The model is trained on a single NVIDIA TITAN Xp GPU for each dataset with Adam optimizer.

Supplementary Analysis

Virtual Node Mechanism Analysis

Figure 1 (left) shows the effect of virtual node count on prediction accuracy. We observe that using only one virtual node leads to suboptimal performance, likely due to limited capacity for modeling diverse global interactions. Performance improves and peaks when using three virtual nodes, indicating this configuration offers a good balance between expressiveness and efficiency. Adding more virtual nodes (e.g., four or five) results in slight degradation, suggesting that excessive

*Corresponding author
Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

Parameter	**ETH**	**HOTEL**	**UNIV**	**ZARA1**	**ZARA2**	**NBA**	**SDD**
Learning Rate	0.001	0.0018	0.001	0.0012	0.0012	0.0005	0.001
Training Epochs	300	300	300	300	300	100	300
Top-P Threshold	0.5	0.6	0.5	0.6	0.5	0.5	0.5
# Expert Router Layers	3	3	3	3	3	2	1
# Virtual Nodes	3	3	3	3	3	2	1

Table 1: Hyperparameter settings for each dataset.

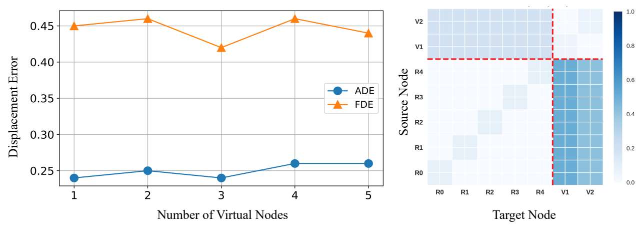

Figure 1: Analysis of virtual node mechanisms. Prediction accuracy (ADE/FDE) with varying virtual node counts on UNIV subset (**Left**). Attention map for message-passing between real nodes (R0–R4) and virtual nodes (V1–V2) (**Right**).

nodes may introduce redundant or noisy interactions that hinder learning.

To better understand how virtual nodes function, we visualize the attention weights between real nodes (R0–R4) and virtual nodes (V1–V2) in Figure 1 (right). The attention map reveals that both virtual nodes maintain strong connections with all real agents, confirming their role as global aggregators. Notably, V1 and V2 exhibit different attention patterns, indicating complementary specialization. The asymmetric structure of the map—where information flows from real nodes to virtual nodes and then back—also reflects our two-stage design, enabling virtual nodes to collect and redistribute high-level context across the graph.

More Qualitative Results

Figure 2 presents qualitative results of our framework across the five ETH/UCY subsets, arranged from left to right in order of increasing scene complexity. These scenarios exhibit diverse and stochastic pedestrian movement behaviors, including nonlinear movements, abrupt direction changes, and densely entangled trajectories. Despite such challenges, our model consistently produces accurate and socially compliant predictions that closely follow ground-truth trajectories.

In relatively simple environments such as **ETH** and **HOTEL**, the predicted trajectories align smoothly with historical trajectories. For mid-level complexity scenes like **ZARA1**, where subtle group interactions and moderate scene dynamics are present, our model effectively maintains prediction accuracy while preserving socially aware movement patterns. In the more complex subsets **ZARA2** and **UNIV**, which contain frequent agent crossings and rapid directional changes, our approach remains robust, accurately anticipating individual behaviors and inter-agent spacing. These results highlight the strong generalization ability of our expert-driven interaction modeling across both structured and highly dynamic pedestrian scenarios.

Failure Case Discussion

While our model demonstrates strong overall performance across diverse scenes, we observe certain failure cases in highly interactive or converging pedestrian scenarios. As shown in Figure 3, when multiple agents approach a shared goal region or intersecting paths, the model may predict future trajectories that lead to potential collisions. This issue often occurs when agents have ambiguous or overlapping motion patterns during the observation period. Without explicit collision avoidance or intent modeling, the model may overfit to historical movements and fail to anticipate social negotiation. While our interaction modeling performs well in general, future work could explore collision-aware objectives or intent-aware prediction to improve physical plausibility and social compliance.

References

Bae, I.; Oh, J.; and Jeon, H.-G. 2023. Eigentrajectory: Low-rank descriptors for multi-modal trajectory forecasting. In *ICCV*, 10017–10029.

Bae, I.; Park, J.-H.; and Jeon, H.-G. 2022. Non-probability sampling network for stochastic human trajectory prediction. In *CVPR*, 6477–6487.

Bae, I.; Park, Y.-J.; and Jeon, H.-G. 2024. Singulartrajectory: Universal trajectory predictor using diffusion model. In *CVPR*, 17890–17901.

Chen, W.; Sang, H.; and Zhao, Z. 2025. PCHGCN: Physically Constrained Higher-Order Graph Convolutional Network for Pedestrian Trajectory Prediction. *IEEE Internet of Things Journal*, 12(13): 25033–25045.

Gu, T.; Chen, G.; Li, J.; Lin, C.; Rao, Y.; Zhou, J.; and Lu, J. 2022. Stochastic Trajectory Prediction via Motion Indeterminacy Diffusion. In *CVPR*.

Kipf, T. N.; and Welling, M. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *ICLR*.

Lee, S.; Lee, J.; Yu, Y.; Kim, T.; and Lee, K. 2024. Mart: Multiscale relational transformer networks for multi-agent trajectory prediction. In *ECCV*, 89–107.

Mangalam, K.; Girase, H.; Agarwal, S.; Lee, K.-H.; Adeli, E.; Malik, J.; and Gaidon, A. 2020. It is not the journey but the destination: Endpoint conditioned trajectory prediction. In *ECCV*, 759–776.

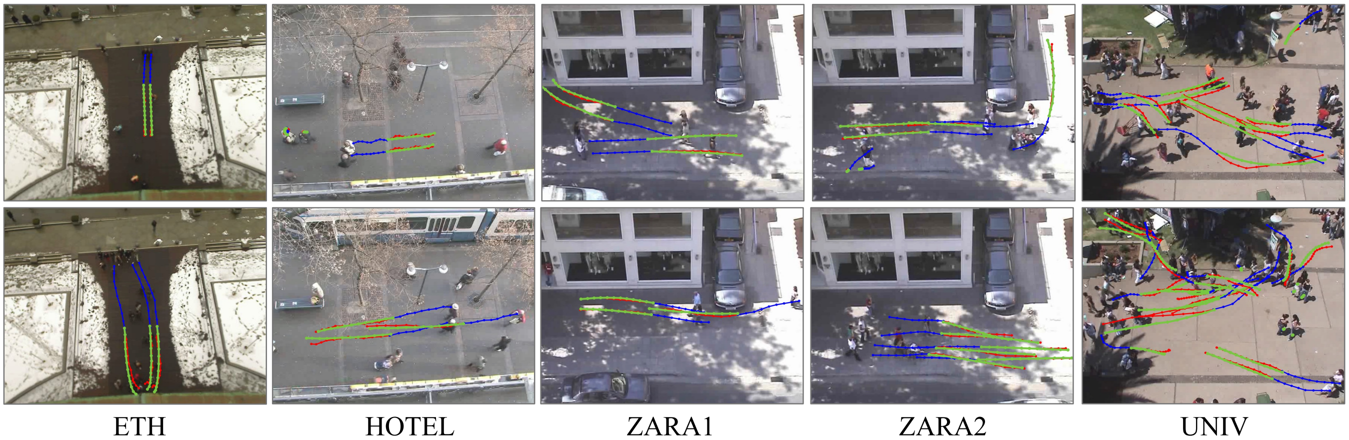

Figure 2: Qualitative results on ETH/UCY datasets. Historical trajectories are in blue, ground-truth trajectories are in red, and predicted trajectories are in green.

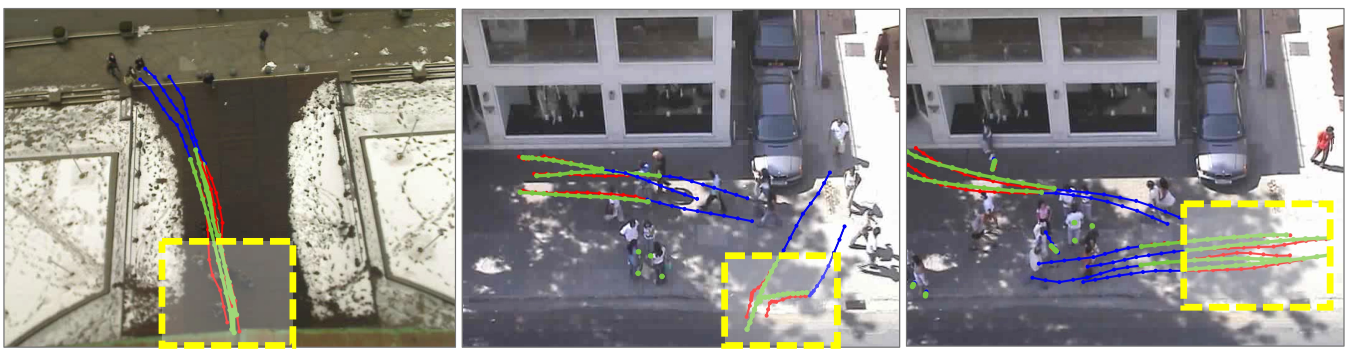

Figure 3: Illustration of a failure case. Historical trajectories are in blue, ground-truth trajectories are in red, and predicted trajectories are in green. Regions with potential trajectory overlap are highlighted by yellow boxes.

Mao, W.; Xu, C.; Zhu, Q.; Chen, S.; and Wang, Y. 2023. Leapfrog Diffusion Model for Stochastic Trajectory Prediction. In *CVPR*.

Mohamed, A.; Zhu, D.; Vu, W.; Elhoseiny, M.; and Claudel, C. 2022. Social-implicit: Rethinking trajectory prediction evaluation and the effectiveness of implicit maximum likelihood estimation. In *ECCV*, 463–479.

Veličković, P.; Cucurull, G.; Casanova, A.; Romero, A.; Lio, P.; and Bengio, Y. 2018. Graph attention networks. In *ICLR*.

Xu, C.; Li, M.; Ni, Z.; Zhang, Y.; and Chen, S. 2022a. GroupNet: Multiscale Hypergraph Neural Networks for Trajectory Prediction with Relational Reasoning. *CVPR*, 6488–6497.

Xu, C.; Mao, W.; Zhang, W.; and Chen, S. 2022b. Remember intentions: Retrospective-memory-based trajectory prediction. In *CVPR*, 6488–6497.

Xu, C.; Tan, R. T.; Tan, Y.; Chen, S.; Wang, Y. G.; Wang, X.; and Wang, Y. 2023. EqMotion: Equivariant Multi-Agent Motion Prediction With Invariant Interaction Reasoning. In *CVPR*.

Xu, C.; Wei, Y.; Tang, B.; Yin, S.; Zhang, Y.; Chen, S.; and Wang, Y. 2024. Dynamic-group-aware networks for multi-agent trajectory prediction with relational reasoning. *Neural Networks*, 170: 564–577.

Yu, C.; Ma, X.; Ren, J.; Zhao, H.; and Yi, S. 2020. Spatio-temporal graph transformer networks for pedestrian trajectory prediction. In *ECCV*, 507–523.