Physics-Guided Multi-Agent Trajectory Prediction with Adaptive Fused Graph Neural Architectures

Jeffrey C. Phillips
Department of Computer Science, University of Alabama at Birmingham, Birmingham, AL, USA.
jeffrey1971@uab.edu

Mateo Thornton
Department of Computer Science, Colorado State University, Fort Collins, CO, USA.
mateo2002@colostate.edu

Cody Wolfe
Department of Computer Science, University of Central Florida, Orlando, FL, USA.
cody.work@ucf.edu

Tongjing Wan
Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, USA.
tongjingwork@ku.edu

Abstract

Accurate trajectory prediction for multiple interacting agents is a cornerstone of autonomous systems, ranging from self-driving vehicles to drone swarms and pedestrian tracking. While graph neural architectures have advanced the modeling of spatiotemporal dependencies among agents, they often produce predictions that violate fundamental physical laws, such as momentum conservation or collision avoidance. This paper proposes a physics-guided multi-agent trajectory prediction framework that integrates adaptive fused graph neural architectures with explicit physical constraints. The system architecture comprises a spatiotemporal graph encoder that captures agent interactions, a physics-guided module that enforces Newtonian dynamics and geometric consistency, and an adaptive fusion mechanism that dynamically balances data-driven and physics-based predictions based on contextual cues. The paper emphasizes system-level considerations, including structural trade-offs between physical fidelity and model flexibility, computational efficiency, and robustness to noisy observations. It further examines governance challenges related to certification and interpretability, fairness implications arising from biased training data or physics priors, and sustainability concerns in deployment across resource-constrained platforms. Through cross-domain case illustrations involving autonomous driving, pedestrian crowds, and aerial swarms, the study highlights how adaptive fusion can improve prediction reliability while maintaining generalizability. The work concludes with a forward-looking perspective on integrating uncertainty quantification, lifelong learning, and policy frameworks to ensure safe and equitable deployment of physics-guided trajectory prediction systems.

Keywords

trajectory prediction, multi-agent systems, graph neural networks, physics-guided machine learning, adaptive fusion, socio-technical infrastructure, robustness, fairness.

1. Introduction

The reliable anticipation of future trajectories for multiple interacting agents is essential for the safe operation of autonomous vehicles, coordinated drone swarms, and pedestrian-aware navigation systems. In recent years, deep learning approaches, particularly recurrent and graph neural networks, have achieved impressive performance in modeling complex spatiotemporal dependencies from observational data [1, 2, 3]. However, these purely data-driven models often produce trajectories that are physically implausible, such as sudden acceleration changes, interpenetration of agents, or violations of momentum conservation. Such inconsistencies undermine trust in autonomous systems operating in safety-critical environments. Physics-guided machine learning has emerged as a promising direction to enforce physical constraints during training and inference, but its integration with graph-based multi-agent architectures remains underexplored from a systems perspective [4, 5].

This paper introduces a physics-guided multi-agent trajectory prediction framework that combines adaptive fused graph neural architectures with explicit physical priors. The core idea is to construct a graph neural encoder that reasons over spatial and temporal edges among agents, while a separate physics-guided module imposes constraints derived from Newtonian dynamics, collision geometry, and domain-specific regularities. An adaptive fusion mechanism learns to weigh these two sources of prediction based on contextual features, such as crowd density, road topology, or sensor noise levels. The framework is designed to be modular and extensible, allowing domain experts to substitute different physical priors or graph structures without retraining the entire system.

The paper adopts a system-level orientation, emphasizing the structural trade-offs inherent in combining physics with learning. These trade-offs include the tension between strict physical adherence and the flexibility to model non-Newtonian behaviors in human crowds, the computational cost of simulating physics versus the efficiency of end-to-end learning, and the risk of introducing bias through oversimplified physics assumptions. Additionally, the paper addresses governance, robustness, fairness, and sustainability considerations that are often neglected in technical treatments. By examining deployment scenarios in autonomous driving, pedestrian flow, and aerial swarms, the study provides concrete insights into how adaptive fusion can enhance prediction quality while mitigating unintended consequences.

The remainder of the paper is organized as follows. Section 2 reviews related work in trajectory prediction, graph neural networks, and physics-guided machine learning. Section 3 describes the system architecture and its key design choices. Section 4 analyzes the structural trade-offs associated with physics-guided constraints. Section 5 details the adaptive fusion mechanism and its dependence on contextual adaptation. Section 6 discusses governance, robustness, and fairness implications. Section 7 explores deployment and sustainability challenges. Section 8 concludes with a summary and future research directions.

2. Related Work

Multi-agent trajectory prediction has evolved from early social LSTM models that pooled hidden states across agents to more expressive graph-based formulations [1]. Alahi et al. introduced social pooling layers to model static interactions, but these methods were limited to pairwise proximity [1]. Vemula et al. proposed a spatiotemporal graph that explicitly edges both space and time [2]. Gupta et al. leveraged generative adversarial networks for multimodal trajectory sampling, but the outputs lacked physical constraints [3]. Graph attention networks have further improved interaction modeling by dynamically weighting edges based on feature similarity [7]. While these approaches capture complex patterns, they do not guarantee physical plausibility.

Physics-guided machine learning adds a layer of inductive bias by incorporating domain knowledge, often in the form of differential equations or energy functions [4, 5]. Raissi et al. demonstrated that physics-informed neural networks could solve partial differential equations while learning from sparse data [4]. For trajectory prediction, researchers have embedded simple kinematic constraints such as constant velocity or acceleration bounds into loss functions [8]. Schölkopf et al. provided a causal perspective, arguing that incorporating physical mechanisms improves generalization to out-of-distribution scenarios [9]. However, these physics-guided methods are typically designed for single agents or small systems, and their extension to large-scale multi-agent settings with heterogeneous interactions remains an open challenge.

The fusion of multiple prediction sources has been explored in ensemble learning and mixture-of-experts frameworks [10]. In the context of trajectory prediction, a few works have proposed adaptive weighting between data-driven and rule-based components. The required reference Zhu et al. introduced a flexible multi-generator model with a fused spatiotemporal graph for trajectory prediction, demonstrating that combining multiple generators with learnable fusion weights can improve accuracy under varying conditions [6]. This work aligns with our adaptive fusion philosophy, though our focus extends to physics guidance, system-level trade-offs, and socio-technical concerns. Other works have employed attention-based gating to combine kinematic and social features [11]. Despite these advances, there is no comprehensive systems analysis that addresses the structural, governance, and fairness dimensions of physics-guided adaptive fusion for multi-agent trajectory prediction.

3. System Architecture and Design

The proposed framework operates in three main stages: graph construction, physics-guided encoding, and adaptive fusion. First, historical trajectories of all agents are used to construct a spatiotemporal graph where each node represents an agent at a given timestep. Edges are defined spatially between agents within a threshold distance and temporally between consecutive timesteps for the same agent. The graph neural encoder employs multiple layers of message passing that aggregate information from neighbors, updating node representations to encode both spatial interactions and temporal dynamics [12]. This encoder is designed to be permutation-invariant and can handle variable numbers of agents.

Simultaneously, a physics-guided module operates on the same input trajectories to produce a set of physically plausible future positions. This module encodes domain-specific constraints such as constant velocity or acceleration, collision avoidance via repulsive potentials, and momentum conservation when mass estimates are available. For autonomous driving, the module may incorporate lane geometry and traffic rules; for pedestrian crowds, it may model social forces that prevent overlap [13]. These constraints are implemented as differentiable penalty terms during training, and during inference they can be applied as corrections to the graph encoder's output.

The adaptive fusion mechanism takes the outputs of the graph encoder and the physics-guided module and combines them into a final trajectory distribution. A gating network, implemented as a small feedforward network, processes contextual features such as scene density, velocity variance, and sensor confidence to produce a scalar per agent at each prediction time step. This scalar determines the relative contribution of the two sources. In low-density, high-velocity scenarios typical of highway driving, the physics-guided module is weighted more heavily because its assumptions hold. In crowded, low-velocity pedestrian zones where human intentions are less predictable, the data-driven graph encoder receives higher weight. The fused output is then used to parameterize a mixture of bivariate Gaussian distributions, enabling multimodal predictions.

4. Physics-Guided Constraints and Structural Trade-offs

Embedding physical knowledge into trajectory prediction introduces a fundamental trade-off between fidelity and flexibility. On one hand, strict enforcement of Newtonian or geometric constraints ensures that predictions are physically realizable, which is critical for safety certification in autonomous systems. On the other hand, such constraints may fail to capture emergent behaviors that deviate from idealized physics, such as humans suddenly stopping or animals making erratic movements. Soft constraints applied as regularization can preserve flexibility but may still bias predictions toward conservative, physics-abiding motions that underestimate genuine variance [14].

The computational cost of physics simulation is another structural trade-off. Solving forward kinematic or collision avoidance models for hundreds of agents in real time can be expensive, especially when using numerical integration. Approximations such as linearized dynamics or grid-based potential fields reduce overhead but may lose accuracy. In the proposed architecture, the physics-guided module is designed as a lightweight feedforward network that learns to approximate the physical constraints, rather than solving them from scratch. This amortized inference trades some exactness for speed, which is acceptable in many deployment contexts where near-real-time prediction is required.

Different application domains impose different physical priors, and the framework must be adaptable. In autonomous driving, the physics of road-following with yaw rate limits is well-understood, and violations are easily detected. In drone swarm coordination, the constraints involve three-dimensional geometry and communication latency, while pedestrian crowds involve social forces that are only partly physical. The system must allow domain experts to define constraint sets without rewriting the entire architecture. This modularity comes at the cost of increased hyperparameter tuning, as the relative importance of each constraint must be calibrated. The adaptive fusion mechanism partially mitigates this by learning to ignore constraints when they are irrelevant, but over-reliance on a fixed set of physics priors can still harm performance in novel environments.

5. Adaptive Fusion Mechanisms

Adaptive fusion aims to dynamically balance the contributions of data-driven and physics-based predictions based on contextual information. The gating network receives a feature vector derived from the current scene, which may include the average speed of agents, the spatial density measured by pairwise distances, the variance of historical accelerations, and external metadata such as weather or road conditions. These features are processed through a small multi-layer perceptron that outputs a single weight per agent per time step. The final prediction is a convex combination of the graph encoder output and the physics-guided output, ensuring that the weight lies between zero and one.

This mechanism introduces a meta-learning capability: the system learns under what conditions physical priors are reliable and when data-driven flexibility is preferred. For example, in a densely crowded square where humans frequently stop and change direction abruptly, the physics-guided module’s constant-velocity assumption would be consistently violated, and the gating network learns to down-weight it. Conversely, on a sparsely trafficked highway with smooth traffic flow, the physics prior becomes highly predictive, and its weight increases. The gating network is trained jointly with the rest of the architecture via gradient descent, using a trajectory prediction loss such as negative log-likelihood [15].

Robustness to sensor noise and occlusions is enhanced through adaptive fusion. When input trajectories are noisy, the data-driven encoder may overfit to spurious patterns, while the physics-guided module provides a regularized baseline. The gating network can react to high noise variance by shifting weight toward the physics prior, reducing the impact of errors. However, if the noise is systematic, such as a biased sensor, the physics prior might incorrectly reinforce the bias. In such cases, the fusion mechanism should ideally detect anomalies through an additional confidence module, which is an area for future work.

Fairness considerations arise because adaptive fusion may treat different agent types or demographic groups disproportionately. For instance, if the training data contains mostly vehicle trajectories, the gating network might learn to rely heavily on physics priors that assume road-following behavior. When applied to cyclists or pedestrians, the same priors may be inappropriate, leading to poorer prediction accuracy for those agents. To mitigate this, the contextual features should include an agent-type indicator, and the training data must be balanced across agent categories. Furthermore, fairness metrics such as equalized prediction error across groups should be monitored during validation.

6. Governance, Robustness, and Fairness Considerations

Deploying a physics-guided trajectory prediction system in safety-critical domains requires robust governance frameworks. Certification bodies need to verify that the system adheres to physical constraints under all foreseeable conditions. The adaptive fusion mechanism introduces a challenge because the effective behavior of the model changes with context, making static certification difficult. One approach is to freeze the gating network at a conservative setting during certification, then allow adaptation only in low-risk operational design domains. Another is to implement an independent monitor that flags when the fusion weight falls outside a safe range, triggering a fallback policy.

Robustness to adversarial perturbations is a key concern. An adversary could inject small perturbations into agent trajectories to cause the gating network to misclassify the context and allocate inappropriate weights. For example, slightly altering velocities might make a dense crowd appear sparse, causing the system to rely too heavily on physics priors and miss sudden stops. Adversarial training that includes worst-case perturbations of the contextual features can harden the gating network [16]. Additionally, the physics-guided module itself can serve as a defense by rejecting inputs that lead to physically impossible outputs; if the predicted trajectory violates basic kinematics despite high data-driven weight, the system can default to the physics prior.

Fairness extends beyond agent types to include geographical and socioeconomic disparities. Training data gathered predominantly from affluent urban areas may encode driving behaviors that differ from rural or low-income regions. The physics priors based on standard road geometry may not apply in informal road networks. The adaptive fusion mechanism could partly compensate by learning to down-weight physics in unfamiliar settings, but if the test distribution is completely alien, the system may still fail. Policy recommendations include requiring diverse training datasets and instituting mandatory fairness audits before deployment. Moreover, transparency about the limitations of physics priors should be communicated to end users and regulators.

7. Deployment and Sustainability

Real-world deployment of multi-agent trajectory prediction systems imposes stringent requirements on latency, memory, and energy consumption. The proposed framework involves multiple neural networks—the graph encoder, physics module, and gating network—which together may exceed the computational budget of embedded platforms. Model compression techniques such as quantization, pruning, and knowledge distillation can reduce the footprint without excessive loss of accuracy [17]. The graph encoder, being the heaviest component, can be replaced with a lightweight spatiotemporal transformer that uses linear attention to scale linearly with the number of agents [18].

Sustainability considerations encompass not only energy efficiency but also the carbon footprint of training and inference. Large graph neural networks trained on months of driving data consume significant energy. Incorporating physics priors can reduce the need for training data, as physical constraints provide a strong inductive bias that requires fewer examples to generalize. This data efficiency contributes to sustainability by shortening training cycles and reducing computing resource usage. However, the physics module itself may require manual tuning or simulation data, which carries its own environmental cost. A life cycle analysis should weigh these factors.

Cross-domain deployment is facilitated by the modular architecture. The same framework can be applied to traffic flow prediction, robot swarm coordination, and sports analytics by swapping the physics module and retraining the gating network. The ability to transfer learned fusion behavior across domains without extensive retraining is an open research problem. Hierarchical graph representations, where agents are grouped into clusters, can improve scalability for swarms with thousands of drones or pedestrians in a stadium [19]. Adaptive fusion at the cluster level reduces computational complexity while preserving overall prediction quality.

8. Conclusion

This paper has presented a system-level perspective on physics-guided multi-agent trajectory prediction using adaptive fused graph neural architectures. The proposed framework integrates a spatiotemporal graph encoder, a physics-guided constraint module, and a context-aware gating mechanism to produce physically plausible and contextually adaptive trajectory predictions. The analysis focused on structural trade-offs such as fidelity versus flexibility, computational cost versus accuracy, and domain specificity versus generality. Governance, robustness, fairness, and sustainability considerations were examined to highlight the socio-technical dimensions that are often overlooked in technical literature. The adaptive fusion mechanism offers a principled way to balance data-driven and physics-based reasoning, but its deployment raises challenges in certification, adversarial robustness, and equitable treatment of all agents. Future work should explore online adaptation of physics priors through continual learning, integration with reinforcement learning for closed-loop control, and the development of standardized fairness benchmarks for trajectory prediction. As autonomous systems become more prevalent, ensuring that their predictive components are physically grounded, contextually aware, and socially responsible will be essential for building trust and enabling safe coexistence.

References

1. Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., & Savarese, S. (2016). Social LSTM: Human trajectory prediction in crowded spaces. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 961–971). IEEE.
2. Vemula, A., Muelling, K., & Oh, J. (2018). Social attention: Modeling attention in human crowds. In Proceedings of the IEEE International Conference on Robotics and Automation (pp. 4601–4607). IEEE.
3. Gupta, A., Johnson, J., Fei-Fei, L., Savarese, S., & Alahi, A. (2018). Social GAN: Socially acceptable trajectories with generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2255–2264). IEEE.
4. Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. Journal of Computational Physics, 378, 686–707.
5. Kipf, T. N., & Welling, M. (2017). Semi-supervised classification with graph convolutional networks. In Proceedings of the International Conference on Learning Representations.
6. Zhu, P., Han, F., & Deng, H. (2023, December). Flexible multi-generator model with fused spatiotemporal graph for trajectory prediction. In IET Conference Proceedings CP874 (Vol. 2023, No. 47, pp. 417-422). Stevenage, UK: The Institution of Engineering and Technology.
7. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In Advances in Neural Information Processing Systems (Vol. 30, pp. 5998–6008).
8. Schölkopf, B., Locatello, F., S., L. (2021). Towards causal representation learning. Proceedings of the IEEE, 109(5), 612–634.
9. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural Computation, 9(8), 1735–1780.
10. Jacobs, R. A., Jordan, M. I., Nowlan, S. J., & Hinton, G. E. (1991). Adaptive mixtures of local experts. Neural Computation, 3(1), 79–87.
11. Lerner, A., Chrysanthou, Y., & Lischinski, D. (2007). Crowds by example. Computer Graphics Forum, 26(3), 655–664.
12. Gilmer, J., Schoenholz, S. S., Riley, P. F., Vinyals, O., & Dahl, G. E. (2017). Neural message passing for quantum chemistry. In Proceedings of the International Conference on Machine Learning (pp. 1263–1272). PMLR.
13. Helbing, D., & Molnar, P. (1995). Social force model for pedestrian dynamics. Physical Review E, 51(5), 4282–4286.
14. Lu, L., Jin, P., Pang, G., Zhang, Z., & Karniadakis, G. E. (2021). Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. Nature Machine Intelligence, 3(3), 218–229.
15. Salakhutdinov, R., & Hinton, G. (2009). Deep Boltzmann machines. In Proceedings of the International Conference on Artificial Intelligence and Statistics (pp. 448–455). PMLR.

16. Madry, A., Makelov, A., Schmidt, L., Tsipras, D., & Vladu, A. (2018). Towards deep learning models resistant to adversarial attacks. In Proceedings of the International Conference on Learning Representations.
17. Cheng, Y., Wang, D., Zhou, P., & Zhang, T. (2017). A survey of model compression and acceleration for deep neural networks. arXiv preprint arXiv:1710.09282.
18. Choromanski, K., Likhosherstov, V., Dohan, D., Song, X., Gane, A., Sarlos, T., ... & Weller, A. (2021). Rethinking attention with performers. In Proceedings of the International Conference on Learning Representations.
19. Schlichtkrull, M., Kipf, T. N., Bloem, P., van den Berg, R., Titov, I., & Welling, M. (2018). Modeling relational data with graph convolutional networks. In Proceedings of the European Semantic Web Conference (pp. 593–607). Springer.