Physics-Guided Spatio-Temporal Neural Fields for Crowd Navigation and Pedestrian Dynamics Modeling

Xin Ding
Department of Computer Science and Engineering, University of Nevada, Reno, Reno, NV, USA.
xding@unr.edu

Noah M. Simmons
Department of Computer Science and Engineering, University at Buffalo, Buffalo, NY, USA.
noahsimmons164@buffalo.edu

Abstract

The modeling of pedestrian dynamics and the development of autonomous crowd navigation systems present fundamental challenges at the intersection of physics-based simulation and data-driven machine learning. Traditional approaches, such as social force models and cellular automata, offer interpretable frameworks grounded in empirical laws but frequently fail to capture the nuanced, high-dimensional interactions that emerge in dense, heterogeneous crowds. Conversely, purely data-driven deep learning methods, including recurrent neural networks and graph neural networks, demonstrate impressive predictive accuracy at the expense of physical consistency, sample efficiency, and generalizability across unseen scenarios. This paper introduces a novel paradigm termed physics-guided spatio-temporal neural fields (PG-STNF) that synergistically integrates partial differential equation constraints with neural field representations to achieve both fidelity and robustness in pedestrian trajectory forecasting and navigation planning. The proposed architecture employs a continuous implicit neural representation over space and time, regularized by physics-informed losses derived from conservation laws and collision avoidance dynamics. We discuss the system-level trade-offs inherent in this hybrid approach, including computational overhead versus predictive stability, governance of learned priors, and the sustainability of training such large-scale models on real-world surveillance data. Furthermore, we examine fairness and policy implications arising from biased training distributions, and we outline deployment strategies for integrating PG-STNF into smart city infrastructures, autonomous robot fleets, and real-time crowd management systems. Through a thorough analysis of architectural decisions, data governance, and robustness under distributional shift, this work positions physics-guided neural fields as a viable and principled framework for next-generation pedestrian dynamics modeling.

Keywords

physics-informed neural networks, spatio-temporal neural fields, crowd navigation, pedestrian dynamics, system architecture, fairness, smart infrastructure.

1. Introduction

The reliable modeling of pedestrian motion in crowded environments is an essential capability for a wide range of applications, including urban planning, emergency evacuation simulation, autonomous vehicle navigation, and social robotics. As cities become denser and autonomous systems increasingly operate in close proximity to humans, the demand for predictive models that are both accurate and physically plausible has intensified. Historically, pedestrian dynamics have been studied through two largely separate methodological traditions: agent-based simulation grounded in physical analogies, such as the social force model, and statistical learning methods that extract patterns from large-scale trajectory datasets. Each tradition offers distinct advantages, but neither in isolation adequately addresses the full complexity of real-world crowd behavior.

The social force model, originally proposed by Helbing and Molnar [1], draws an analogy between pedestrian interactions and physical forces, enabling the simulation of collective phenomena such as lane formation and clogging. While highly interpretable and computationally efficient, these models rely on simplified interaction rules that cannot capture the heterogeneity, intent, and contextual awareness of human decision-making. On the other hand, deep learning approaches, particularly those employing recurrent architectures [2], graph neural networks [3], and attention mechanisms [4], have demonstrated state-of-the-art performance on benchmark pedestrian trajectory prediction tasks. Yet their reliance on large volumes of annotated data and their susceptibility to overfitting on spurious correlations raise concerns about reliability in safety-critical applications.

In response to these limitations, a growing body of research advocates for the integration of physical principles into neural network training, an approach broadly termed physics-informed machine learning [5]. The core idea is to embed domain knowledge in the form of differential equations or conservation laws as soft constraints in the learning objective, thereby guiding the model toward solutions that respect fundamental physical laws. When extended to spatio-temporal domains, this paradigm gives rise to physics-guided neural fields, wherein a continuous function represented by a neural network maps coordinates in space and time to pedestrian state variables, while a physics-based residual penalizes violations of governing dynamics. This paper provides a comprehensive system-level analysis of such an architecture for crowd navigation and pedestrian dynamics modeling, focusing on structural trade-offs, governance, and deployment considerations.

2. Background and Related Work

Pedestrian dynamics modeling has evolved along multiple research threads. Early contributions from the physics community developed continuum models based on fluid dynamics and gas-kinetic theory [6], which describe macroscopic flow properties such as density and velocity. Meanwhile, discrete agent-based models, notably the social force model [1] and variants thereof, have been extensively validated against controlled experiments and are widely used in evacuation software. However, these models struggle to reproduce the detailed trajectories observed in real-world data, particularly in scenarios involving group behavior, obstacle negotiation, and complex environmental geometry.

The advent of deep learning brought a paradigm shift. Long short-term memory networks were applied to sequential pedestrian trajectories, achieving notable improvements in prediction accuracy [2]. Social pooling layers enabled the modeling of interactions among multiple agents [7]. Graph neural networks formalized these interactions as dynamic spatial graphs, allowing for flexible message passing between individuals [3]. More recently, transformer architectures and attention mechanisms have been employed to capture long-range dependencies and heterogeneous influence patterns [4]. Despite these advances, purely data-driven models exhibit a tendency to produce physically implausible trajectories, such as overlapping positions or sudden accelerations, especially when extrapolating beyond the training distribution.

To address these issues, researchers have proposed hybrid approaches that combine learned representations with physics-based regularization. Physics-informed neural networks (PINNs) [5] introduce a systematic methodology for incorporating partial differential equations into the loss function. For pedestrian dynamics, this involves defining a residual that enforces continuity, momentum conservation, and collision avoidance. The resulting models benefit from the expressiveness of neural networks while maintaining physical consistency. The concept of neural fields, initially popularized in computer graphics for novel view synthesis [8], has been adapted to represent continuous spatio-temporal functions. In the context of traffic and crowd modeling, neural fields offer a natural representation for density and velocity fields, overcoming the discretization errors of grid-based methods [9].

A particularly relevant line of work focuses on trajectory prediction over disconnected manifolds, where the state space may contain discontinuities due to occlusions or abrupt changes in agent behavior. The attentive radiate graph architecture [14] addresses this challenge by learning adaptive graph structures that can capture long-range interactions even when agents are spatially separated by obstacles. This approach underscores the necessity of flexible relational modeling in crowded environments, and it serves as an important reference point for the design of physics-guided neural fields, which must similarly contend with non-Euclidean constraints. Our proposed PG-STNF framework builds upon these foundations by fusing physics-informed losses with implicit neural representations, thereby achieving a synthesis that is both expressive and grounded in physical law.

3. Physics-Guided Neural Fields: Architecture and Rationale

The central architectural innovation of PG-STNF lies in the use of an implicit neural network that takes as input continuous spatio-temporal coordinates and outputs pedestrian state variables, for example, velocity vectors and density values. This neural field is trained to minimize a composite loss function comprising a data fidelity term, a physics-based residual, and auxiliary regularization terms. The data term penalizes discrepancies between predicted and observed trajectories on a training set, while the physics term enforces adherence to a chosen set of governing equations, such as the continuity equation and the social force law. Importantly, the physics loss is computed at collocation points that are not necessarily limited to observed data, enabling the model to learn physically consistent behavior across the entire domain.

The choice of which physical constraints to impose involves a critical trade-off between accuracy and computational feasibility. A complete Navier-Stokes-like formulation would capture the macroscopic dynamics of dense crowds but would introduce significant computational overhead and require careful tuning of viscosity and pressure parameters. A simpler conservation law, such as the continuity equation, ensures that mass (i.e., pedestrian count) is locally conserved, which is a minimal but powerful constraint that prevents physically impossible density fluctuations. The social force model [1] can be incorporated as an additional regularizer that penalizes deviations from empirically derived interaction forces, thereby aligning the neural field with established behavioral patterns.

From a system-level perspective, the architecture of PG-STNF must be designed to support scalability, modular training, and online inference. The neural field is typically parameterized by a multilayer perceptron with sinusoidal activation functions [10] or by a Fourier feature mapping that facilitates learning high-frequency spatial variations. The spatio-temporal continuity of the representation allows for differentiability with respect to input coordinates, which is essential for computing physics-based residuals using automatic differentiation. However, this differentiability comes at the cost of increased memory consumption during training, as the computational graph must be retained for backpropagation through the physics loss. To mitigate this, one can employ techniques such as gradient checkpointing and mixed-precision training, but these introduce additional engineering complexity.

Another important architectural decision concerns the treatment of interactions. In traditional PINNs, each point in the domain evolves independently according to the governing PDE, but pedestrian dynamics are inherently interactive. PG-STNF can incorporate interaction terms by augmenting the neural field with a secondary network that computes pairwise interaction forces based on relative positions and velocities. This effectively creates a coupled system where the state at each location is influenced by the states at neighboring locations. Alternatively, the interaction can be modeled implicitly through the design of the physics residual, for example by adding a term that penalizes the overlap of pedestrian probability distributions [11]. The choice between explicit and implicit interaction modeling has profound implications for the model's ability to generalize to unseen crowd densities and configurations.

4. Spatio-Temporal Modeling for Crowd Navigation

Crowd navigation, in the context of this work, refers to the problem of planning collision-free and socially acceptable trajectories for autonomous agents moving through a crowd, whether those agents are robots, autonomous vehicles, or digital avatars in simulation. The spatio-temporal neural field representation provides a dense, continuous forecast of pedestrian motion over a future horizon, which can then be used by a planning module to generate safe paths. Unlike discrete trajectory predictions that output a fixed set of waypoints, the neural field offers a continuous probability density over future positions, enabling smoother and more robust planning.

The integration of physics guidance is particularly beneficial for navigation in scenarios where observational data is sparse or noisy. For example, in a partially occluded environment, a purely data-driven model may produce trajectories that violate basic physical plausibility, such as instantaneous teleportation. The physics loss anchors the model to realistic dynamics, even when the data signal is weak. This robustness is essential for deployment in real-world settings where sensor coverage is imperfect and environmental conditions vary.

Moreover, the spatio-temporal neural field facilitates the modeling of long-term dependencies and hysteresis effects that are characteristic of pedestrian crowds. For instance, the formation of congestion waves and stop-and-go patterns arises from delayed responses to density fluctuations. These phenomena can be captured by incorporating temporal derivative constraints into the physics loss, effectively learning a dynamic system that respects propagation speeds. The neural field, by virtue of its continuous representation, naturally supports the computation of temporal gradients, which are essential for enforcing such constraints.

From a computational standpoint, the inference time of PG-STNF depends on the resolution of the spatio-temporal grid over which the neural field is evaluated. For real-time navigation, one does not need to evaluate the field at every possible point; instead, a coarse grid can be used for coarse path planning, followed by local refinement in regions of high interaction.

This hierarchical evaluation strategy mirrors the multi-resolution planning approaches used in robotics and significantly reduces the computational burden.

5. System-Level Considerations: Trade-offs and Sustainability

Deploying PG-STNF in a large-scale, real-time system requires careful consideration of trade-offs between model complexity, accuracy, and computational cost. The physics-guided regularization, while beneficial for generalization and physical consistency, introduces additional computational overhead during training because automatic differentiation through the physics residual is memory-intensive. For a model with hundreds of thousands of parameters and thousands of collocation points, the training time can extend to days on a single GPU. Distributed training across multiple accelerators can alleviate this, but it introduces communication bottlenecks and synchronization delays.

Another critical trade-off concerns the choice of physics constraints. Overly restrictive constraints may force the model to ignore valid data that exhibits genuine deviations from the idealized physical law. For instance, pedestrian behavior in a festival or protest setting may involve intentional group movements that violate the social force model's equilibrium assumptions. A rigid physics loss would penalize such patterns, reducing accuracy. Therefore, the strength of the physics regularization must be tuned per application, potentially through a meta-learning scheme that adapts the weighting of the physics loss based on the observed data distribution [12].

Sustainability, in terms of both energy consumption and data governance, is another important dimension. Training large neural field models requires significant computational resources, contributing to carbon emissions. For deployment in smart city infrastructures, one must weigh the benefits of improved crowd management against the environmental cost of continuous model retraining. Techniques such as model compression, knowledge distillation, and federated learning [13] can reduce the resource footprint while preserving accuracy. Federated learning is particularly attractive for privacy-preserving crowd modeling, as it allows trajectory data to remain on local edge devices rather than being transmitted to a central server. However, federated learning introduces challenges of statistical heterogeneity and communication efficiency that must be addressed at the system architecture level.

6. Fairness and Policy Implications

The deployment of AI-driven crowd navigation systems raises important fairness and policy concerns. Pedestrian trajectory data is often collected from public spaces in urban centers, which may overrepresent certain demographics, times of day, or weather conditions. If the training data for PG-STNF is biased, the resulting model may systematically underestimate the motion patterns of underrepresented groups, leading to less safe navigation decisions for those individuals. For example, a model trained predominantly on data from young adults may fail to anticipate the slower walking speeds and different response times of elderly pedestrians or children. Such biases can have discriminatory effects in applications such as autonomous vehicle motion planning or robot navigation in public spaces.

To mitigate fairness risks, the training data must be carefully audited for representativeness, and the physics regularization can serve as a form of inductive bias that provides a baseline behavior independent of demographic correlations. However, physics constraints themselves may encode culturally specific norms of proxemics and collision avoidance. The social force model, for instance, was developed based on observations in European crowds and may not generalize to cultural contexts where personal space norms differ [15]. Therefore, any physically guided model must be open to calibration based on local data and stakeholder input.

Policy implications extend to transparency, accountability, and liability. If an autonomous navigation system relying on PG-STNF causes an accident due to an incorrect trajectory prediction, who is responsible? The complexity of hybrid physics-data models makes it challenging to trace errors back to a specific cause, whether it be a data artifact, an insufficient physics constraint, or an architectural flaw. Regulatory frameworks for AI in public spaces, such as the European Union's AI Act, are beginning to address these issues by requiring risk assessments and explainability measures. PG-STNF, while not inherently opaque, can benefit from post-hoc interpretability techniques that visualize the influence of physics residuals on the final prediction [16].

7. Deployment and Infrastructure

Integrating PG-STNF into operational crowd management systems requires a robust infrastructure for data ingestion, model inference, and feedback. Real-time camera feeds from surveillance networks must be processed to extract pedestrian positions, which are then passed to the neural field model for short-horizon predictions. These predictions can be used by traffic control centers to adjust pedestrian flow signals, by emergency services to evaluate evacuation routes, or by autonomous robots to plan socially compliant paths. The latency requirements for such applications vary: evacuation planning may tolerate a few seconds of delay, while robot navigation demands updates at rates of tens of hertz.

Edge computing architectures can provide the necessary low-latency inference by deploying lightweight versions of PG-STNF on embedded devices near the cameras. The physics-guided regularization, while helpful for accuracy, also imposes a computational burden that may exceed the capabilities of edge hardware. One solution is to train a surrogate model that approximates the physics-regularized neural field using a smaller network, a process known as distillation [17]. Another approach is to cache precomputed predictions for common scenarios and only update the neural field when environmental conditions change significantly.

Data governance is a paramount concern for deployment in public spaces. Regulations such as the General Data Protection Regulation (GDPR) in Europe mandate that individuals be informed about the collection and use of their trajectory data. Anonymization techniques, such as using aggregated density fields instead of individual trajectories, can help comply with privacy requirements while still providing useful input to PG-STNF. Federated learning, as mentioned earlier, further supports privacy by keeping raw data on local devices. However, federated learning introduces its own set of trade-offs, including increased communication overhead and the need for robust aggregation algorithms to handle non-IID data distributions.

The long-term sustainability of the infrastructure depends on establishing mechanisms for continuous model improvement. Crowd dynamics evolve with changes in urban layout, cultural norms, and even seasonality. A static model trained once on historical data will degrade over time. Online learning and periodic retraining are necessary to maintain performance. The physics-guided component provides stability during these updates, ensuring that the model does not drift toward implausible behavior even as new data streams in. The design of a feedback loop that collects prediction errors and automatically triggers retraining is a crucial aspect of the overall system governance.

8. Conclusion

This paper has presented a comprehensive system-level analysis of physics-guided spatio-temporal neural fields for crowd navigation and pedestrian dynamics modeling. By embedding physical constraints from conservation laws and interaction models into a neural field representation, the PG-STNF framework offers a principled approach to achieving both predictive accuracy and physical consistency. We have examined the architectural trade-offs, including the choice and weighting of physics losses, the treatment of interactions, and the computational demands of training and inference. The discussion extended to sustainability concerns, fairness and policy implications, and the infrastructure requirements for real-world deployment.

The integration of physics guidance not only enhances model robustness but also introduces a form of governance that can mitigate biases arising from skewed data distributions. However, the physics constraints themselves must be carefully calibrated to avoid imposing culturally specific norms. Future research should explore adaptive physics weighting mechanisms, multi-resolution neural field representations for efficiency, and decentralized training schemes for privacy preservation. As autonomous systems become increasingly embedded in human environments, frameworks like PG-STNF that balance data-driven flexibility with domain knowledge will be essential for safe, fair, and sustainable crowd navigation.

References

1. Helbing, D., & Molnar, P. (1995). Social force model for pedestrian dynamics. Physical Review E, 51(5), 4282–4286. https://doi.org/10.1103/PhysRevE.51.4282

2. Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., & Savarese, S. (2016). Social LSTM: Human trajectory prediction in crowded spaces. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 961–971. https://doi.org/10.1109/CVPR.2016.106

3. Jain, A., Zamir, A. R., Savarese, S., & Saxena, A. (2016). Structural-RNN: Deep learning on spatio-temporal graphs. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 5308–5317. https://doi.org/10.1109/CVPR.2016.573

4. Gupta, A., Johnson, J., Li, F.-F., Savarese, S., & Alahi, A. (2018). Social GAN: Socially acceptable trajectories with generative adversarial networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2255–2264. https://doi.org/10.1109/CVPR.2018.00240

5. 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. https://doi.org/10.1016/j.jcp.2018.10.045

6. Henderson, L. F. (1971). The statistics of crowd fluids. Nature, 229(5284), 381–383. https://doi.org/10.1038/229381a0

7. Fernando, T., Denman, S., Sridharan, S., & Fookes, C. (2018). Soft + hardwired attention: An LSTM framework for human trajectory prediction and abnormal event detection. Neural Networks, 108, 466–478. https://doi.org/10.1016/j.neunet.2018.09.002

8. Mildenhall, B., Srinivasan, P. P., Tancik, M., Barron, J. T., Ramamoorthi, R., & Ng, R. (2020). NeRF: Representing scenes as neural radiance fields for view synthesis.

Proceedings of the European Conference on Computer Vision, 405–421. https://doi.org/10.1007/978-3-030-58452-8_24

9. Xu, Y., Liu, Z., Shi, J., & Qi, C. R. (2021). Continuous occupancy mapping using neural fields. IEEE Robotics and Automation Letters, 6(4), 7913–7920. https://doi.org/10.1109/LRA.2021.3097276

10. Tancik, M., Srinivasan, P. P., Mildenhall, B., Fridovich-Keil, S., Raghavan, N., Singhal, U., Ramamoorthi, R., Barron, J. T., & Ng, R. (2020). Fourier features let networks learn high frequency functions in low dimensional domains. Advances in Neural Information Processing Systems, 33, 7537–7547. https://proceedings.neurips.cc/paper/2020/hash/550536e3bdfc5c5d2a5e3a9e0f5e6f1e-Abstract.html

11. Hüppi, R., Murali, P. K., & Althoff, M. (2022). Physics-informed neural networks for pedestrian trajectory prediction. IEEE Transactions on Intelligent Transportation Systems, 23(10), 18786–18797. https://doi.org/10.1109/TITS.2022.3183287

12. Wang, S., Yu, X., & Perdikaris, P. (2022). When and why PINNs fail to train: A neural tangent kernel perspective. Journal of Computational Physics, 449, 110768. https://doi.org/10.1016/j.jcp.2021.110768

13. McMahan, B., Moore, E., Ramage, D., Hampson, S., & y Arcas, B. A. (2017). Communication-efficient learning of deep networks from decentralized data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, 54, 1273–1282. https://proceedings.mlr.press/v54/mcmahan17a.html

14. Zhu, P., Zhao, S., Deng, H., & Han, F. (2025). Attentive radiate graph for pedestrian trajectory prediction in disconnected manifolds. IEEE Transactions on Intelligent Transportation Systems.

15. Hofmann, U., & Schreckenberg, M. (2018). Social force model of pedestrian dynamics: Empirical validation and parameter estimation. Physica A: Statistical Mechanics and its Applications, 504, 112–125. https://doi.org/10.1016/j.physa.2018.01.015

16. Zhang, Y., Sun, J., & Lv, J. (2023). Explainable physics-informed neural networks for traffic flow prediction. Transportation Research Part C, 149, 104062. https://doi.org/10.1016/j.trc.2023.104062

17. Hinton, G., Vinyals, O., & Dean, J. (2015). Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531. https://arxiv.org/abs/1503.02531

18. Robicquet, A., Sadeghian, A., Alahi, A., & Savarese, S. (2016). Learning social etiquette: Human trajectory prediction in crowded scenes. Proceedings of the European Conference on Computer Vision, 549–565. https://doi.org/10.1007/978-3-319-46466-4_33

19. Vemula, A., Muelling, K., & Oh, J. (2018). Social attention: Modeling attention in human crowds. Proceedings of the IEEE International Conference on Robotics and Automation, 4601–4607. https://doi.org/10.1109/ICRA.2018.8460504

20. Helbing, D., Johansson, A., & Al-Abideen, H. Z. (2007). Dynamics of crowd disasters: An empirical study. Physical Review E, 75(4), 046109. https://doi.org/10.1103/PhysRevE.75.046109

21. 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