Geometric Graph Neural Network Modeling of Human Interactions in Crowded Environments [1]

Sara Honarvar * **Yancy Diaz-Mercado** *

* *Department of Mechanical Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: honarvar@ umd.edu, yancy@ umd.edu)*

Abstract: Modeling human trajectories in crowded environments is challenging due to the complex nature of pedestrian behavior and interactions. This paper proposes a geometric graph neural network (GNN) architecture that integrates domain knowledge from psychological studies to model pedestrian interactions and predict future trajectories. Unlike prior studies using complete graphs, we define interaction neighborhoods using pedestrians' field of view, motion direction, and distance-based kernel functions to construct graph representations of crowds. Evaluations across multiple datasets demonstrate improved prediction accuracy through reduced average and final displacement error metrics. Our findings underscore the importance of integrating domain knowledge with data-driven approaches for effective modeling of human interactions in crowds.

Keywords: Machine Learning in modeling, estimation, and control; Modeling and Validation; Multi-agent and Networked Systems; Graph Neural Network; Crowd Navigation

1. INTRODUCTION

Over the last decade, predicting how humans move and interact has become an active field of research, driven by applications in autonomous driving, surveillance, and the creation of socially aware robots (see Rudenko et al. (2020); Mavrogiannis et al. (2023) and references therein). The challenge lies in anticipating human motion due to its inherently complex and stochastic nature, influenced by internal and external stimuli, social dynamics, and psychological decision-making factors as described in Wirth et al. (2022). This challenge is magnified in crowded spaces, where the interplay of these factors becomes even more intricate (see Korbmacher and Tordeux (2022) for a review).

Despite this complexity, humans demonstrate an impressive ability to navigate crowds. As we move, we instinctively stay close to our group while maintaining a social distance from others. Our behavior adapts based on the type of interaction at play. Understanding these implicit interactions is crucial for autonomous systems, enabling them to predict and adapt to the unpredictable behaviors of pedestrians in the absence of explicit communication.

Various modeling-based and learning-based algorithms have been proposed to understand human interactions when navigating through crowds, as reviewed in Korbmacher and Tordeux (2022). Model-based approaches, such as the social force model by Helbing and Molnar (1995) and velocity-based methods like the reciprocal n-body collision avoidance technique introduced by Van Den Berg et al. (2011), aim to develop intuitive strategies for explaining human interactions based on first principles and social factors. In contrast, learning-based methods leverage deep learning algorithms, including long short-term memory (LSTM) networks Alahi et al. (2016), convolutional neural networks (CNNs), generative adversarial networks (GANs) Gupta et al. (2018), and graph neural networks (GNNs) Mohamed et al. (2020), to predict human motions directly from data.

Both model-based and learning-based approaches have their advantages and disadvantages. Model-based algorithms typically have fewer parameters, making them computationally efficient and enabling careful reasoning about human intentions. However, they may overlook complex aspects of human behavior. Our proposed model takes the form of a geometric graph neural network, a structure inspired by the Social Spatio-Temporal Graph Neural Network (STGCNN) architecture proposed by Mohamed et al. (2020). However, we go beyond conventional GNN formulations by integrating geometric kernel functions and a refined neighborhood of interaction definition, drawing insights from psychological studies on human crowd behaviors (Rio et al. (2018); Wirth et al. (2022)). Our hybrid approach aims to combine the strengths of both model-based and learning-based methods by incorporating geometric kernel functions and refined interaction neighborhood definitions derived from psychological studies. These enhancements contribute to a more interpretable and context-aware weighted adjacency matrix, enabling a comprehensive framework for modeling pedestrian interactions.

⋆ This work was supported by the Diversity in Robotics and AI PhD Fellowships by Microsoft and Amazon Lab 126 from the Maryland Robotics Center.

[1] © 2024 the authors. This work has been accepted to IFAC for publication under a Creative Commons Licence CC-BY-NC-ND.

2. RELATED WORK

In recent years, deep learning approaches for predicting human motion have gained significant attention. These methods typically assume that pedestrian trajectories follow a bivariate Gaussian distribution and aim to predict trajectory distributions based on observed frames.

Social-LSTM proposed by Alahi et al. (2016) represents one of the early methods using a recurrent neural network to predict individual pedestrian motion. To incorporate human interactions, it introduces a social pooling layer that aggregates recurrent outputs and concatenates hidden states of nearby agents. Extending this idea, Social-GAN by Gupta et al. (2018) employs generative networks and a loss function encouraging multi-modality of pedestrian trajectory distributions, as agents' trajectories may have multiple possible futures. However, it assigns uniform weights to all surrounding pedestrians, failing to capture the importance of close neighbors.

Given the advancements of GNNs introduced in Kipf and Welling (2016) for explaining and modeling interactions between different entities, recent works have integrated graph representation GNNs to capture the interaction mechanism between humans, departing from traditional pooling methods (see Korbmacher and Tordeux (2022) for an extensive review).

GNN is a powerful tool for learning representations of network systems, but they rely heavily on the given network structure which may not accurately reflect the true interactions. Most approaches construct an initial graph based on node feature similarities, then refine it using metric (e.g., neural (e.g., attention mechanism such as the works by Huang et al. (2019) and Kosaraju et al. (2019)), kernel functions as in Mohamed et al. (2020)), or direct methods (e.g., optimization methods enumerated in Zhu et al. (2021)).

For instance, STGAT proposed by Huang et al. (2019) and social-BiGAT by Kosaraju et al. (2019) treat each agent as a node in a graph, and use Graph Attention Networks (GAT) to describe their interaction by weighing the importance of each pedestrian's contribution. Although attention mechanism is beneficial in terms of modeling the importance of each neighboring pedestrian, they exhibit drawbacks such as a lack of explainability and a substantial increase in training time and learnable parameters. In response to these limitations, alternative approaches like social STGCNN (Mohamed et al. (2020)) have emerged. This method proposes a graph representation for modeling crowd interaction based on spatio-temporal information, opting for a kernel function instead of an attention mechanism to weigh the importance of agents before constructing a graph representation of pedestrians. This strategic choice not only overcomes the shortcomings associated with attention mechanisms but also introduces a more interpretable and computationally efficient approach to assessing the importance of individual agents within a crowd.

Our work builds upon the kernel-based GNN framework proposed in Social STGCNN by Mohamed et al. (2020), but explores the impact of various geometric kernel functions to identify the neighborhood of interaction in crowds, determining which neighbors influence each pedestrian's behavior and to what extent. These proposed kernel functions draw inspiration from psychological studies (Rio et al. (2018)) that conducted experiments on human crowds to understand the underlying mechanisms of their interactions. By leveraging this domain knowledge, we aim to enhance the model's ability to reason about pedestrian interactions in a more interpretable and context-aware manner, paving the way for improved prediction and navigation capabilities in crowded environments.

3. METHOD

3.1 Problem Formulation

The main objective in this paper is to predict the future trajectory positions of N pedestrians over a time horizon of T_{pred}, given their observed trajectories from $t = 1$ to $t = T_{\text{obs}}$. This task is particularly challenging in crowded environments, where pedestrian motion is governed by complex interactions and social dynamics.

Agent Dynamics: Following the work in Mohamed et al. (2020), we assume pedestrian trajectories follow a bivariate Gaussian distribution. The parameters $(\mu_t^i, \sigma_t^i, \rho_t^i)$ represent the mean, standard deviation, and correlation coefficient of the Gaussian distribution, respectively, where we parameterize mean $\mu_t^i = (\mu_x, \mu_y)_t^i$, standard deviation $\sigma_t^i = (\sigma_x, \sigma_y)_t^i$, and correlation coefficient $\rho_t^i = (\rho_x, \rho_y)_t^i$, for $i = 1, \ldots, N$. The pedestrian trajectories are given by $p_t^n = (x_t^i, y_t^i) \sim \mathcal{N}(\mu_t^i, \sigma_t^i, \rho_t^i)$, where (x_t^i, y_t^i) are random variables capturing the probability distribution of pedestrian i's position at time t in the 2D space. The model is trained by minimizing the negative log-likelihood loss function:

$$J^i(W) = -\sum_{t=1+T_{\text{obs}}}^{T_{\text{pred}}} \log(\mathbb{P}(p_t^n | \hat{\mu}_t^i, \hat{\sigma}_t^i, \hat{\rho}_t^i)), \tag{1}$$

Here, $\hat{\mu}_t^i, \hat{\sigma}_t^i, \hat{\rho}_t^i$ are the estimated parameters of the predicted trajectory distribution, $\hat{p}_t^i$.

Graph Representation of Crowd Interaction: We employ a graph representation rooted in algebraic graph theory to model pedestrian interactions within crowd Mesbahi and Egerstedt (2010). Consider $G_t = (\mathcal{V}_t, E_t)$ as a spatial graph at time t, comprising N nodes representing the relative positions of pedestrians in the scene at time t. The node set is $\mathcal{V}_t = \{\nu_t^1, \ldots, \nu_t^N\}$, and edge set $E = \{e_t^{ij} = (\nu_t^i, \nu_t^j)\} \subseteq \mathcal{V}_t \times \mathcal{V}_t$ captures the interactions among agents. The adjacency matrix is defined as $A_t = [a_t^{ij}]$, determining the neighborhood of interaction, where $a_t^{ij} = w_t^{ij}$ if ν_t^i and ν_t^j are connected (agent i can receive information from agent j); otherwise, $a_{ij} = 0$. The interaction neighborhood will be defined in more details later. The weights w_t^{ij} quantify how strongly agents influence each other's movements and are determined through geometric kernel functions inspired by psychological studies on human crowd behaviors. The graph Laplacian is defined as $L_t = D_t - A_t$, where D_t denotes the diagonal matrices of node degrees, taken as the sum of the weights for each neighbor: $D_t = [d_t^{ij}] = \sum_{j=1}^{N} a_t^{ij}$.

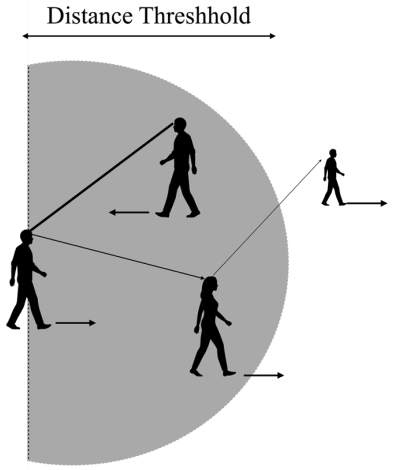

Fig. 1. A visual representation of the neighborhood of interaction for a pedestrian (shown in black), incorporating a 180-degree field of view and a 5-meter distance threshold, selected based on empirical observations from psychological studies as described in Rio et al. (2018). Pedestrians within the shaded region, which extends up to 90 degrees on either side of the walking direction, are considered part of the neighborhood. The thickness of the connecting lines decreases with increasing distance, indicating that closer neighbors have a stronger influence on the pedestrian's motion (corresponding to Graph 2 neighborhood definition with distance-based weighting according to our kernel definitions)

3.2 Topology of Interaction

Diverging from the approach of social-STGCNN Mohamed et al. (2020), which assumes a complete graph to establish the neighborhood set for each agent and weighs the importance of neighboring agents based solely on their distance, this work takes a more nuanced perspective. Our goal is to experiment with different neighborhood sets inspired by psychological literature on human interactions in crowds, as described in Rio et al. (2018), aiming to enhance model performance in predicting future pedestrian motion. This exploration seeks to identify the crucial factors in accurately modeling human interactions across various datasets.

Neighborhood of Interaction:

According to Grasso et al. (1998), people have a 180° horizontal field of view and face toward their walking direction ($\Delta p_t^i = p_{t+1}^i - p_t^i$). Therefore, interactions are expected to be influenced by neighbors within their field of view ($\pm 90°$) who are approaching them. Empirical experiments on human crowds by Rio et al. (2018) reveal that the impact of neighbors diminishes exponentially to zero at a distance of 4–5m. The following sections explain how we incorporate these key points in constructing our weighted adjacency matrix.

(1) ***Graph 1 (View):***
To account for the 180° field of view, we introduce the following kernel-based weight. It stipulates that if the angle between walking direction of pedestrian i and j is smaller than 90°, they mutually influence each other; otherwise, the weight is set to zero.

$$w_{view,t}^{ij} = \begin{cases} K_{ij} & \Delta p_t^i \cdot \Delta p_t^j > 0 \\ 0 & \text{o.w.} \end{cases}. \tag{2}$$

K_{ij} will be defined shortly.

(2) ***Graph 2 (View, Thresh):***
In a different scenario, we incorporate a threshold of $\epsilon = 5m$ distance, selected based on empirical observations from psychological studies according to Rio et al. (2018) to ensure that only nearby pedestrians within the field of view influence each other (Fig. 1):

$$w_{view,thresh,t}^{ij} = \begin{cases} K_{ij} & \Delta p_t^i \cdot \Delta p_t^j > 0, d_t^{ij} < \epsilon \\ 0 & \text{o.w.} \end{cases} \tag{3}$$

where $d_t^{ij} = \|p_t^i - p_t^j\|$ is the distance between agent i and j.

(3) ***Graph 3 (Approach):***
In this scenario, when the distance between agents decreases, signifying that they are approaching each other and likely to influence each other, we define the kernel-based weight as:

$$w_{appr,t}^{ij} = \begin{cases} K_{ij} & d_{t+1}^{ij} > d_t^{ij} \\ 0 & \text{o.w.} \end{cases} \tag{4}$$

(4) ***Graph 4 (View, Approach):*** We also consider a case that encompasses both of the above scenarios:

$$w_{view,appr,t}^{ij} = \begin{cases} K_{ij} & \Delta p_t^i \cdot \Delta p_t^j > 0, d_{t+1}^{ij} > d_t^{ij} \\ 0 & \text{o.w.} \end{cases} \tag{5}$$

It is worth noting that Yang et al. (2023) also considers the field of view and direction of movement. However, they consider each of the factors in separate graphs, and then use a multi-layer perceptron (MLP) to combine them into a single graph. In this work, we choose to combine the two conditions into a single graph to follow psychological studies and better explain the adjacancy matrix.

Kernel Functions for Assigning Neighbors Importance Through Edge Weights:

We employ kernel functions to capture the strength of interaction between pedestrians. We examine the impact of two distinct kernel functions on the model's accuracy in predicting pedestrian trajectories.

(1) **Kernel 1 (Inverse of Norm):** Similar to Mohamed et al. (2020), we use the inverse of the L_2 norm as a similarity measure, considering that further away neighbors exert less influence:

$$K_{ij} = \begin{cases} \dfrac{1}{\|p_t^i - p_t^j\|} & \|p_t^i - p_t^j\| \neq 0 \\ 0 & o.w. \end{cases}$$

(2) Alternatively, in this case, we use the exponential decay as the kernel function:

$$K_{ij} = \begin{cases} \exp(-\|p_t^i - p_t^j\|) & \|p_t^i - p_t^j\| \neq 0 \\ 0 & o.w. \end{cases}$$

to reflect the fact that the weight decays as agents move away from each other, according to Rio et al. (2018).

3.3 ***Model Architecture:***

Our model leverages the same network architecture as introduced in Mohamed et al. (2020), consisting of spatial

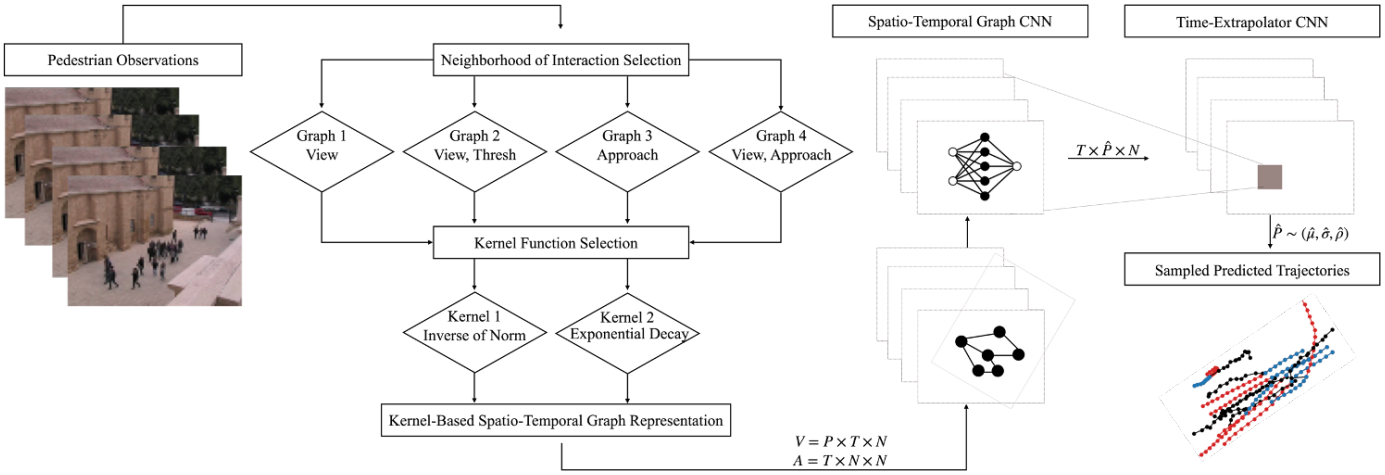

Fig. 2. Illustration of our geometric GNN model. From observed pedestrian trajectories over T frames, we construct a kernel-based spatio-temporal graph by defining neighborhood relationships and weighting edges using kernel functions. This graph is processed by a spatio-temporal graph CNN, similar to the architecture proposed by Mohamed et al. (2020), to generate a spatio-temporal embedding. A temporal CNN then predicts the distribution of future trajectories. V represents the set of nodes, $P = 2$ is the dimension of pedestrian positions, and $\hat{P}$ denotes the dimensions of the embedding from the predicted Gaussian distribution. Our approach integrates domain knowledge from psychological studies, resulting in a more context-aware and interpretable adjacency matrix.

and temporal components. The key difference from previous work is in the graph representation, which incorporates geometric kernels and neighborhood definitions inspired by psychological studies on crowd dynamics, as explained earlier. Fig. 2 provides a visual overview of the network architecture adopted in this paper.

(1) **Kernel-based Spatio-Temporal Graph Representation:** The first step in our network is to construct the graph representation of pedestrian interactions. We utilize the proposed neighborhood of interactions and kernel functions one by one and report the modeling accuracy for each of them.

(2) **Spatio-Temporal Graph CNN:**
In the second step, to extract meaningful features from the graph representation, we employ spatio-temporal graph convolutions, as described by Yan et al. (2018). This approach is particularly suitable for our problem as it effectively captures both spatial and temporal dependencies in pedestrian trajectories. The colvolution operation compute node embeddings by aggregating information from neighboring nodes in a recursive fashion as follows:

$$f(V^{(l)}, A) = \sigma\left(\mathcal{L}_t V^{(l)} W^{(l)}\right), \tag{6}$$

where $V^{(l)}$ is the stack of all node values at time t and layer l, and $W^{(l)}$ is the matrix of trainable parameters at layer l. $\mathcal{L}_t = D_t^{-1/2} L_t D_t^{-1/2}$ is the normalized graph Laplacian matrix at time t, as explained in Kipf and Welling (2016). Finally, σ is the activation function.

(3) **Time-Extrapolator CNN:**
Taking the features from the previous step as input, this network applies temporal convolutions to predict probability distributions of future trajectories for all pedestrians within the time interval from $t = T_{\text{obs}+1}$ to $t = T_{\text{pred}}$.

3.4 Datasets and evaluation metrics

To train and assess the performance of our model, we utilize two well-established pedestrian datasets, in line with established methodologies in prior research. These datasets include: (1) ETH Dataset, collected by Pellegrini et al. (2009), with two scenes: ETH and Hotel, and (2) UCY Dataset, given by Lerner et al. (2007): with 3 scenarios: ZARA-01, ZARA-02, and UNIV. In total, our model is evaluated across five distinct datasets, encompassing diverse nonlinear trajectories from 1536 pedestrians within real-world crowded environments. To ensure consistency in data representation across various datasets, we adopt the preprocessing approach outlined in Social-GAN by Gupta et al. (2018). To facilitate fair comparisons, we employ a leave-one-out training methodology, as suggested Alahi et al. (2016). This involves training and validating on four datasets and subsequently testing on the fifth. This procedure is consistently applied across the **four** interaction neighborhood scenarios and the **two** proposed kernel functions as well as other baseline methods for best evaluation.

For the quantitative assessment of our approach, we employ two widely recognized metrics Alahi et al. (2016):

Average Displacement Error (ADE) computes the average L2 distance between the ground truth and our model's predictions over all predicted time steps. It serves as a measure of trajectory accuracy along the path.

Final Displacement Error (FDE) quantifies the distance between the predicted final position and the ground truth final position at the end of the prediction period T_{pred}.

Given our assumption of Gaussian distribution for pedestrian trajectories, the estimated trajectory is a distribution. For fair comparison with previous work, we follow Mohamed et al. (2020)'s evaluation approach. We generate 20 samples from the predicted trajectory distribution and

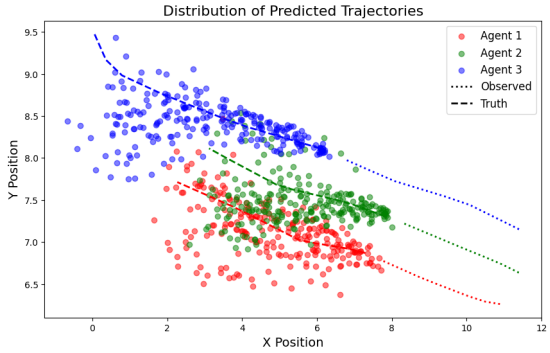

Fig. 3. Visualizing the predicted distribution of pedestrian trajectories for a representative scene (UNIV). Sampled trajectories from the predicted Gaussian distribution are depicted as dot points in the scatter plot, while observed and true trajectories are shown as dashed and dotted lines respectively.

compare the best-performing sample (closest to the ground truth) against the ground truth to calculate ADE and FDE. This evaluation approach has been widely adopted by previous studies, including those by Gupta et al. (2018); Huang et al. (2019).

Training setup: Our training setup closely aligns with the architecture and parameters outlined in Mohamed et al. (2020), ensuring a consistent baseline for evaluation. We employ the Parametric Rectified Linear Unit (PReLU) as our activation function, denoted by σ. The training batch size is set to 128 and the model is trained for a total of 250 epochs for each dataset. The learning rate is initialized at 0.01 and adjusted to 0.002 after 150 epochs to enhance the model's adaptability over the training duration. Stochastic Gradient Descent (SGD) is employed for optimization, in line with the methodology proposed by Mohamed et al. (2020). Consistent with the referenced work, our model comprises 1 layer of Spatio-Temporal Graph CNN and 5 layers of Temporal-Extrapolator CNN. This allows evaluating the impact of kernel functions and interaction topologies.

4. RESULTS

The results presented in Table 1 show that our proposed methods, incorporating different neighborhood definitions and kernel functions, achieve competitive performance compared to previous approaches across various pedestrian datasets. Our model has $7,563$ parameters, with an average inference time of 0.002 seconds, based on multiple single inference steps.

On the ETH dataset, our method with Graph 1 (using pedestrian's field of view in neighborhood definition) and Kernel 1 (inverse of the norm) achieves the best performance, following the Social-STGCNN ADE of 0.64 and improving the FDE to 1.09. For the HOTEL dataset, our method with Graph 3 (using approach dynamics in neighborhood definition) and Kernel 2 (exponential decay) significantly outperforms all existing methods, achieving the best ADE of 0.34 and FDE of 0.47. On the UNIV dataset, the previous Social-STGCNN method by Mohamed et al. (2020) maintains the best performance with an ADE of 0.44 and FDE of 0.79. However, our method with Graph 2 (using field of view and $5m$ distance threshold for neighborhood definition) and Kernel 2 closely follows the best performance with an ADE of 0.45 and FDE of 0.82. Other graph variations and kernel functions also show very close performance on this dataset, further highlighting the importance of using geometric definitions. For the ZARA01 dataset, the Social-BiGAT method by Kosaraju et al. (2019) achieves the best ADE of 0.30, while our method with Graph 2 and Kernel 1 achieves the best FDE of 0.52. On the ZARA02 dataset, our method with Graph 2 and Kernel 2 outperforms all existing methods, achieving the best ADE of 0.27 and FDE of 0.45.

Overall, our method with Graph 1 and Kernel 1 achieves the best average ADE of 0.43 and FDE of 0.73 across all datasets, surpassing previous methods. Our results suggest that methods incorporating attention mechanisms, whether through softmax (e.g., social-BiGAT by Kosaraju et al. (2019)) or geometric kernel functions, consistently enhance accuracy, underscoring the importance of kernel functions for graph representations in GNN methods. Furthermore, our results highlight the effectiveness of integrating domain knowledge, such as pedestrian's field of view, $5m$ distance threshold, and approach dynamics, into neighborhood definitions for modeling human interactions in crowded environments. While Mohamed et al. (2020) found no improvement with the exponential decay kernel, our results suggest that performance outcomes can vary and improve depending on the interaction neighborhood definition.

While no single method consistently outperforms others across all datasets, our proposed methods exhibit competitive performance. Performance discrepancies may arise from unaccounted factors like landmarks or scene information, highlighting the potential impact of external factors. Notably, Lv et al. (2023) demonstrated significant accuracy improvements by integrating scene information into their graph attention network.

Our primary objective was to demonstrate how incorporating domain knowledge can enhance data-driven trajectory prediction, rather than finding the best-performing model. Thus, we maintained the same parameters as the referenced work. Additionally, our model estimates uncertainty using the covariance matrix in the bivariate Gaussian distribution, capturing the inherent uncertainty in predictions. Future work can explore Bayesian approaches and the impact of data and domain uncertainties on model performance.

4.1 Conclusion

In summary, this case study introduces a geometric GNN architecture featuring various kernel functions and interaction topologies inspired by psychological studies. This framework lays a foundation for better understanding and modeling human interactions. Integrating a 180° field of view in the interaction neighborhood and employing the inverse of norm kernel function to weigh neighboring agents' importance contribute to overall accuracy improvements, reflected in enhanced ADE and FDE metrics across all datasets. Our results suggest that different topologies of interaction and kernel functions may perform better

Table 1. ADE/FDE Results for Various Methods across 5 Pedestrian Datasets

Method	ETH	HOTEL	UNIV	ZARA01	ZARA02	Average
social-LSTM (Alahi et al. (2016))	1.09/2.35	0.79/1.76	0.67/1.40	0.47/1.00	0.56/1.17	0.72/1.54
social-GAN (Gupta et al. (2018))	0.81/1.52	0.72/1.61	0.60/1.26	0.34/0.69	0.42/0.84	0.58/1.18
Social-BiGAT (Kosaraju et al. (2019))	0.69/1.29	0.49/1.01	0.55/1.32	**0.30**/0.62	0.36/0.75	0.48/1.00
social-STGCNN (Mohamed et al. (2020))	**0.64**/1.11	0.49/0.85	**0.44**/**0.79**	0.34/0.53	0.30/0.48	0.44/0.75
Ours (Kernel 1: Inverse of Norm)						
- View (Graph 1- Kernel 1)	**0.64**/**1.09**	0.41/0.63	**0.45**/0.85	0.35/0.59	0.30/0.5	**0.43**/**0.73**
- View, thresh (Graph 2- Kernel 1)	0.76/1.48	0.40/0.69	**0.45**/0.85	0.33/**0.52**	0.29/0.47	0.44/0.80
- Approach (Graph 3- Kernel 1)	0.82/1.29	0.39/0.58	0.46/0.868	0.34/0.53	0.30/0.50	0.46/0.75
- View, approach (Graph 4- Kernel 1)	0.75/1.37	0.38/0.57	0.47/0.88	0.34/0.57	0.30/0.50	0.44/0.77
Ours (Kernel 2: Exponential Decay)						
- View (Graph 1- Kernel 2)	0.80/1.54	0.47/0.77	0.46/0.84	0.37/0.58	0.43/0.67	0.51/0.88
- View, thresh (Graph 2- Kernel 2)	0.80/1.40	0.47/0.77	**0.45**/**0.82**	**0.32**/0.53	**0.27**/**0.45**	0.47/0.80
- Approach (Graph 3- Kernel 2)	0.86/1.37	**0.34**/**0.47**	0.47/0.86	0.36/0.56	0.28/0.46	0.47/0.74
- View, approach (Graph 4- Kernel 2)	0.76/1.36	0.47/0.83	0.48/0.88	0.33/**0.52**	0.30/0.48	0.47/0.81

under different scenarios. However, determining the optimal designs that best explain complex interactions across domains remains an open challenge.

REFERENCES

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

Grasso, R., Prévost, P., Ivanenko, Y.P., and Berthoz, A. (1998). Eye-head coordination for the steering of locomotion in humans: an anticipatory synergy. *Neuroscience Letters*, 253(2), 115–118.

Gupta, A., Johnson, J., Fei-Fei, L., Savarese, S., and Alahi, A. (2018). Social gan: Socially acceptable trajectories with generative adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2255–2264.

Helbing, D. and Molnar, P. (1995). Social force model for pedestrian dynamics. *Physical review E*, 51(5), 4282.

Huang, Y., Bi, H., Li, Z., Mao, T., and Wang, Z. (2019). Stgat: Modeling spatial-temporal interactions for human trajectory prediction. In *Proceedings of the IEEE/CVF international conference on computer vision*, 6272–6281.

Kipf, T.N. and Welling, M. (2016). Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Korbmacher, R. and Tordeux, A. (2022). Review of pedestrian trajectory prediction methods: Comparing deep learning and knowledge-based approaches. *IEEE Transactions on Intelligent Transportation Systems*.

Kosaraju, V., Sadeghian, A., Martín-Martín, R., Reid, I., Rezatofighi, H., and Savarese, S. (2019). Social-bigat: Multimodal trajectory forecasting using bicycle-gan and graph attention networks. *Advances in Neural Information Processing Systems*, 32.

Lerner, A., Chrysanthou, Y., and Lischinski, D. (2007). Crowds by example. In *Computer graphics forum*, volume 26, 655–664. Wiley Online Library.

Lv, P., Wang, W., Wang, Y., Zhang, Y., Xu, M., and Xu, C. (2023). Ssagcn: social soft attention graph convolution network for pedestrian trajectory prediction. *IEEE transactions on neural networks and learning systems*.

Mavrogiannis, C., Baldini, F., Wang, A., Zhao, D., Trautman, P., Steinfeld, A., and Oh, J. (2023). Core challenges of social robot navigation: A survey. *ACM Transactions on Human-Robot Interaction*, 12(3), 1–39.

Mesbahi, M. and Egerstedt, M. (2010). *Graph theoretic methods in multiagent networks*. Princeton University Press.

Mohamed, A., Qian, K., Elhoseiny, M., and Claudel, C. (2020). Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 14424–14432.

Pellegrini, S., Ess, A., Schindler, K., and Van Gool, L. (2009). You'll never walk alone: Modeling social behavior for multi-target tracking. In *2009 IEEE 12th international conference on computer vision*, 261–268. IEEE.

Rio, K.W., Dachner, G.C., and Warren, W.H. (2018). Local interactions underlying collective motion in human crowds. *Proceedings of the Royal Society B: Biological Sciences*, 285(1878), 20180611.

Rudenko, A., Palmieri, L., Herman, M., Kitani, K.M., Gavrila, D.M., and Arras, K.O. (2020). Human motion trajectory prediction: A survey. *The International Journal of Robotics Research*, 39(8), 895–935.

Van Den Berg, J., Guy, S.J., Lin, M., and Manocha, D. (2011). Reciprocal n-body collision avoidance. In *Robotics Research: The 14th International Symposium ISRR*, 3–19. Springer.

Wirth, T.D., Dachner, G.C., Rio, K.W., and Warren, W.H. (2022). The neighborhood of interaction in human crowds is neither metric nor topological, but visual. *bioRxiv*, 2022–08.

Yan, S., Xiong, Y., and Lin, D. (2018). Spatial temporal graph convolutional networks for skeleton-based action recognition. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.

Yang, Z., Pang, C., and Zeng, X. (2023). Trajectory forecasting using graph convolutional neural networks based on prior awareness and information fusion. *ISPRS International Journal of Geo-Information*, 12(2), 77.

Zhu, Y., Xu, W., Zhang, J., Du, Y., Zhang, J., Liu, Q., Yang, C., and Wu, S. (2021). A survey on graph structure learning: Progress and opportunities. *arXiv preprint arXiv:2103.03036*.