

Neural Interaction Energy for Multi-Agent Trajectory Prediction

Kaixin Shen, Ruijie Quan, Linchao Zhu, Jun Xiao, Yi Yang
Zhejiang University

ABSTRACT

Maintaining temporal stability is crucial in multi-agent trajectory prediction. Insufficient regularization to uphold this stability often results in fluctuations in kinematic states, leading to inconsistent predictions and the amplification of errors. In this study, we introduce a framework called **Multi-Agent Trajectory prediction via neural interaction Energy (MATE)**. This framework assesses the interactive motion of agents by employing neural interaction energy, which captures the dynamics of interactions and illustrates their influence on the future trajectories of agents. To bolster temporal stability, we introduce two constraints: inter-agent interaction constraint and intra-agent motion constraint. These constraints work together to ensure temporal stability at both the system and agent levels, effectively mitigating prediction fluctuations inherent in multi-agent systems. Comparative evaluations against previous methods on four diverse datasets highlight the superior prediction accuracy and generalization capabilities of our model.

KEYWORDS

Trajectory Prediction, Multi-agent System, Partial Differential Equations

1 INTRODUCTION

Multi-agent trajectory prediction is a fundamental research task with many applications such as autonomous driving [23], interactive robotics [18], particle simulation [24], and team sports [13]. It refers to predicting the future positions of multiple agents in a dynamic environment simultaneously, based on their historical positions and agent interactions. This task is critical in many real-world scenarios where a group of agents interacts with each other, giving rise to complicated behavior patterns at the level of both individuals and the whole system. For this reason, it is more challenging to simultaneously predict the future positions of multiple agents compared to single-agent trajectory prediction [25].

One of the major obstacles in multi-agent trajectory prediction lies in the frequent fluctuations of kinematic states across consecutive time intervals. These fluctuations pose a significant challenge as they can lead to inconsistencies in the trajectory predictions for each agent. The variability in kinematic states, such as sudden changes in position or velocity, further causes error amplification in sequential predictions [1, 39, 46]. Such fluctuations can be mainly attributed to the absence of proper regularization to temporal dynamics in multi-agent systems.

Motivated by this, we aim to regularize the multi-agent trajectory prediction systems to be temporally stable, enhancing the prediction robustness across diverse scenarios. Temporal stability refers to the property that the system’s kinematic states do not change abruptly or inconsistently over time, but rather approach a steady-state distribution [5]. Temporal stability offers two advantages for trajectory prediction in multi-agent systems: 1) Improved

prediction accuracy. By enforcing temporal stability in the multi-agent systems, the prediction results would be more resilient to observation noises and error propagation. 2) Enhanced adaptability to diverse environments. The presence of temporal stability in multi-agent systems results in more uniform and coordinated movement patterns, enabling the system to adapt seamlessly to a variety of environmental conditions.

In this paper, we propose a framework named MATE for multi-agent trajectory prediction. We enhance the temporal stability of the multi-agent system with an inter-agent constraint and an intra-agent constraint. These two constraints help to preserve temporal stability at the system level and agent level, respectively.

First, we introduce the **inter-agent interaction constraint** to enhance the stability of agent interactions at the system level. We represent the interactions between agents as the changes of their **neural interaction energy**, while the system-level dynamics can be considered as the aggregation of neural interaction energy from all agents. Specifically, the neural interaction energy quantifies agents’ motion interactiveness at each step. It also determines the action of agents in future steps. The agents with low neural interaction energy may exhibit reduced movement speed, whereas those with high energy are prone to be more dynamic. The inter-agent interaction constraint minimizes the change of system-level neural interaction energy, stabilizing the multi-agent movement patterns and ensuring a more coherent evolution of the multi-agent system over time.

Second, we introduce the **intra-agent motion constraint** to enforce the agent-level stability. For each agent, we leverage a temporal motion variance term to measure the consistency of their movement patterns over time. We derive an approximated motion based on the agent’s past kinematic states and its interactions with surrounding agents. The temporal motion variance term quantifies the discrepancy between the approximated motion and the predicted motion, which helps to identify the agent movement inconsistency. By minimizing this variance, it ensures that the predicted motion aligns closely with the approximated motion, restricting the kinematic states of agents to be coherent and may vary within a plausible range.

Our model outperforms the state-of-the-art methods in multi-agent trajectory prediction across four datasets, i.e., PHASE [30], Socialnav [3], Charged [25], and NBA datasets. We further validate the advantages of our method in zero-shot generalization to unseen scenarios. In summary, the main contributions of this paper are threefold:

- We propose to use neural interaction energy to model the interactions between agents, providing an effective framework to capture the complexities of agent interactions.
- We establish an inter-agent interaction constraint and an intra-agent motion constraint, which well preserve the temporal stability at the system level and agent level.

- Extensive experiments demonstrate that our model accurately predicts future trajectories and improves the generalization abilities in unseen scenarios.

2 RELATED WORK

Trajectory Prediction. Trajectory prediction approaches fall into model-based [14] and model-free methods [1, 2, 8–10, 12, 15, 21, 24, 25, 27, 28, 33, 36, 38, 40, 41, 45, 46]. Before the deep learning era, previous works employ various physical models to conduct trajectory prediction. Social force [14] and energy [19] are commonly used to model pedestrian motion. Some works [42] exploit the reciprocal velocity obstacles (RVO) model for trajectory prediction. [44] predicts transitions in pedestrians by applying spatiotemporal graphs, where nodes and edges are represented by RNNs. [32] computes joint motion predictions based on the time of possible collision between pairs of agents. [48] propose the Multi-Agent Tensor Fusion encoding, which fuses contextual images of the environment with sequential trajectories of agents. We will include the above discussion and references in the revision. The rise of deep learning [3, 6, 12, 31, 34] has introduced model-free methods into this field. Such a method shifts the focus of modeling agents’ trajectories to fitting the distribution of data [9, 13, 16, 29, 47], which improves computational efficiency while reducing the model interpretability. GAT-LSTM [43] employs graph attention layers along with LSTMs for prediction. NRI [22] designs a VAE model with recurrent GNN networks to encode and predict trajectories. EvolveGraph [24] and fNRI [45] expand the framework of NRI and introduce additional modules to improve performance. RFM [40] uses a recurrent GNN model with a supervised loss to realize trajectory prediction. IMMA [39] introduces a multiplex graph and proposes a progressive layer training strategy. GRIN [25] employs generative models to learn the distribution of trajectories. Inspired by traditional dynamics, we combine a novel concept named neural interaction energy and sequential models to predict the changes in agents’ trajectories and introduce an inter-agent interaction constraint and an intra-agent motion constraint for preserving temporal stability.

Differential Equations Informed Neural Network. Recently the combination of deep learning and differentiable equations has earned significant interest. Related research [20, 37] can be categorized into different subfields such as deep learning assisted DE [7], differentiable physics, neural differential equations [4], and physics-informed neural networks (PINNs) [26, 35, 49]. PINNs are based on the physical prior that constrains the output by calculating the means of a partial differential equation (PDE) system with the assistance of a neural network. Inspired by such research, we propose novel neural differential equations for modeling interactions and establishing motion constraints for agents to predict coherent future trajectories.

3 METHODOLOGY

3.1 Problem Formulation.

Given an observed trajectory $\mathbf{X}_i^T = \{\mathbf{x}_i^1, \mathbf{x}_i^2, \dots, \mathbf{x}_i^{T_{obs}}\}$ of an agent $i \in N$ across period T_{obs} , our goal is to predict the future trajectories of all agents simultaneously. $\mathbf{x}_i^t \in \mathbb{R}^2$ is the 2D coordinate position of the agent i at time step t . We denote the future trajectory across T_{pred} steps as $\mathbf{Y}_i^T = \{\mathbf{x}_i^{T_{obs}+1}, \mathbf{x}_i^{T_{obs}+2}, \dots, \mathbf{x}_i^T\}$, where $T = T_{obs} +$

T_{pred} . Note that for each agent, we only predict a single future trajectory across T_{pred} steps. In this paper, kinematic states include position and velocity. At each step, the position $\mathbf{x}$ of an agent indicates its location in the 2D coordinate system. In addition, to represent movement speed, we calculate the velocity $\mathbf{v}$ of agents at each step based on their position and time interval Δt .

3.2 Inter-agent Interaction Modeling

Predicting future trajectories in multi-agent systems necessitates an accurate modeling of agent interactions. Existing approaches often rely on graph neural networks (GNNs) to depict these interactions as latent features of graph edges. However, this representation ignores ensuring the stability of agent interactions. Also, it does not focus on promising a temporally coherent evolution of the system. To address these overlooked aspects, we employ a neural interaction energy mechanism to model the interactions of agents, further preserving the temporal stability at the system level.

Neural Interaction Energy E . We first present the neural interaction energy to quantify the agent’s motion interactiveness, thereby capturing the dynamics of multi-agent interactions and illustrating how interactions influence agents’ future trajectories.

Neural interaction energy quantifies the motion interactiveness at each step and affects the agent movement in future steps. This motion interactiveness is a function concerning both an agent’s kinematic states (position and velocity) and its interaction representation. For clarity, we use E_i^t to denote the neural interaction energy scalar for agent i at step t and use $\mathbf{E}_i^t$ to denote the corresponding neural interaction energy features in the neural networks.

Intuitively, the reduction of neural interaction energy will decrease the agent interactiveness, while agents with high neural interaction energy are more likely to move. In a multi-agent system, the motion of the target agent is affected by other agents. The change in neural interaction energy, reflecting each agent’s interactions with others, acts as a measure of the dynamic exchanges within the system. Moreover, the overall dynamics of the multi-agent system can be viewed as the aggregation of the neural interaction energies of all participating agents, providing a comprehensive view of the system’s interactive behavior.

Inter-agent Interaction Constraint. We introduce an inter-agent interaction constraint to preserve system-level temporal stability based on the interaction between agents. Considering a multi-agent system, agents are moving concurrently and their behaviors are mutually influenced through complex social interactions. The social interactions among agents indicate the transfer of neural interaction energy, accelerating or decelerating agents. However, due to the complexity of interactions between agents in systems, the neural interaction energy transfer process becomes difficult to analyze from the perspective of the agent level.

Nevertheless, we model the interactions at the system level. We aim to preserve system-level temporal stability. It means the changes in its neural interaction energy maintain temporal continuity. The target formulation for system-level temporal stability is,

$$E_i^{t+\Delta t} - E_i^t \rightarrow 0, \quad \text{when } \Delta t \rightarrow 0. \quad (1)$$

Therefore, over a short time interval, the aggregated neural interaction energy of the system should experience negligible changes,

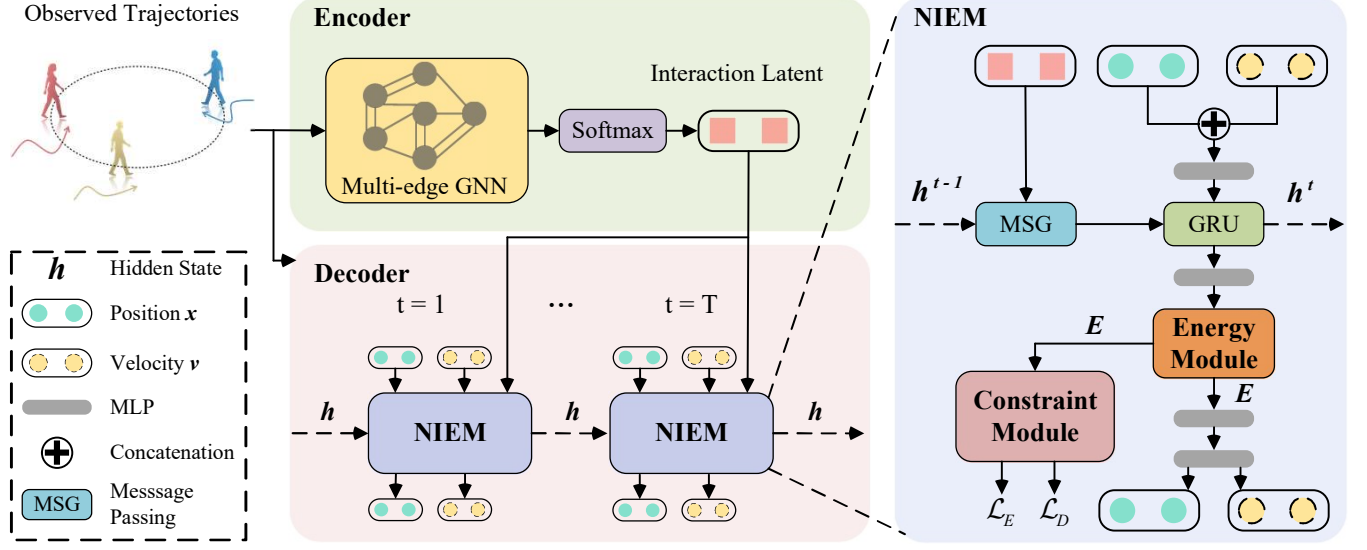


Figure 1: The MATE model consists of (i) a MATE-Encoder responsible for extracting interaction latent between agents, and (ii) a MATE-Decoder integrated with a Neural Interaction Energy Module (NIEM). The NIEM incorporates an Energy Module that extracts the neural interaction energy features E of agents. These features are used for the subsequent networks. They are also leveraged to calculate the inter-agent interaction and intra-agent motion constraints through a Constraint Module. Refer to the Appendix for more details about the NIEM.

resulting in a nearly constant sum of neural interaction energy changes of agents. That is,

$$\Delta \sum_{i=1}^N E_i = \sum_{i=1}^N E_i^{t+\Delta t} - \sum_{i=1}^N E_i^t \rightarrow 0, \quad \text{when } \Delta t \rightarrow 0. \quad (2)$$

Note that the magnitude of change in neural interaction energy differs among different types of agents, necessitating the assignment of weights to achieve normalization. Considering that different types of agents have different ranges of positional changes within the same time interval, we employ Δx as the normalization weight. We have

$$\sum_{i=1}^N \frac{\Delta E_i}{\Delta x_i} \rightarrow 0, \quad \text{when } \Delta t \rightarrow 0. \quad (3)$$

ΔE_i and Δx_i denote the change in neural interaction energy and the change in position of agent i , respectively.

Consequently, we can model the interactions between agents while preserving the temporal stability at the system level. Formally, the system would have minimal neural interaction energy change in its aggregated interactions of all agents over a short period. Since the time interval is short, Δx tends to be zero. Not that we use the derivative of the neural interaction energy features concerning the position $\nabla_x E_i(x)$ in implementation to replace the expression in Eq. (3). That is

$$\sum_{i=1}^N \nabla_x E_i(x) = \sum_{i=1}^N \lim_{\Delta x \rightarrow 0} \frac{\Delta E_i}{\Delta x_i} \rightarrow 0. \quad (4)$$

We omit the superscript t for simplification.

In summary, we leverage Eq. (4) to ensure the temporal stability of a multi-agent system by minimizing the aggregated changes

in neural interaction energy over a short period. Therefore, by incorporating this inter-agent interaction constraint into model optimization, we preserve system-level temporal stability. We add this constraint in the form like

$$\mathcal{L}_E = \frac{1}{N} \frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T \|\nabla_x E_i^t(x)\|_2. \quad (5)$$

3.3 Intra-agent Motion Modeling

This section introduces to improve intra-agent motion coherence and avoid catastrophic erroneous predictions, preserving agent-level temporal stability. If the value of the target agent's kinematic states deviates from the normal range (e.g., a pedestrian moves at the speed of a car), it indicates unstable changes and a potential prediction error. We aim to mitigate the prediction discrepancy at each step. Technically, we define a temporal motion variance term and regularize this variance to be coherent. This helps to restrict the value of the agent's kinematic states to vary within a plausible range.

The kinematic states of the target agent in the current time step (x^t, v^t) can be determined by the kinematic states from the previous time step and the neural interaction energy. We calculate the derivative of the predicted position x^t and the change in neural interaction energy $\nabla_x E^t(x)$ concerning the previous step velocity v^{t-1} . We formulate the following temporal motion variance u for the target agent,

$$u^t = \gamma \cdot \frac{\partial \nabla_x E^t(x)}{\partial v^{t-1}} + \beta \cdot \frac{\partial x^t}{\partial v^{t-1}} + \alpha, \quad (6)$$

where γ and β are scaling scalars and α is the bias term representing a constant correlation between the position and velocity of agents.

For clarity, we omit the subscripts i of the variables as agents share the same formula. Within a multi-agent environment, an agent’s position is not solely determined by its own intention but is also influenced by the interactions with other agents. Consequently, in addition to $\frac{\partial \mathbf{x}^t}{\partial \mathbf{v}^{t-1}}$, our motion variance u introduces an additional term to encapsulate external influences on the agent. Hyperparameters γ and β are introduced to weigh the significance of these two terms, catering to scenarios where agent behaviors vary based on differing extents of external influences. This refined approach renders our proposed constraints more adaptable for trajectory prediction within multi-agent systems.

In the temporal motion variance u , we use the derivative of the predicted output ($\nabla_{\mathbf{x}} E^t(\mathbf{x})$ and $\mathbf{x}^t$) concerning $\mathbf{v}^{t-1}$ to measure the prediction’s sensitivity. A high sensitivity means that a minor miscalculation in estimating $\mathbf{v}^{t-1}$ can magnify into a prominent prediction error. The detailed mathematical derivation of Eq. (6) and determination of hyperparameters γ , β , and α can be found in the Appendix.

By minimizing the temporal motion variance u :

$$\mathcal{L}_D = \frac{1}{N} \frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T \|u_i^t\|_2, \quad (7)$$

we reduce the prediction sensitivity and penalize inaccurate or incoherent predictions, achieving agent-level temporal stability. We add this intra-agent motion constraint to model optimization.

3.4 Model Architecture

MATE-Encoder with Multi-Edge Graph. Previous research employs a single-edge GNN to represent interactions between agents. In such a framework, the interaction latent $\mathbf{z}$ produced by the GNN is normalized and then utilized as a one-hot vector to symbolize interactions for proceeding prediction. However, multiple types of interactions frequently coexist between agents in real scenarios, and a simple one-hot vector fails to encapsulate the intricate properties of these interactions. For instance, in social event systems, factors like familiarity, status, and identities concurrently influence agents’ interactions with others.

Addressing this, as shown in Fig. 1, we present a multi-edge GNN, where multiple edges are assigned between agents. Each edge symbolizes a distinct type of interaction and is separately encoded through a unique Multi-Layer Perceptron (MLP) layer. We hypothesize that all types of interactions influence the agents’ neural interaction energy and their future trajectories. This fosters a richer representation of the interactions in a given multi-agent system.

Mathematically, given the observation of all agents $\mathbf{X}^{1:T_{obs}}$, we employ a GNN model to learn the interaction latent as $\mathbf{z}_{ij} = \text{GNN}(\mathbf{X}_i, \mathbf{X}_j, \theta)$. Here $\mathbf{X}_i$ and $\mathbf{X}_j$ represent the observed trajectory of agent i and j respectively, θ denotes model parameters, and $\mathbf{z}_{ij}$ represents the interaction latent between agent i and agent j . The interaction latent $\mathbf{z}_{ij}$ has K unnormalized dimensions, each indicating a type of interaction. We assume that agents can maintain diverse interactions and that the neighboring agents of the target agent have different importance considering each type of interaction. Consequently, instead of normalizing the dimensions

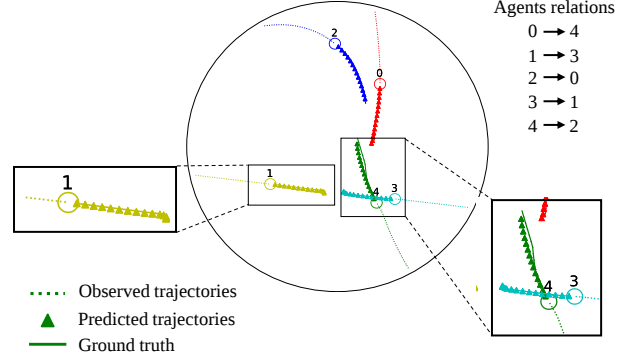


Figure 2: Qualitative analysis of our model on Socialnav dataset. Agent relations mean that the agent to the left of the arrow moves towards the agent to the right of the arrow.

of interaction latent $\mathbf{z}_{ij}$, we normalize the same dimension of interaction latent between agent i and all its neighbors j as:

$$z_{ij}^k = \frac{\exp(z_{ij}^k)}{\sum_{j \in N_i} \exp(z_{ij}^k)}, \quad \forall k \in K. \quad (8)$$

N_i denotes the neighbor set of agent i , and K represents the number of dimensions of the latent code $\mathbf{z}_{ij}$, we take the $\mathbf{z} = \{\mathbf{z}_{ij} \mid 1 \leq i, j \leq N\}$ as the output of the MATE-Encoder.

MATE-Decoder with Neural Interaction Energy Module. To optimize the prediction of the future trajectory, our MATE-Decoder incorporates a sequential model, incorporating our proposed constraints. Specifically in the MATE-Decoder, we introduce a Neural Interaction Energy Module (NIEM) unit (depicted in Fig. 1) that processes the previous position $\mathbf{x}$, velocity $\mathbf{v}$, interaction latent $\mathbf{z}$, and previous hidden state $\mathbf{h}$, yielding the new position, velocity, and hidden state for the predicted step. To enhance the performance of the MATE-Decoder, we adopt a burn-in strategy [22] that initiates the decoding process from the first observation step during training, thereby prolonging the prediction horizon and enhancing model robustness.

Structure of the NIEM. Initially, a message passing (MSG) block is leveraged to integrate the information between the hidden state $\mathbf{h}$ and interaction latent $\mathbf{z}$. This is followed by the processing of the embedding from the previous step’s kinematic states along with the output from the MSG block through a GRU module. The GRU module results in the generation of an updated hidden state. Subsequently, an Energy Module calculates the neural interaction energy features $\mathbf{E}$ for each agent. These features serve as inputs for subsequent networks, facilitating the generation of agent trajectories. Additionally, a Constraint Module exploits these neural interaction energy features to formulate the inter-agent interaction constraint $\mathcal{L}_E$ and the intra-agent motion constraint $\mathcal{L}_D$. It is noteworthy that, in our implementation, the inter-agent interaction constraint is integrated into the trajectory prediction through two distinct methods. For each dataset, the method yielding the most favorable result is selected for final result presentation. Refer to the Appendix for more details about the NIEM.

Model	PHASE			Socialnav		
	ADE↓	FDE↓	Graph.%↑	ADE↓	FDE↓	Graph.%↑
MLP	1.024	1.763	-	0.241	0.513	-
GAT-LSTM [43]	1.545	2.527	52.94	0.306	0.527	21.83
NRI [22]	0.986	1.772	55.30	0.217	0.386	57.18
EvolveGraph [24]	0.848	1.522	58.96	0.160	0.321	70.23
RFM(skip 1) [40]	0.870	1.581	55.30	0.160	0.325	70.05
RFM [40]	0.892	1.630	54.71	0.156	0.317	71.53
fNRI [45]	0.883	1.607	55.49	0.151	0.308	33.97
IMMA [39]	0.801	1.484	79.21	0.139	0.279	81.38
Ours	0.660	1.281	91.18	0.111	0.229	91.21

Table 1: Quantitative results of trajectory prediction on PHASE and Socialnav datasets.

3.5 Loss for Training

We employ an end-to-end training method. In our approach, we compute the mean squared error (MSE), denoted as $\mathcal{L}_P$, between the predicted position and their corresponding ground truth across period T . That is,

$$\mathcal{L}_P = \frac{1}{N} \frac{1}{T} \sum_{i=1}^N \sum_{t=1}^T \|\mathbf{x}_i^t - \mathbf{d}_i^t\|_2, \quad (9)$$

where $\mathbf{d}_i^t$ means ground truth position for agent i at time step t . Moreover, we integrate the proposed constraints $\mathcal{L}_E$ and $\mathcal{L}_D$ into our optimization. Above all, the loss function for our framework is shown below and λ_1, λ_2 are hyperparameters for scale balance,

$$\mathcal{L} = \mathcal{L}_P + \lambda_1 \mathcal{L}_E + \lambda_2 \mathcal{L}_D. \quad (10)$$

4 EXPERIMENTS

4.1 Datasets

We validate our method on four datasets: **PHASE** [30], **Socialnav** [3], **Charged** [25], **NBA**. For simulated datasets, we set the same simulation configurations with previous work for fair Comparison. The detailed description each dataset are as follows:

PHASE. The PHASE is a dataset of physically grounded abstract social events that resemble a wide range of real-life social interactions by including social concepts such as helping another agent. For comparison with previous work, we choose the ‘‘collaboration’’ task in that two agents collaborate to move one of the balls to a pre-set position. Note that we additionally take the label information indicating the difference between agents and balls as part of the input. We do the same data augmentation on the training set as previous work, by flipping the environment and rotating it by 90 degrees, 180 degrees, and 270 degrees, resulting in a training set with 8x more instances. We use 80% for training, 10% for validation, and 10% for testing, and predict 10 steps based on the observation of 24 steps. The timestep of this dataset is 0.25.

Social Navigation Environment. The Social Navigation Environment (Socialnav) dataset includes annotations for each trajectory, and it also includes annotations for social interactions between agents, such as whether two agents are nearby or whether one agent is following another. Complex unseen trajectories can be easily generated by varying different parameters and configurations. We follow the environmental configurations of previous work. In this simulation, agents are controlled by ORCA policy. We set the

radius of agents to 0.3, the radius of the circular environment to 8, and the preferred speed of agents to 1. For Social Navigation we use 100k multi-agent trajectories, in total for training, validation, and testing. We also simulate 50k, 25k, 12.5k, and 10k samples to validate the performance of our model in different dataset sizes. We use 80% for training, 10% for validation, and 10% for testing, and predict 10 steps based on the observation of 24 steps. The timestep of this dataset is 0.25.

Charged. The Charged datasets are usually used to illustrate the performance of graph relation prediction in previous work, as the relation type between agents can be set manually. We simulate a dynamic system controlled by physical laws. There are 5 charged particles in each scene, each particle has a positive or negative charge with equal probability, attracting the particles with different charges and vice versa. We experiment with the generalization of models on the Charged dataset by modifying configurations. We generate 50k trajectories for training, 10k trajectories for validation, and 10k trajectories for testing. In the basic environment, we set the side length of the box square to 5, the number of particles to 5, and the sampling timestep to 0.2. To compare with GRIN and previous work, we use our model to predict 20 steps based on the observation of 80 steps. The timestep of this dataset is 0.2.

NBA. The National Basketball Association (NBA) uses the SportVU tracking system to collect player-tracking data, where each frame contains the location of all ten players and the ball at each time step. We use 300k multi-agent trajectories of players and the ball in total for training, validation, and testing. We use 80% for training, 10% for validation, and 10% for testing, and predict 10 steps based on the observation of 24 steps. The timestep of this dataset is 0.4.

4.2 Setup

Baselines. Following IMMA [39] and GRIN [25], we compare our results with approaches conducting multi-agent trajectory prediction. In specific, we choose MLP, GAT-LSTM [43], NRI [22], EvolveGraph [24], RFM [40], fNRI [45], IMMA, FQA [17], dNRI [11], SocialGAN [12], and GRIN as baselines in quantitative evaluation. We choose GAT-LSTM and IMMA for qualitative analysis.

Hyperparameter Choice. For all datasets, we set λ_1 to 1 and λ_2 to 0.001 in Eq. (10). λ_2 is set to a smaller value to balance the regularization term. When comparing with baselines, we only report this set of hyperparameters as they have the highest average performance across four datasets.

Evaluation Metrics. We quantitatively compare our model with other methods using the following metrics: 1) Average Displacement Error (**ADE**) is the L2 distance between the ground truth trajectories and predicted trajectories. 2) Final Displacement Error (**FDE**) measures the L2 distance between the ground truth of the destination and the predicted destination. 3) Graph Accuracy (**Graph.**) measures the consistency between the predicted interaction graph and the input interaction graph. To ensure a fair comparison with previous methods [39], we use Graph Accuracy to evaluate our model on the PHASE and Socialnav datasets since these datasets have the corresponding ground-truth social interaction graph.

Implementation Details. For generalization in the Socialnav dataset, we change the environment by setting the number of agents to 10 (doubling the number of agents) and setting the preferred speed of agents to 2 (doubling the speed of agents). For generalization in the Charged dataset, we change the environment by setting the number of agents to 10 (doubling the number of agents), setting the time step to 0.1 (halving the sampling timestep), setting the side length of the box square to 3.5 (halving the environment space size).

We do experiments for PHASE, Socialnav, and Charged datasets with Nvidia 2080Ti GPU, with Ubuntu 18.04 system. Our networks were implemented using the PyTorch framework. For Socialnav and Charged datasets, the input dimension is 2 (position for each agent) and it is 4 for the PHASE dataset (position and label for each agent). We train the model with Adam optimizer with learning rate 1×10^{-3} and decay the learning rate by a factor of 0.9 if the validation performance has not improved in 5 epochs. We use a hidden size of 96, 128, and 128 for the Social Navigation Environment, PHASE, and the Charged dataset, and we use 4 latent graph layers for PHASE and Socialnav datasets (the dimension of edge latent $\mathbf{z}$ is 4) and 4 latent graph layers for the Charged dataset (the dimension of edge latent $\mathbf{z}$ is 2).

As the size of the NBA dataset is much larger than the other three datasets, we do an experiment using Nvidia 3090 GPU, with Adam optimizer with learning rate 5×10^{-4} and decay the learning rate by a factor of 0.9 if the validation performance has not improved in 5 epochs. We use a hidden size of 256 and 5 latent graph layers for the NBA dataset (the dimension of edge latent $\mathbf{z}$ is 5).

For implementing IMMA, RFM, RFM (skip 1), and GAT models for comparison in the Charged dataset, we use the default configurations mentioned in IMMA, that is, we use Adam optimizer with an initial learning rate of $1e-6$ and decay the learning rate by a factor of 0.9 if the validation performance has not improved in 5 epochs, and we use 2 latent graph layers (the dimension of edge latent $\mathbf{z}$ is 2).

4.3 Results for Trajectory Prediction

Quantitative Evaluation. In Table 1, we report ADE, FDE, and Graph Accuracy for each model on PHASE and Socialnav datasets. Our model achieves the best performance on both datasets. Compared with the IMMA model, we achieved a relative improvement by 17.6% / 13.7% in ADE / FDE on the PHASE dataset, respectively. On the Socialnav dataset, the improvement is 20.1% / 17.9% in ADE / FDE. In the aspect of the Graph Accuracy metric, we compare the ability of each model to predict social interaction graphs. The

Model	ADE↓	FDE↓
NRI [22]	0.63	1.30
FQA [17]	0.82	1.76
dNRI [11]	0.94	1.93
GAT-LSTM [43]	0.74	1.45
Social-GAN [12]	0.66	1.25
RFM(skip 1) [40]	0.78	1.53
RFM [40]	0.79	1.55
IMMA [39]	0.73	1.44
GRIN [25]	0.52	1.09
Ours	0.48	1.05

Table 2: Quantitative results on the Charged dataset.

Model	ADE↓	FDE↓
MLP	1.113	1.990
GAT-LSTM [43]	0.978	1.733
NRI [22]	0.946	1.818
EvolveGraph [24]	0.896	1.695
RFM(skip 1) [40]	0.938	1.756
RFM [40]	0.839	1.572
fNRI [45]	0.804	1.517
IMMA [39]	0.769	1.438
Ours	0.659	1.266

Table 3: Quantitative results on NBA dataset.

Dataset	Method	ADE↓	FDE↓
PHASE	MATE	0.706	1.462
	MATE + $\mathcal{L}_E$	0.688	1.357
	MATE + $\mathcal{L}_D$	0.690	1.382
	MATE + $\mathcal{L}_E + \mathcal{L}_D$	0.660	1.281

Table 4: Ablation study on PHASE dataset.

results show that our model achieves an accuracy of over 90% on both datasets. It means that our method can not only accurately predict the future trajectories of each agent, but also clearly express the relationships between agents, which helps explain how agents affect other’s trajectories. Our model and other models are also tested on Charged dataset and NBA dataset. Results are presented in Table 2 and Table 3. As can be seen, our model yields the best performance compared with existing methods, indicating that our method is robust to both simulated and real-world scenarios.

Qualitative Analysis. Fig. 2 visualizes the trajectory prediction results of our model on Socialnav dataset. Note that each agent has a target agent and moves towards its target. It can be seen that our model accurately predicts the future trajectory of agents with the assistance of their past information and interactions with others. We also visualize the prediction results of our method and others in Fig. 3 on Charged dataset. It can be seen that interactions exist between charges that repel and attract each other. We accurately predict such interactions and future trajectories of charges. We visualize the results on NBA dataset in Fig. 5 and show the trajectories of the ball and the players in different colored lines. See more visualizations of our model in the Appendix.

Model	Double agents			Double speed		
	ADE↓	FDE↓	Graph.%↑	ADE↓	FDE↓	Graph.%↑
MLP	-	-	-	0.303	0.632	-
GAT-LSTM [43]	1.486	2.538	19.84	0.361	0.635	22.04
NRI [22]	0.527	1.096	34.57	0.269	0.485	55.10
EvolveGraph [24]	0.354	0.988	50.56	0.219	0.421	67.30
RFM(skip 1) [40]	0.441	0.792	49.41	0.209	0.418	68.03
RFM [40]	0.333	0.762	51.37	0.205	0.411	69.54
fNRI [45]	0.310	0.659	19.71	0.206	0.410	33.48
IMMA [39]	0.195	0.406	64.25	0.192	0.383	78.84
Ours	0.173	0.367	78.34	0.159	0.329	89.02

Table 5: Quantitative results of zero-shot generalization on Socialnav dataset.

Model	Double agents		Half sampling Δt		Half space	
	ADE↓	FDE↓	ADE↓	FDE↓	ADE↓	FDE↓
GAT-LSTM [43]	1.198	2.282	0.381	0.738	1.435	2.667
RFM(skip 1) [40]	4.538	10.44	0.406	0.822	1.380	2.5286
RFM [40]	1.897	4.019	0.414	0.834	1.378	2.5287
IMMA [39]	1.140	2.227	0.402	0.777	1.400	2.5290
Ours	1.011	2.148	0.348	0.745	1.341	2.604

Table 6: Quantitative results of zero-shot generalization on Charged dataset.

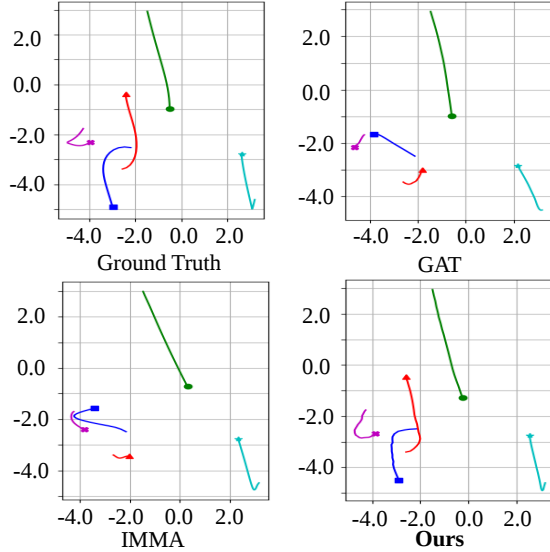


Figure 3: Qualitative analysis of models on the Charged dataset. Different colored lines denote the trajectories of different charges.

4.4 Ablation Study

Table 4 shows the results of our ablation study on PHASE dataset, where we verify the effect of different components of our model. Specifically, we compare the performance of our model with different variants: 1) MATE (base model), 2) MATE + $\mathcal{L}_E$ (inter-agent interaction constraint), 3) MATE + $\mathcal{L}_D$ (intra-agent motion constraint), and 4) MATE + $\mathcal{L}_E + \mathcal{L}_D$. We depict the results in Table 4. Compared with MATE (base model), our model performs better with the assistance of the inter-agent interaction constraint

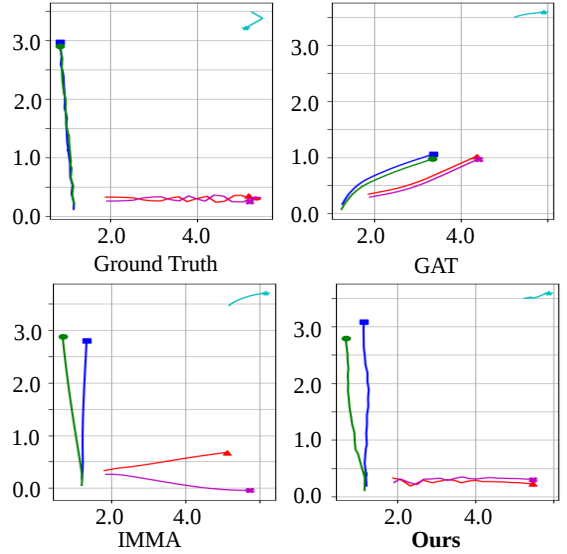


Figure 4: Visualization of generalization on the Charged dataset. Different colored lines denote the trajectories of different charges.

on PHASE dataset, indicating that modeling interactions between agents through neural interaction energy and constraining it as prior knowledge preserves a temporally stable system, improving the prediction accuracy and robustness of our model. Moreover, adding the intra-agent motion constraint further boosts the performance, indicating that punishing the incoherent motion prediction and connecting the relation between interactions and motions of agents helps to preserve the agent-level temporal stability.

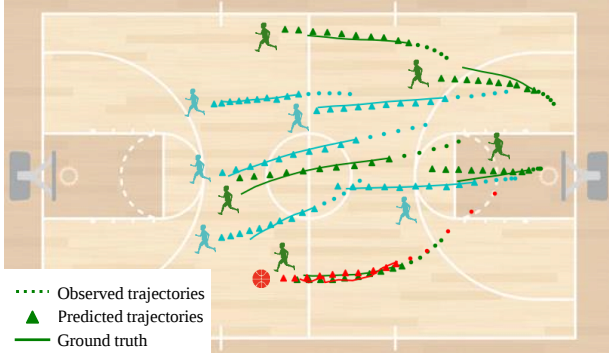


Figure 5: Qualitative analysis of our model on NBA dataset. The red agent is the basketball, and the other colored agents represent players.

4.5 Generalization to Unseen Scenarios

Our method focuses on preserving temporal stability, which exists in various multi-agent systems. Consequently, our approach is robust to unseen scenarios. We conduct zero-shot experiments to present the generalization ability of our method. Table 5 and Table 6 depict the zero-shot experiments on Socialnav and Charged datasets, where we test the trajectory prediction accuracy of our model and other existing models to unseen situations.

On Socialnav dataset, we create two new environments by 1) doubling the number of agents and 2) doubling the speed of agents. As shown in Table 5, our model outperforms other methods in both environments, demonstrating that our model is generalizable to different scenarios of complexity and dynamics in the social navigation environment.

On Charged dataset, we test our model on three novel situations: 1) doubling the number of charges, 2) halving the sampling timestep, and 3) halving the environment space size. As shown in Table 6, our model achieves the best performance in most situations. In Fig. 4, we also visualize the results of our model and others in the unseen situation by halving the environment space size. Though the trajectory prediction accuracy of each model drops, our model performs the best. This indicates that our model has a better generalization ability.

4.6 Results of Different Dataset Sizes

To investigate the impact of different size datasets on our models, we simulate datasets of 50k, 25k, 12.5k, and 10k samples, and use the same network structure and hyperparameters. The result is in Fig. 6. Compared with the strongest baseline (IMMA), our model outperforms on all datasets of different sizes, and the prediction performance remains stable.

5 CONCLUSION

In this work, we proposed a framework named MATE for multi-agent trajectory prediction, aiming at eliminating the prediction fluctuations and preserving the temporal stability of multi-agent systems. Our framework introduces the concept of neural interaction energy to model the interactions and motions of agents in a dynamic environment. Furthermore, we employ partial differential

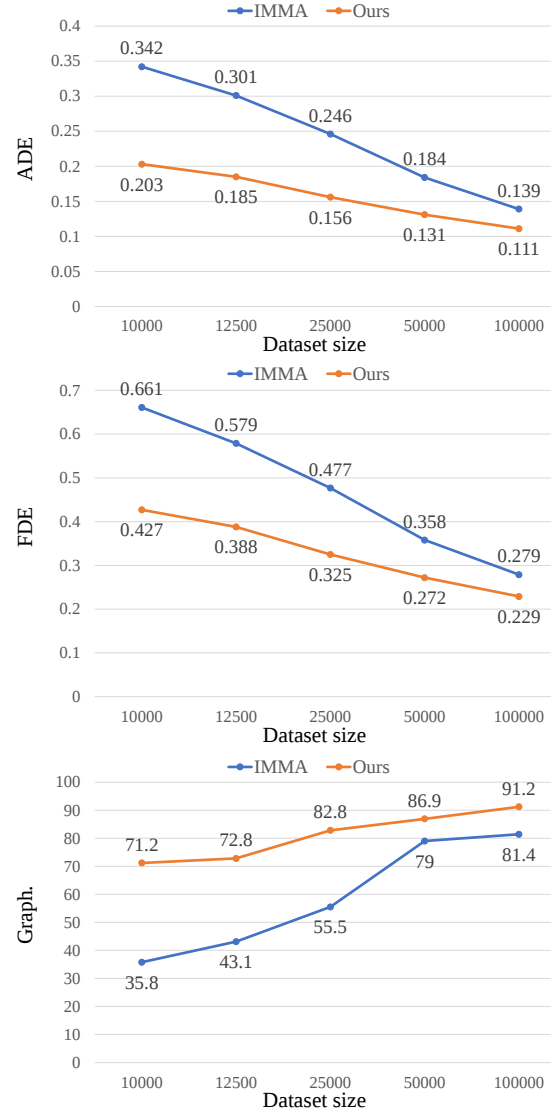


Figure 6: Results of our model and IMMA on different sizes of Socialnav datasets. Lower ADEs and FDEs and higher graph accuracies (Graph.) are better.

equations to establish the inter-agent interaction constraint and the intra-agent motion constraint, which preserve the temporal stability of the multi-agent system and improve the prediction accuracy of our model. Through quantitative and qualitative experiments, we evaluate our framework on four datasets and demonstrate that it outperforms the state-of-the-art methods in trajectory prediction. We also show the advantages of our framework in zero-shot generalization capacity. Our work opens up new possibilities for representing interactions between agents for multi-agent systems and illustrating the mechanism of how interactions influence agents' future trajectories.

REFERENCES

- [1] Alexandre Alahi, Kratharth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. 2016. Social lstm: Human trajectory prediction in crowded spaces. In *CVPR*.
- [2] Defu Cao, Yujing Wang, Juanyong Duan, Ce Zhang, Xia Zhu, Congrui Huang, Yunhai Tong, Bixiong Xu, Jing Bai, Jie Tong, et al. 2020. Spectral temporal graph neural network for multivariate time-series forecasting. In *NeurIPS*.
- [3] Changan Chen, Yuejiang Liu, Sven Kreiss, and Alexandre Alahi. 2019. Crowd-robot interaction: Crowd-aware robot navigation with attention-based deep reinforcement learning. In *ICRA*.
- [4] Ricky TQ Chen, Yulia Rubanova, Jesse Bettencourt, and David K Duvenaud. 2018. Neural ordinary differential equations. In *NeurIPS*.
- [5] Maria Chli, Philippe De Wilde, Jan Goossenaerts, Vladimir Abramov, Nick Szirbik, Luis Correia, Pedro Mariano, and Rita Ribeiro. 2003. Stability of multi-agent systems. In *SMC'03 Conference Proceedings. 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme-System Security and Assurance*.
- [6] Nachiket Deo and Mohan M Trivedi. 2018. Convolutional social pooling for vehicle trajectory prediction. In *CVPR Workshops*.
- [7] Yaofeng Desmond Zhong, Biswadip Dey, and Amit Chakraborty. 2020. Symplectic ODE-Net: Learning Hamiltonian Dynamics with Control. In *ICLR*.
- [8] Dennis Fassmeyer, Pascal Fassmeyer, and Ulf Brefeld. 2022. Semi-Supervised Generative Models for Multiagent Trajectories. In *NeurIPS*.
- [9] Harshayu Girase, Haiming Gang, Srikanth Malla, Jiachen Li, Akira Kanehara, Karttikeya Mangalam, and Chiho Choi. 2019. Loki: Long term and key intentions for trajectory prediction. In *CVPR*.
- [10] Francesco Giuliani, Irtiza Hasan, Marco Cristani, and Fabio Galasso. 2021. Transformer networks for trajectory forecasting. In *ICPR*.
- [11] Colin Graber and Alexander Schwing. 2020. Dynamic neural relational inference for forecasting trajectories. In *CVPR Workshops*.
- [12] Agrim Gupta, Justin Johnson, Li Fei-Fei, Silvio Savarese, and Alexandre Alahi. 2018. Social gan: Socially acceptable trajectories with generative adversarial networks. In *CVPR*.
- [13] Sandro Hauri, Nemanja Djuric, Vladan Radosavljevic, and Slobodan Vucetic. 2021. Multi-Modal Trajectory Prediction of NBA Players. In *CVPR*.
- [14] Dirk Helbing and Peter Molnar. 1995. Social force model for pedestrian dynamics. *Physical review E* 51, 5 (1995), 4282.
- [15] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. 2017. beta-VAE: Learning basic visual concepts with a constrained variational framework. In *ICLR*.
- [16] Yingfan Huang, Huikun Bi, Zhaoxin Li, Tianlu Mao, and Zhaoqi Wang. 2019. STGAT: Modeling spatial-temporal interactions for human trajectory prediction. In *ICCV*.
- [17] Nitin Kamra, Hao Zhu, Dweep Kumarbhai Trivedi, Ming Zhang, and Yan Liu. 2020. Multi-agent trajectory prediction with fuzzy query attention. In *NeurIPS*.
- [18] Takayuki Kanda, Hiroshi Ishiguro, Tetsuo Ono, Michita Imai, and Ryohei Nakatsu. 2002. Development and evaluation of an interactive humanoid robot "Robovie". In *ICRA*.
- [19] Ioannis Karamouzas, Brian Skinner, and Stephen J Guy. 2014. Universal power law governing pedestrian interactions. *Physical review letters* 113, 23 (2014), 238701.
- [20] George Em Karniadakis, Ioannis G Kevrekidis, Lu Lu, Paris Perdikaris, Sifan Wang, and Liu Yang. 2021. Physics-informed machine learning. *Nature Reviews Physics* 3, 6 (2021), 422–440.
- [21] Diederik P Kingma and Max Welling. 2014. Auto-encoding variational bayes. In *ICLR*.
- [22] Thomas N. Kipf, Ethan Fetaya, Kuan-Chieh Wang, Max Welling, and Richard S. Zemel. 2018. Neural Relational Inference for Interacting Systems. In *ICML*.
- [23] Jesse Levinson, Jake Askeland, Jan Becker, Jennifer Dolson, David Held, Soeren Kammel, J Zico Kolter, Dirk Langer, Oliver Pink, Vaughan Pratt, et al. 2011. Towards fully autonomous driving: Systems and algorithms. In *IV*.
- [24] Jiachen Li, Fan Yang, Masayoshi Tomizuka, and Chiho Choi. 2020. Evolvegraph: Multi-agent trajectory prediction with dynamic relational reasoning. In *NeurIPS*.
- [25] Longyuan Li, Jian Yao, Li Wenliang, Tong He, Tianjun Xiao, Junchi Yan, David Wipf, and Zheng Zhang. 2021. GRIN Generative Relation and Intention Network for Multi-agent Trajectory Prediction. In *NeurIPS*.
- [26] Zongyi Li, Hongkai Zheng, Nikola Kovachki, David Jin, Haoxuan Chen, Burigede Liu, Kamyar Azizzadenesheli, and Anima Anandkumar. 2021. Physics-informed neural operator for learning partial differential equations. *arXiv preprint arXiv:2111.03794* (2021).
- [27] Huynh Manh and Gita Alaghband. 2018. Scene-lstm: A model for human trajectory prediction. *arXiv preprint arXiv:1808.04018* (2018).
- [28] Weibo Mao, Chenxin Xu, Qi Zhu, Siheng Chen, and Yanfeng Wang. 2023. Leapfrog Diffusion Model for Stochastic Trajectory Prediction. In *CVPR*.
- [29] Abdullah Mohamed, Kun Qian, Mohamed Elhoseiny, and Christian Claudel. 2020. Social-stgcn: A social spatio-temporal graph convolutional neural network for human trajectory prediction. In *CVPR*.
- [30] Aviv Netanyahu, Tianmin Shu, Boris Katz, Andrei Barbu, and Joshua B Tenenbaum. 2021. PHASE: PHysically-grounded Abstract Social Events for Machine Social Perception. In *AAAI*.
- [31] Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. 2017. Automatic differentiation in pytorch. In *NeurIPS*.
- [32] Julien Pettré, Jan Ondřej, Anne-Hélène Olivier, Armel Cretual, and Stéphane Donikian. 2009. Experiment-based modeling, simulation and validation of interactions between virtual walkers. In *Proceedings of the 2009 ACM SIGGRAPH/Eurographics symposium on computer animation*.
- [33] Ruijie Quan, Linchao Zhu, Yu Wu, and Yi Yang. 2021. Holistic LSTM for pedestrian trajectory prediction. *IEEE transactions on image processing* 30 (2021), 3229–3239.
- [34] Neil Rabinowitz, Frank Perbet, Francis Song, Chiyuan Zhang, SM Ali Eslami, and Matthew Botvinick. 2018. Machine theory of mind. In *ICML*.
- [35] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. 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 (2019), 686–707.
- [36] Alvaro Sanchez-Gonzalez, Jonathan Godwin, Tobias Pfaff, Rex Ying, Jure Leskovec, and Peter Battaglia. 2020. Learning to simulate complex physics with graph networks. In *ICML*.
- [37] Shyam Sankaran, Hanwen Wang, Leonardo Ferreira Guilhoto, and Paris Perdikaris. 2022. On the impact of larger batch size in the training of Physics Informed Neural Networks. In *NeurIPS Workshop*.
- [38] Kihyuk Sohn, Honglak Lee, and Xinchun Yan. 2015. Learning structured output representation using deep conditional generative models. In *NeurIPS*.
- [39] Fan-Yun Sun, Isaac Kaurav, Ruohan Zhang, Jiachen Li, Mykel J Kochenderfer, Jiajun Wu, and Nick Haber. 2022. Interaction modeling with multiplex attention. In *NeurIPS*.
- [40] Andrea Tacchetti, H. Francis Song, Pedro A. M. Mediano, Vinicius Zambaldi, János Kramár, Neil C. Rabinowitz, Thore Graepel, Matthew Botvinick, and Peter W. Battaglia. 2019. Relational Forward Models for Multi-Agent Learning. In *ICLR*.
- [41] Bohan Tang, Yiqi Zhong, Chenxin Xu, Wei-Tao Wu, Ulrich Neumann, Ya Zhang, Siheng Chen, and Yanfeng Wang. 2023. Collaborative uncertainty benefits multi-agent multi-modal trajectory forecasting. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2023).
- [42] Jur Van Den Berg, Stephen J Guy, Ming Lin, and Dinesh Manocha. 2011. Reciprocal n-body collision avoidance. In *Robotics Research: The 14th International Symposium ISRR*.
- [43] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. 2018. Graph attention networks. In *ICLR*.
- [44] Anirudh Vemula, Katharina Muelling, and Jean Oh. 2018. Social attention: Modeling attention in human crowds. In *ICRA*.
- [45] Ezra Webb, Ben Day, Helena Andres-Terre, and Pietro Liò. 2019. Factorised neural relational inference for multi-interaction systems. In *ICML Workshops*.
- [46] Jiangbei Yue, Dinesh Manocha, and He Wang. 2022. Human trajectory prediction via neural social physics. In *ECCV*.
- [47] Eric Zhan, Stephan Zheng, Yisong Yue, Long Sha, and Patrick Lucey. 2019. Generating multi-agent trajectories using programmatic weak supervision. In *ICLR*.
- [48] Tianyang Zhao, Yifei Xu, Mathew Monfort, Wongun Choi, Chris Baker, Yibiao Zhao, Yizhou Wang, and Ying Nian Wu. [n. d.]. Multi-agent tensor fusion for contextual trajectory prediction. In *CVPR*.
- [49] Kirill Zubov, Zoe McCarthy, Yingbo Ma, Francesco Calisto, Valerio Pagliarino, Simone Azeglio, Luca Bottero, Emmanuel Luján, Valentin Sulzer, Ashutosh Bharambe, et al. 2021. Neuralpde: Automating physics-informed neural networks (pinns) with error approximations. *arXiv preprint arXiv:2107.09443* (2021).