Sparse Multi-Relational Graph Convolutional Network for Multi-Type Object Trajectory Prediction

Jianhui Zhang[1] , **Jun Yao**[1] , **Liqi Yan**[1*] , **Yanhong Xu**[1] and **Zheng Wang**[2]

[1]College of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018 China
[2]School of Computer Science, Wuhan University, Wuhan 430072 China
{jh_zhang, jun_yao, ylq, yhxu}@hdu.edu.cn, wangzwhu@whu.edu.cn

Abstract

Object trajectory prediction is a hot research issue with wide applications in video surveillance and autonomous driving. The previous studies consider the interaction sparsity mainly among the pedestrians instead of multi-type of objects, which brings new types of interactions and consequently superfluous ones. This paper proposes a Multi-type Object Trajectory Prediction (MOTP) method with a Sparse Multi-relational Graph Convolutional Network (**SMGCN**) and a novel multi-round Global Temporal Aggregation (GTA). MOTP introduces a novel adaptive sparsification and multi-scale division method to model interactions among multi-type of objects. It further incorporates a Sparse Multi-relational Temporal Graph to capture the temporal division of multi-type trajectories, along with a multi-round Global Temporal Aggregation (GTA) mechanism to mitigate error accumulation, and enhances the trajectory prediction accuracy. The extensive evaluation on the ETH, UCY and SDD datasets shows that our method outperforms the typical state-of-the-art works by significant margins. *Codes will be available in https://github.com/sounio/SMGCN*.

1 Introduction

Trajectory prediction [Rudenko *et al.*, 2020] primarily utilizes the information from the observed trajectories of objects to analyze a sequence of future location coordinates of them, which has drawn considerable data analysis researches and critial applications in various data processing hot areas, including autonomous driving [Yuan *et al.*, 2018], smart transportation [Zhou *et al.*, 2021][Lv *et al.*, 2021] and visual recognition [Hu *et al.*, 2021][Liang *et al.*, 2022].

Despite of the recent development in the area, trajectory prediction is still a challenging and hot video analysis task due to the multiple types of objects and complex interactions among them. The existing works mainly focus on the pedestrian trajectory prediction [Huang *et al.*, 2022][Shi *et al.*, 2021], which usually study the motion of pedestrians and is disturbed easily [Gupta *et al.*, 2018], like to group with friends [Mohamed *et al.*, 2020] or to take social behaviours similar to the others [Sun *et al.*, 2020]. The previous works put a strong assumption on the interaction among pedestrians that each has an interaction with all the rest ones, which generates superfluous interactions and ignores trajectory motion tendency [Shi *et al.*, 2021]. Shi *et al.* argues that the interactions among pedestrians present are a sparse and directed [Shi *et al.*, 2021]. Meanwhile, the existing works usually adopts the metrics, distance or attention, to model the interactions among pedestrians [Shi *et al.*, 2021][Mohamed *et al.*, 2020][Alahi *et al.*, 2014][Bae and Jeon, 2021].

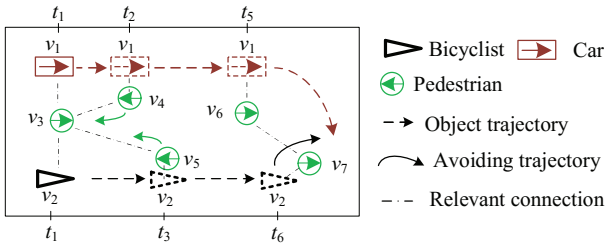

Figure 1: **Sparse Multi-relational Directed Interaction & Multi-type Motion Tendency. (I.1)** Multi-type objects have quite different speed and create diverse multi-relational interactions, such as the ones among pedestrian, car and bicycle. **(I.2)** The interaction between the same kind of objects is directional, sparse, and among different types of objects which tend to more directional than the the former one and dynamically sparse. **(I.3)** The motion tendency is distinct among multi-type objects because of their unique position, speed and type.

This paper take more types of objects into account, which gives a rise to intricate interactions, encompassing different types and diverse trajectories of objects, as vividly depicted in Figure 1. The challenging nature of these object trajectories goes beyond what's typically encountered, primarily due to the complex multi-relational dynamics among diverse types of objects, such as pedestrians, cars, and bicyclists, in the context of city video surveillance. Furthermore, these multi-type objects normally engage with neighboring ones in an asymmetric manner, resulting in sparse multi-relational directional connections among them. However, prior works focused on collision avoidance or motion tendency [Shi *et al.*, 2021], which regularly considers short-term trajectories

*Corresponding author.

and assumes gradual changes in directions for the same object type (*e.g.*, pedestrians). Figure 1 illustrates ***three specific indications***, and exposes the limitations of these previous methods: **(I.1)** car v_1 and bicyclist v_2 have higher speeds than the pedestrians so v_1 connects with v_4 at t_2 and disconnects with v_3, and v_2 connects with v_5 at t_3 and disconnects with v_3. Meanwhile, the connection of pedestrians, such as v_3, v_4 and v_5, is relatively stable; **(I.2)** when car v_1 or bicycle v_2 heads towards from the opposite direction with a pedestrian v_4 or v_5 at t_2 or t_5, the trajectory of pedestrian usually detours to avoid the collision; **(I.3)** cars usually have a larger radius of turning circle than bicycles, while bicycle has larger one than pedestrian so the trajectory may be quite different between different types of object. It is obvious that the precious works on sparse directed interaction methods which only consider pedestrians [Shi *et al.*, 2021] cannot describe the interactions among multi-type objects as **I.1**. The traditional works for socially entangled pedestrian trajectory prediction [Bae and Jeon, 2021] cannot deal with the social interactions among various types of objects as **I.1** and the asymmetric collision avoid among different types of objects as **I.2**. Figure 2 shows that the precious works, which divide pedestrians into several groups, are not suitable for the case of multi-type objects. Although motion tendency can facilitate the prediction with temporal information from trajectories [Shi *et al.*, 2021], it is insufficient to describe the motion tendency of multi-type objects since the difference of them as **I.3**.

To address the aforementioned challenges, this paper introduces a novel approach known as **S**parse **M**ulti-Relational **G**raph **C**onvolutional **N**etwork (**SMGCN**), depicted in Figure 3. The SMGCN is designed to capture intricate interactions in multi-type objects, encompassing both inter-type and inner-type dynamics, which describes object trajectories from the input videos by a spatial graph and a temporal graph as shown in Figure 3. Compared with previous methods, the competitiveness of our approach lies in the following five aspects. **Firstly**, we introduce a novel sparse multi-relational spatial graph, grouping objects based on distance and relative displacement while integrating position and label information. It prunes unnecessary connections in different types of objects and models the sparse directed interactions among them, as depicted in Figure 2. **Secondly**, we leverage the multi-head self-attention mechanism [Vaswani *et al.*, 2017] to learn asymmetric inter and inner-grouped directed interaction scores among multi-type objects. These scores combine distance with relative displacement fusions, and use asymmetric convolutional network to yield the fusion interaction mask matrix. **Thirdly**, we propose an adaptive sparsification method that computes thresholds by averaging each matrix dimension and fusing all averages, avoiding fixed thresholds. It yields sparse interaction score adjacency matrices, representing the sparse multi-relational spatial graph, which obtains by using a method similar to [Shi *et al.*, 2021]. **Fourthly**, our SMGCN model introduces multi-motion tendencies, representing trajectory patterns for different types of objects. Objects with low speed tend to follow smaller turning radii when avoiding collisions, while objects with higher speeds take larger turns. For example, pedestrians tend to avoid bicyclists, whereas the latter tend to avoid cars. **Finally**, the object trajectories can be represented by the sparse multi-relational spatial and temporal graphs, with a series of GCNs [Kipf and Welling, 2017]. These graphs estimate bi-Gaussian distribution parameters by the Time Convolution Network[Bai *et al.*, 2018], and enable trajectory prediction for each object.

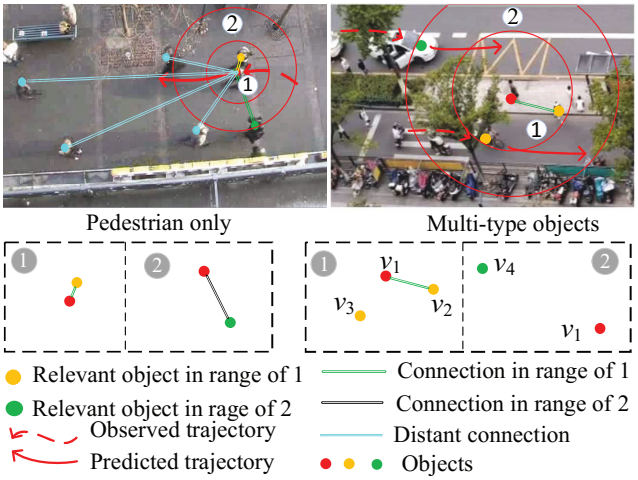

Figure 2: **Combining object grouping with interaction sparsity.** (1) Pedestrians can be grouped by distance or relative displacement in the left subgraph. (2) Multi-type objects are not suitable to be grouped by the two metrics. Car v_4 may be separated by road shoulder with pedestrian v_1. Bicyclist v_3 has more frequent interaction with other object than pedestrians for its high speed.

To our best knowledge, this paper contributes a quite novel method to model the multi-type objects interaction with the Multi-Relational Graph Convolutional Network and the Sparse Directed Interaction explicitly. Our contributions state as the following three-fold:

- Propose a new model, Sparse Multi-relational Spatial Graph, combines intricate multi-scale divisions with effective interaction sparsification, effectively pruning away unnecessary interactions among diverse types of objects.
- Design an adaptive sparsification method, which not only incorporates diverse information such as distance, relative displacement, object position and label, but also employs multi-layer GTA, leading to substantial enhancements in the accuracy of trajectory predictions.
- Propose a Sparse Multi-relational Temporal Graph to capture object trajectory temporal division and accurately model multi-type motion tendencies. This feature enables a more nuanced understanding of object behaviors in various scenarios.

Our SMGCN offers an intuitive, yet versatile solution for multi-type trajectory prediction that definitively surpasses a diverse array of architectural approaches. This study extensively engages in numerical experiments employing ETH [Pellegrini *et al.*, 2009], UCY [Lerner *et al.*, 2007], and SDD datasets [Robicquet *et al.*, 2016]. Our experiments illustrate the superiority of our method compared to state-of-

the-art alternatives. Specifically, our approach showcases a reduction of ↓ **5.71%** in Average Displacement Error (ADE) and ↓ **8.33%** in Final Displacement Error (FDE) across the ETH and UCY datasets. Similarly, in the SDD datasets, our method achieves a substantial decrease of ↓ **26.66%** in Minimum ADE and ↓ **12.49%** in Minimum FDE.

2 Related Work

2.1 Pedestrian Trajectory Prediction

Advanced over the previous works on trajectory prediction with social forces [Mehran *et al.*, 2009], the recent works introduce the machine learning such as CNNs and RNNs to predict the trajectory of pedestrians like Social-GAN [Gupta *et al.*, 2018], Social-LSTM [Alahi *et al.*, 2016], and so on. They present new social pooling schemes to compute all pedestrians' importance by attention modules. Gupta *et al.* presents the Social-GAN model based on a generative model to predict trajectories in a multi-modal way and accordingly a socially-acceptable path [Gupta *et al.*, 2018]Sadeghian *et al.* designs the SoPhie model to capture the interactions between human and environment respectively [Sadeghian *et al.*, 2019]. Shi *et al.* describes the relationship among pedestrians by an attention module, and predicts the next coordinates for a pedestrian by Gaussian mixture model [Shi *et al.*, 2020]. Sun *et al.* develops a reciprocal learning, for human forward and backward trajectory prediction with two prediction networks [Sun *et al.*, 2020]. Bae *et al.* presents a disentangled multi-relational GCN, and socially entangled pedestrian trajectory prediction [Bae and Jeon, 2021]. Furthermore, Shi *et al.* presents a Sparse Graph Convolution Network (SGCN) for pedestrian trajectory prediction by modeling the sparse directed interaction among pedestrians and the motion tendency [Shi *et al.*, 2021].

2.2 Multiple-Class Trajectory Prediction

Prior works mainly model sparse interactions among pedestrians, and only a few of them focus on multi-type users *e.g.*, vehicle, cyclist, *etc.*. Li *et al.* proposes a sparse graph convolution network based approach for multi-class trajectory prediction to decide the spatial and temporal connections of agents [Li *et al.*, 2022]. However, the generated relation graph is always overly dense with unnecessary connections between objects, without addressing sparsity to eliminate redundant interactions among them. This paper presents a sparse multi-relational directed interaction which divides objects into several spatial-temporal groups based on metrics (*e.g.*, position, distance and relative displacement) and enables the removal of superfluous interactions within and between groups. This methodology effectively describes interactions among multi-type objects and leverages adaptive sparsification.

3 Our Method

This paper proposes an end-to-end MOTP method based on a sparse multi-relational GCN as shown in Figure 3. It contributes as follows: (1) spatial multi-relation division to eliminate the irrelevant relations and dividing the relevant relations among the multi-type objects into several groups; (2) an adaptive threshold based sparsification to optimize the sparsity of graph representation for interactions among objects; (3) multi-relational GCN to extract multi-relational spatial and temporal interactions respectively.

3.1 Sparse Multi-Relational Graph Learning

Graph input. Given a series of observations sampled from continuous video frames over time moments $t \in \{1, \cdots, T\}$, where T is the period of observation, all objects in the video can be detected to obtain their coordinates $\{(x_i^t, y_i^t)\}_{i=1}^N$ with certain algorithms. This section describes the sequence of input frames, which contains multi-type objects, with a spatial graph denoted by G_s and a temporal graph denoted by G_t, and takes the two graphs as the input of the MSG. Let v denote object and V denote the set of objects, among which is the set E of edges denoted by e_j, $j = 1, \cdots$. This paper considers that each object v_k may belong to different types, defined τ_k, $k = 1, \cdots, Y$, which is represented with one-hot encoded semantic labels. The spatial graph $G_s(V^t, E^t)$ and the temporal graph $G_t(V_i, E_i)$ composes of the object locations at time t and the trajectories corresponding to object v_i respectively, where $V^t = \{v_i^t, i = 1, \cdots, N\}$ and $V_i = \{v_i^t, t = 1, \cdots, T\}$. v_i^t means the attributes of object v_i at time t containing its coordinates and type. The sets of edges among the two sets of objects are expressed as $E^t = \{e_{i,j}^t, i, j = 1, \cdots, N\}$ for G_s and $E_i = \{e_i^{l,m}, l, m = 1, \cdots, T\}$ for G_t respectively, where $e_{i,j}^t$ and $e^{l,m}$ are set as 1 when objects v_i^t, v_j^t or v_i^l, v_i^m are connected, and 0 otherwise.

Multi-relation Spatial Division. One key contribution of ours involves a novel approach to categorize multi-type objects into group and analyze the relations within each group, as well as among different groups, to eliminate irrelevant relations. This distinguishes our work from prior approaches, which usually analyze the relation in single group and all relations are assumed to be relevant [Bae and Jeon, 2021]. For the purpose, we present a Sparse Multi-relation Spatial Graph, which contains several key procedures: multi-relation spatial division, spatial self-attention, spatial-temporal fusion and sparse, as shown in Figure 3.

Different from the previous works which learn complementary features by combining the information of distance and the relative displacement [Bae and Jeon, 2021], we utilize the object information including the Euclidean distance, the relative displacement, position and label among objects so as to increase the sparsity for the spatial graph inputs, *i.e.*, to eliminate the irrelevant relations and select out the exact ones among objects from the input spatial graph. The information from labels indicates the type of specific object. The position is the geometric location of objects and offers extra information for determining relation besides distance and relative displacement.

Inspired by the reference [Bae and Jeon, 2021], which proposes disentangled multi-scale aggregation of social relations on a weighted graph, we use two groups of weighted sub-graphs for scale metrics: distance and relative displacement. These sub-graphs are fused separately with position and label information. Each sub-graph represents an object division, with the total number of divisions set to B for both scale met-

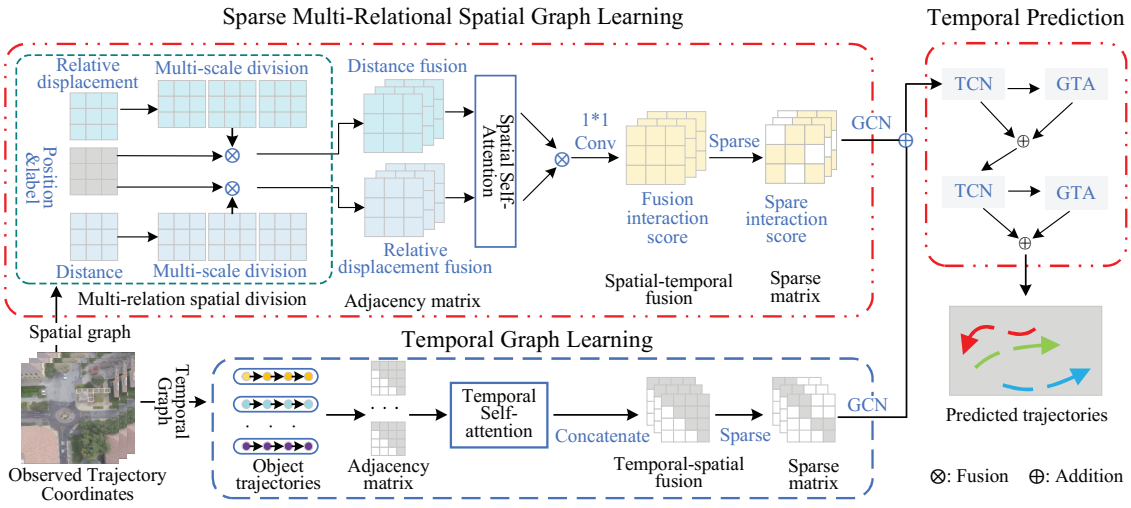

Figure 3: **Sparse multi-relational graph learning.** It consists of two procedures including the sparse multi-relational spatial and temporal graph learning. The multi-scale division divides the objects into several groups, and the fusion procedure generates two sequences of matrices, based on which the spatial self-attention generates the scores for the spatial interaction among objects and yields a interaction fusion matrix by 1×1 convolution. The temporal self-attention generates the scores for the temporal interaction among object trajectories.

rics. Let $A^k_{\chi,ij}$ denote the adjacent matrix of the sub-graph representing the k^{th} object division. It represents relations among objects with edge weights measured by distance or displacement scales. This section defines it for each observation moment by following equation:

$$A^k_{\chi,ij} = \begin{cases} 1 & r[k] \leq a_{ij} < r[k+1] \text{ or } i = j; \\ 0 & \text{otherwise;} \end{cases} \quad (1)$$

where a_{ij} is the distance or relative displacement relation between object v_i and v_j, χ={distance,relative displacement} indicates the operation in under the metric of either the distance or relative displacement. $r[k]$ is a scale factor, $k = 1, 2, \cdots, B$ and responses for two cases: distance scale and relative displacement. For the distance scale, a_{ij} is the Euclidean distance between object v_i and v_j, and $r[k] \in \{0, 0.5, 1, 2, 4\}$. For the relative displacement scale, a_{ij} is the relative displacement between object v_i and v_j, and accordingly, $r[k] \in \{0, 0.25, 0.5, 0.75, 1\}$. For the two cases, the equation (1) leads to two groups of matrices A^k_d and A^k_r for the distance scale and relative displacement scale respectively, where $k = 1, \cdots, B$.

Self-attention mechanism. Previous works primarily use the distance as the input for the relation analysis among objects, neglecting the integration of environmental geometric positions, where predicting object trajectory can benefit. Let H^t be a matrix denoting the information of position and label at moment t. To fuse H^t with the information of distance and relative displacement for each group of objects, *i.e.*, the two matrices A^t_d and A^t_r, we obtain a distance fusion adjacent matrix $\tilde{A}^t_d$ and relative displacement adjacent matrices $\tilde{A}^t_r$ by the multiplication fusion, *i.e.*, $\tilde{A}^t_d = A^t_d \cdot H^t$ and $\tilde{A}^t_r = A^t_r \cdot H^t$.

The another information fusion is in the self-attention as shown in Figure 3, called interaction score fusion which aims to score the interaction among objects in order to obtain matrices that describe the relation weights. To calculate the weight for the relation among objects, this section introduces the attention score matrix based on the two matrices $\tilde{A}_d = A_d \cdot H$ and $\tilde{A}_r = A_r \cdot H$ by the self-attention mechanism [Vaswani *et al.*, 2017][Bae and Jeon, 2021]. The spatial interaction for both matrices are dense and has the same space denoted by $R \in \mathbb{R}^{N \times N}$ among objects for group k at each moment, which is given by embedding one-hot encoded object type labels as following equations:

$$\begin{aligned} &M^k_\chi = \phi(G^k_\chi, W^\chi_M), Q^k_\chi = \phi(M^k_\chi, W^k_{\chi,M}), \\ &K^k_\chi = \phi(M^k_\chi, W^\chi_K), R^k_\chi = Softmax(Q^k_\chi K^T_\chi / \sqrt{d_\chi}), \end{aligned} \quad (2)$$

where $\phi(\cdot, \cdot)$ is a linear transformer. M^k_χ is the embedding of sub-graph A^k_d or A^k_r. W^χ_M,$W^k_{\chi,M}$ and W^χ_K are learnable weight matrices. Accordingly, Q^k_χ and K^k_χ are the query and key of the self-attention mechanism for the sub-graph, respectively. $W^k_{\chi,M} \in \mathbb{R}^{D \times D_K}$, where D is the dimension of spatial graph G_χ. In Equation (2), R^k_χ is calculated step by step for each moment and independent across time steps for both distance and relative displacement fusion matrices.

Interaction score fusion. To capture high-level interaction features by integrating both distance and relative information, $R^{t,k}_d$ and $R^{t,k}_r$ are fused across different sub-graph with the operation of the multiplication fusion and obtain a sequence of fused spatial interaction matrix denoted by $R^{t,k}$, *i.e.*, $R^{t,k} = R^{t,k}_d \otimes R^{t,k}_r$, $\forall, t = 1, \cdots, T, k = 1, \cdots, B$. $R^{t,k}$ represents the interactions among objects at time t for the division scale $r[k]$. To thoroughly explore spatial multi-type object interaction information across various granularity, we fuse the interaction feature across different division scale with a 1×1 convolution and obtain division-spatial-temporal interaction matrix denoted by $[R^{t,B}]^T_{t=1}$. Over the entire observation duration T, we organize the spatial matrices R^t, where $t = 1, \cdots, T$, into a unified matrix with an additional

dimension, $[R^t]_{t=1}^T$, and fuse the interactions belong to the same pair of objects along the temporal channel in $[R^t]_{t=1}^T$ by 1×1 convolution, which leads to a spatial-temporal interaction matrix denoted by $\hat{R}^{t,B}$ with a 1×1 convolution. We follow the equation (2) of the reference [Shi *et al.*, 2021] to implement a cascade of asymmetric convolution operation on the rows and columns of $\hat{R}^{t,B}$, and obtain the high-level interaction feature denoted by F with size $T \times N \times N$.

Sparse. Densely connected graphs suffer from superfluous spatial interactions and temporal dependencies. To address this issue, we introduce an adaptive sparse metric for multi-type relationships, which deviates from conventional approaches typically employ a fixed threshold η to eliminate implicit connections and optimize graph representation sparsity [Li *et al.*, 2022]. Our method proposes a dynamic approach to calculate the adaptive metric from the sparse multi-relational matrix denoted by $\mathcal{M}$ as following equation:

$$\mathcal{M}_{ij} = \begin{cases} 1 & F_{ij} > \sum_{j=1}^{N}\sum_{i=1}^{N} F_{ij}/(N \cdot N), \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

The matrix $\mathcal{M}$ in Equation (3) is fused with the high-level interaction feature by element-wise multiplication, and we can obtain the sparse multi-relational spatial adjacency matrix A as given by the following equation:

$$A = (\mathcal{M} + I) \odot F, \tag{4}$$

where I is a unit matrix to indicate that each object is self-connected, and $\odot$ is an element-wise multiplication operator. A is then normalized with Zero-Softmax function to keep the sparsity as the reference [Kipf and Welling, 2017]. By denoting the normalized adjacency matrix A by $\hat{A}$, we can obtain the sparse multi-relational spatial graph $G(V^t, \hat{A}_s)$.

Sparse multi-relational temporal graph learning. Similar to sparse multi-relational spatial graph, we delve into the interplay between distinct object types from a temporal perspective. With the temporal graph as input, the model in Figure 3 processes the sparse multi-relational temporal graph so as to learn the multi-type motion tendency, which is different from the motion tendency considering only pedestrian [Shi *et al.*, 2021]. In departure from prior research, our study incorporates object types into consideration and introduces the position encoding tensor [Vaswani *et al.*, 2017] to represent the Sparse Multi-relational Temporal Graph, *i.e.*, $E_t = \phi(G_t, W_E^t) + \xi_\tau$, where ξ_τ is the position encoding tensor and τ is the object type. Mirroring the sparse multi-relational spatial graph learning paradigm, we adopt the self-attention mechanism to compute the asymmetric attention score matrices, and concatenate the matrices to the temporal-spatial fusion matrices. We also take the sparsification method, similar to that for the sparse multi-relational spatial graph learning and obtain the sparse matrices. Finally, we obtain a sparse multi-relational temporal graph denoted $G(V_i, \hat{A}_t), v_i \in V$, where $\hat{A}_t$ is the normalized adjacency matrix for v_i.

3.2 Multi-type Trajectory Prediction

The prior subsection derives spatial-temporal features via sparse multi-relational spatial and temporal graphs utilizing multi-relational GCNs. In this section, we adopt a trajectory prediction method inspired by DMRGCN [Bae and Jeon, 2021], employing multi-layer Temporal Convolutional Network (TCN) augmented with Global Temporal Aggregation (GTA). A departure from previous approaches, we introduce the utilization of GTA after each layer of TCN to effectively extract and propagate global temporal information.

Multi-layer TCN. We employ TCN [Bai *et al.*, 2018] to achieve spatial-temporal aggregation by integrating convolutional and recurrent architectures with long-term memory.

Multi-layer GTA. GTA is proposed to learn to compensate for the accumulated error. Different from those typical methods, we update spatio-temporal feature by GTA for every TCN layer to minimize the residual error for multi-type object interactions, which is has a more complex temporal trends compared to single-type object:

$$o'_{l,t,i,c} = o_{l,t,i,c} + \sigma(\sum_{p=1}^{T}\sum_{q=1}^{C}(o_{l,p,i,q} \times w_{l,p,q,c}) + b_{l,i,c}), \tag{5}$$

where $o_{l,t,i,c}$ denotes an element of the spatio-temporal feature O_l output from the l-th TCN layer at time t of object v_i, and C is the total number of channels in the feature. The symbol σ represents the Sigmoid function, while w and b denote the learnable kernel weight and bias, respectively. In our SMGCN, GTA plays a pivotal role in rectifying path predictions, particularly during multi-type object interactions, with a specific emphasis on addressing long-term routes.

Learning Objective. As outlined in previous works such as Social-LSTM [Alahi *et al.*, 2016] and SGCN [Shi *et al.*, 2021], the trajectory coordinates (x_t, y_t) at time t of object v_i follow a bi-variate Gaussian distribution $\mathcal{N}(\mu_i^t, \theta_i^t, \rho_i^t)$, where μ_i^t and θ_i^t are the mean and standard deviation of the distribution respectively, and ρ_i^t is the correlation coefficient. Hence, our model is trained to predict the trajectory of object v_i by minimizing the negative log-likelihood loss as:

$$L^i(W) = \Sigma_{t=T+1}^{T_p} \log P((x_{t,i}, y_{t,i})|\mu_i^t, \theta_i^t, \rho_i^t), \tag{6}$$

where T_p is the prediction time window and $(x_{t,i}, y_{t,i})$ is the coordinates of object v_i at t.

4 Experiment and Analysis

Datasets. We conduct experiments using three public datasets: ETH [Pellegrini *et al.*, 2009], UCY [Lerner *et al.*, 2007], and SDD [Robicquet *et al.*, 2016]. These datasets offer diverse pedestrian trajectory data, capturing real-world interactions. The ETH dataset comprises videos from scenes ETH and HOTEL, while UCY includes scenes UNIV, ZARA1, and ZARA2. The SDD dataset covers six categories (pedestrian, cyclist, cart, car, skater, bus) within Stanford University. For evaluation, we employe the leave-one-out [Cawley and Talbot, 2003] strategy for ETH and UCY, and uniformly distribute SDD data among test, train, and validation sets. Our approach, following [Shi *et al.*, 2021], utilizes a 0.4-second

Methods	Publ.	ETH		UCY			Average
		ETH	HOTEL	UNIV	ZARA1	ZARA2	
Social-LSTM [Alahi *et al.*, 2016]	CVPR	1.09 / 2.35	0.79 / 1.76	0.67 / 1.40	0.47 / 1.00	0.56 / 1.17	0.72 / 1.54
Vanilla-LSTM [Alahi *et al.*, 2016]	CVPR	1.09 / 2.41	0.86 / 1.91	0.61 / 1.31	0.41 / 0.88	0.52 / 1.11	0.70 / 1.52
Social-GAN [Gupta *et al.*, 2018]	CVPR	0.87 / 1.62	0.67 / 1.37	0.76 / 1.52	0.35 / 0.68	0.42 / 0.84	0.61 / 1.21
Sophie [Sadeghian *et al.*, 2019]	CVPR	0.70 / 1.43	0.76 / 1.67	0.54 / 1.24	0.30 / 0.63	0.38 / 0.78	0.51 / 1.15
PITF [Liang *et al.*, 2019]	CVPR	0.73 / 1.65	0.30 / 0.59	0.60 / 1.27	0.38 / 0.81	0.31 / 0.68	0.46 / 1.00
GAT [Kosaraju *et al.*, 2019]	NIPS	0.68 / 1.29	0.68 / 1.40	0.57 / 1.29	0.29 / 0.60	0.37 / 0.75	0.52 / 1.07
Social-BiGAT [Kosaraju *et al.*, 2019]	NIPS	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]	CVPR	0.64 / 1.11	0.49 / 0.85	0.44 / 0.79	0.34 / 0.53	0.30 / 0.48	0.44 / 0.75
STAR [Yu *et al.*, 2020]	ECCV	0.56 / 1.11	0.26 / 0.50	0.52 / 1.15	0.41 / 0.90	0.31 / 0.71	0.41 / 0.87
RSBG w/o context [Sun *et al.*, 2020]	CVPR	0.80 / 1.53	0.33 / 0.64	0.59 / 1.25	0.40 / 0.86	0.30 / 0.65	0.48 / 0.99
SGCN [Shi *et al.*, 2021]	CVPR	0.63 / 1.03	0.32 / 0.55	**0.37** / 0.70	0.29 / 0.53	0.25 / 0.45	0.37 / 0.65
CSCNet [Xia *et al.*, 2022]	PR	**0.51** / 1.05	**0.22 / 0.42**	0.47 / 1.02	0.36 / 0.81	0.31 / 0.68	0.37 / 0.79
Social-SSL [Tsao *et al.*, 2022]	ECCV	0.69 / 1.37	0.24 / 0.44	0.51 / 0.93	0.42 / 0.84	0.34 / 0.67	0.44 / 0.85
Social-DualCVAE [Gao *et al.*, 2022]	ArXiv	0.66 / 1.18	0.34 / 0.61	0.39 / 0.74	**0.27** / 0.48	0.24 / 0.42	0.38 / 0.69
SEEM [Wang *et al.*, 2023]	TPAMI	0.62 / 1.20	0.61 / 1.21	0.50 / 1.04	0.31 / 0.61	0.36 / 0.68	0.48 / 0.95
SMGCN (ours)	-	0.63 / **1.02**	0.27 / 0.45	**0.37 / 0.66**	**0.27 / 0.47**	**0.22 / 0.40**	**0.35 / 0.60**

Table 1: **Comparison with state-of-the-art methods on the ETH and UCY dataset** under the metrics of ADE and FDE. We observe that our SMGCN demonstrates the best average error scores in both ADE / FDE metrics (lower is better).

Methods	minADE	minFDE
Linear [Li *et al.*, 2022]	37.11	63.51
Social-LSTM [Alahi *et al.*, 2016]	31.19	56.97
SF [Yamaguchi *et al.*, 2011]	36.48	58.14
CAR-Net [Sadeghian *et al.*, 2018]	25.72	51.80
Social-GAN [Gupta *et al.*, 2018]	27.25	41.44
Social-STGCNN [Mohamed *et al.*, 2020]	26.46	42.71
SMGCN (ours)	**20.89**	**36.84**

Table 2: **Comparison with state-of-the-art methods on SDD dataset** under the metrics of Minimum ADE (minADE) and Minimum FDE (minFDE) as defined in [Chang *et al.*, 2019].

observation time interval, adopting 8 frames (*i.e.*, 3.2 seconds) for training and predicting 12 subsequent frames (*i.e.*, 4.8 seconds) in trajectory prediction tasks.

Metrics. We evaluate prediction results using Average Displacement Error (ADE) [Raksincharoensak *et al.*, 2016] and Final Displacement Error (FDE) [Alahi *et al.*, 2016]. ADE is the average L-2 distance between predicted and ground-truth trajectory points, while FDE is the L-2 distance between their final destinations. For quantitative evaluation, we normalize the best prediction from 20 samples by L-2, aligning with previous work [Shi *et al.*, 2021], ensuring comparability with typical benchmark results.

Experimental settings. In our experiments, both of the embedding dimensions for graph embeddings and the self-attention mechanism are set as 64. The feature enhancement component employs an asymmetric convolutional network with 7 convolutional layers, using a kernel size of 3. We cascade 1 layer of GCN and 2 layers of TCN. The nonlinear activation function is the PRelu [He *et al.*, 2015]. This section trains the model for 600 epochs using the Adam optimizer with a batch size of 128. The attenuation factor set to 0.1 for every 50 rounds. During inference, 20 sample values were extracted from the double Gaussian distribution, and the sample values closest to the true values were used as metrics. The learning rate is set as 0.001.

4.1 Comparison with State-of-The-Arts

Single-type trajectory prediction. This section evaluates the experimental performance under the ETH and UCY dataset, selecting notable methods published in the last eight years. Results in Table 1 depict our method's superior performance over selected methods and state-of-the-art ones. Specifically, our method outperforms the existing best method SGCN [Shi *et al.*, 2021] and CSCNet [Xia *et al.*, 2022] by ↑ **5.71%** averaging on ETH and UCY datasets under the ADE metric. Our method's effectiveness lies in reducing the impact of unnecessary interactions within inner and inter groups. This is achieved through learning with our Sparse Multi-relational Spatial Graph.

Multi-type trajectory prediction. Furthermore, we assess the influence of object types on our methods through experiments on the SDD dataset, comparing the results with typical state-of-the-art methods. Table 2 displays the comparison results of ADE and FDE metrics between those methods and ours. For example, under the minADE metric, our method outperforms the previous leading method CAR-Net [Sadeghian *et al.*, 2018] by ↑ **23.12%**, and surpasses the previous second-best method Social-STGCNN [Mohamed *et al.*, 2020] by ↑ **26.66%** on the SDD datasets. The reason is that our method considers the interactions among multi-type objects, *i.e.*, dividing objects into several groups enables a more nuanced understanding of object behaviors in various scenarios, leading to substantial enhancements in the accuracy of multi-type trajectory predictions. Figure 4 displays visualization results from the SDD dataset scenario: red lines for observed, green for predicted, and blue for actual trajectories of Multi-type objects.

4.2 Ablation Study

Effect of adaptive sparse threshold η. We commence by contrasting our adaptive sparse threshold, denoted as η, with

Datasets	Scenes	MOTP-V1 ($\eta = 0.5$)	MOTP-V2 ($\eta = 0$)	**MOTP (Adaptive η)**
ETH	ETH	0.72 / 1.13	0.68 / 1.19	**0.63 / 1.02**
	HOTEL	0.39 / 0.55	0.86 / 1.37	**0.27 / 0.45**
UCY	UNIV	0.39 / 0.74	0.79 / 1.35	**0.37** / 0.66
	ZARA1	**0.27** / 0.50	0.40 / 0.65	**0.27 / 0.47**
	ZARA2	**0.22** / 0.44	0.38 / 0.65	**0.22 / 0.40**
Average		0.41 / 0.67	0.62 / 1.04	**0.35 / 0.60**

Table 3: **Impact of different sparse threshold η.** Results are reported in the forms of ADE and FDE. The adaptive η is calculated from the learnable sparse multi-relational matrix.

Variants	ETH		UCY		
	ETH	HOTEL	UNIV	ZARA1	ZARA2
w/o Relative	0.72	0.33	0.38	0.28	0.24
w/o Distance	0.69	0.35	0.40	**0.27**	0.23
w/o GTA	0.69	0.32	0.39	**0.27**	0.26
Single-layer GTA	0.67	0.30	0.40	**0.27**	0.26
SMGCN	**0.63**	**0.27**	**0.37**	**0.27**	**0.22**

Table 4: **Impact of each component of SMGCN in ADE.** The ultimate SMGCN seamlessly incorporates both relative displacement and distance relations while leveraging a multi-layer GTA.

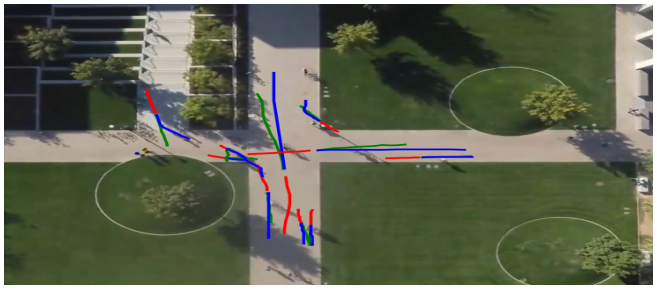

Figure 4: **Visualization of trajectory prediction.** Based on past trajectory, our model predicts future paths. Results show close alignment with actual trajectories.

Variants	ETH		UCY		
	ETH	HOTEL	UNIV	ZARA1	ZARA2
w/o Relative	1.29	0.56	0.67	0.5	0.43
w/o Distance	1.03	0.58	0.75	0.5	0.41
w/o GTA	1.27	0.55	0.73	**0.47**	0.48
Single-layer GTA	1.27	0.49	0.75	**0.47**	0.44
SMGCN	**1.02**	**0.45**	**0.66**	**0.47**	**0.40**

Table 5: **Impact of each component of SMGCN in FDE.** The ultimate SMGCN seamlessly incorporates both relative displacement and distance relations while leveraging a multi-layer GTA.

a fixed threshold ($\eta = 0.5$ or $\eta = 0$) to assess the efficacy of the proposed multi-relational sparse graph under varying sparsity levels. The outcomes are presented in Table 3. We examine two different variants of our approach: (1) MOTP-V1: exploring the impact of removing the adaptive threshold η, *i.e.* the threshold η is not fixed and setting is as $\eta = 0.5$, which means that the sparsity among object is assumed to be fixed; (2) MOTP-V2: investigating the impact of removing the sparsification, *i.e.* the threshold η is set as $\eta = 0$. By comparing the results, we find that the adaptive threshold can impact the trajectory prediction performance. Specifically, in contrast to MOTP-V1, MOTP with the adaptive threshold exhibits a notable improvement, averaging $\uparrow$ **17.05%** in ADE and $\uparrow$ **11.67%** in FDE, respectively. Furthermore, juxtaposed with MOTP-V2, MOTP with sparsification adeptly prunes superfluous interactions stemming from multi-type objects, yielding a substantial $\uparrow$ **76.7%** improvement in ADE and $\uparrow$ **73.67%** in FDE, respectively. This underscores the pivotal role of sparse directed multi-type interactions in the accurate prediction of trajectories for multi-type objects.

Effect of multi-relation spatial division. We delve into the significance of the multi-relational spatial division, examining their impact on the ADE metric, as depicted in Table 4 and Table 5. We analyze the effects of removing the relative and distance displacement relations from the multi-relation spatial graph and evaluate the performance of SMGCN. The results reveal a noteworthy enhancement across both the ETH and UCY datasets. For instance, on the challenging HOTEL scene of the ETH dataset, with our final SMGCN method showcases remarkable superiority over its variant lacking the relative displacement relation, boasting an impressive improvement ($\uparrow$ **18.18%** in ADE, $\uparrow$ **19.64%** in FDE). Furthermore, it surpasses the variant without the distance displacement relation, achieving a substantial advancement ($\uparrow$ **22.86%** in ADE, $\uparrow$ **22.41%** in FDE). This demonstrates the critical significance of both the relative and distance displacement relations.

Effect of multi-layer GTA. As illustrated in Table 4 and Table 5, our investigation extends to various configurations of GTA, encompassing scenarios both with and without GTA, utilizing a single-layer GTA, and deploying our comprehensive SMGCN framework with a multi-layer GTA.

The results divulge significant insights into the performance on the HOTEL scene of the ETH dataset. Notably, SMGCN with a multi-layer GTA exhibits enhanced performance compared to its counterpart without GTA, achieving a remarkable improvement ($\uparrow$ **15.62%** in ADE, $\uparrow$ **18.18%** in FDE). Furthermore, it surpasses the single-layer GTA variant with a substantial enhancement ($\uparrow$ **10.00%** in ADE, $\uparrow$ **8.16%** in FDE). These findings strongly indicate that, with the proposed multi-layer GTA, our SMGCN exhibits the capability to predict more accurate trajectories for multi-type objects compared to using a single GTA or none at all.

5 Conclusion

This paper studies proposes SMGCN to leverage the sparsity of multi-relation and multi-scale directed interactions and motion tendency. We conducts extensive experimental evaluations based on three public datasets, and the results show that our method can predict trajectories of multi-type objects more accurately than the typical state-of-the-art methods even under some complex scenes, such as different types of objects moving in a group and interacting with each other simultaneously. SMGCN can identify the sparse multi-relational directed interactions among multi-type objects.

Acknowledgments

This work is supported by the National Key R&D Program of China under No.2021YFC3320301.

References

[Alahi *et al.*, 2014] Alexandre Alahi, Vignesh Ramanathan, and Li Fei-Fei. Socially-aware large-scale crowd forecasting. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2211–2218, 2014.

[Alahi *et al.*, 2016] Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. Social LSTM: Human trajectory prediction in crowded spaces. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 961–971, 2016.

[Bae and Jeon, 2021] Inhwan Bae and Hae-Gon Jeon. Disentangled multi-relational graph convolutional network for pedestrian trajectory prediction. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(2):911–919, May 2021.

[Bai *et al.*, 2018] Shaojie Bai, J. Zico Kolter, and Vladlen Koltun. An empirical evaluation of generic convolutional and recurrent networks for sequence modeling, 2018.

[Cawley and Talbot, 2003] Gavin C. Cawley and Nicola L. C. Talbot. Efficient leave-one-out cross-validation of kernel fisher discriminant classifiers. *Pattern Recognit.*, 36(11):2585–2592, 2003.

[Chang *et al.*, 2019] Ming-Fang Chang, John Lambert, Patsorn Sangkloy, Jagjeet Singh, Slawomir Bak, Andrew Hartnett, De Wang, Peter Carr, Simon Lucey, Deva Ramanan, and James Hays. Argoverse: 3d tracking and forecasting with rich maps. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 8748–8757. Computer Vision Foundation / IEEE, 2019.

[Gao *et al.*, 2022] Jiashi Gao, Xinming Shi, and James J. Q. Yu. Social-dualcvae: Multimodal trajectory forecasting based on social interactions pattern aware and dual conditional variational auto-encoder. *CoRR*, abs/2202.03954, 2022.

[Gupta *et al.*, 2018] Agrim Gupta, Justin Johnson, Li Fei-Fei, Silvio Savarese, and Alexandre Alahi. Social-gan: Socially acceptable trajectories with generative adversarial networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2255–2264, 2018.

[He *et al.*, 2015] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 1026–1034, 2015.

[Hu *et al.*, 2021] Tao Hu, Chengjiang Long, and Chunxia Xiao. A novel visual representation on text using diverse conditional gan for visual recognition. *IEEE Transactions on Image Processing*, 30:3499–3512, 2021.

[Huang *et al.*, 2022] Yanjun Huang, Jiatong Du, Ziru Yang, Zewei Zhou, Lin Zhang, and Hong Chen. A survey on trajectory-prediction methods for autonomous driving. *IEEE Transactions on Intelligent Vehicles*, 7(3):652–674, 2022.

[Kipf and Welling, 2017] Thomas N. Kipf and Max Welling. Semi-supervised classification with graph convolutional networks, 2017.

[Kosaraju *et al.*, 2019] Vineet Kosaraju, Amir Sadeghian, Roberto Martín-Martín, Ian Reid, S. Hamid Rezatofighi, and Silvio Savarese. *Social-BiGAT: Multimodal Trajectory Forecasting Using Bicycle-GAN and Graph Attention Networks*, volume 13. Curran Associates Inc., Red Hook, NY, USA, 2019.

[Lerner *et al.*, 2007] Alon Lerner, Yiorgos Chrysanthou, and Dani Lischinski. Crowds by example. *Computer Graphics Forum*, 26(3), 2007.

[Li *et al.*, 2022] Ruochen Li, Stamos Katsigiannis, and Hubert P. H. Shum. Multiclass-sgcn: Sparse graph-based trajectory prediction with agent class embedding. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 2346–2350, 2022.

[Liang *et al.*, 2019] Junwei Liang, Lu Jiang, Juan Carlos Niebles, Alexander G. Hauptmann, and Li Fei-Fei. Peeking into the future: Predicting future person activities and locations in videos. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5718–5727, 2019.

[Liang *et al.*, 2022] Yuanzhi Liang, Linchao Zhu, Xiaohan Wang, and Yi Yang. A simple episodic linear probe improves visual recognition in the wild. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9549–9559, 2022.

[Lv *et al.*, 2021] Zhihan Lv, Dongliang Chen, Ranran Lou, and Qingjun Wang. Intelligent edge computing based on machine learning for smart city. *Future Generation Computer Systems*, 115:90–99, 2021.

[Mehran *et al.*, 2009] Ramin Mehran, Alexis Oyama, and Mubarak Shah. Abnormal crowd behavior detection using social force model. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 935–942, 2009.

[Mohamed *et al.*, 2020] Abduallah Mohamed, Kun Qian, Mohamed Elhoseiny, and Christian Claudel. Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14412–14420, 2020.

[Pellegrini *et al.*, 2009] Stefano Pellegrini, Andreas Ess, Konrad Schindler, and Luc Van Gool. You'll never walk alone: Modeling social behavior for multi-target tracking. In *2009 IEEE 12th International Conference on Computer Vision*, pages 261–268, 2009.

[Raksincharoensak *et al.*, 2016] Pongsathorn Raksincharoensak, Takahiro Hasegawa, and Masao Nagai.

Motion planning and control of autonomous driving intelligence system based on risk potential optimization framework. *International Journal of Automotive Engineering*, 7(1):53–60, 2016.

[Robicquet *et al.*, 2016] Alexandre Robicquet, Amir Sadeghian, Alexandre Alahi, and Silvio Savarese. Learning social etiquette: Human trajectory understanding in crowded scenes. In *European Conference on Computer Vision (ECCV)*, pages 549–565, 2016.

[Rudenko *et al.*, 2020] Andrey Rudenko, Luigi Palmieri, Michael Herman, Kris M. Kitani, Dariu M. Gavrila, and Kai Oliver Arras. Human motion trajectory prediction: A survey. *The International Journal of Robotics Research*, 39(8):895–935, 2020.

[Sadeghian *et al.*, 2018] Amir Sadeghian, Ferdinand Legros, Maxime Voisin, Ricky Vesel, Alexandre Alahi, and Silvio Savarese. Car-net: Clairvoyant attentive recurrent network. In *Computer Vision – ECCV 2018*, pages 162–180, 2018.

[Sadeghian *et al.*, 2019] Amir Sadeghian, Vineet Kosaraju, Ali Sadeghian, Noriaki Hirose, Hamid Rezatofighi, and Silvio Savarese. Sophie: An attentive gan for predicting paths compliant to social and physical constraints. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[Shi *et al.*, 2020] Xiaodan Shi, Xiaowei Shao, Zipei Fan, Renhe Jiang, and Ryosuke Shibasaki. Multimodal interaction-aware trajectory prediction in crowded space. In *In Thirty-Fourth AAAI Conference on Artificial Intelligence.*, pages 11982–11989, 2020.

[Shi *et al.*, 2021] Liushuai Shi, Le Wang, Chengjiang Long, Sanping Zhou, Mo Zhou, Zhenxing Niu, and Gang Hua. Sgcn: Sparse graph convolution network for pedestrian trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 8994–9003, June 2021.

[Sun *et al.*, 2020] Hao Sun, Zhiqun Zhao, and Zhihai He. Reciprocal learning networks for human trajectory prediction. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7414–7423, 2020.

[Tsao *et al.*, 2022] Li-Wu Tsao, Yan-Kai Wang, Hao-Siang Lin, Hong-Han Shuai, Lai-Kuan Wong, and Wen-Huang Cheng. Social-ssl: Self-supervised cross-sequence representation learning based on transformers for multi-agent trajectory prediction. In *Computer Vision – ECCV 2022*, pages 234–250, 2022.

[Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30, pages 5998–6008, Long Beach, California, 4-9 December 2017.

[Wang *et al.*, 2023] Dafeng Wang, Hongbo Liu, Naiyao Wang, Yiyang Wang, Hua Wang, and Seán F. McLoone. SEEM: A sequence entropy energy-based model for pedestrian trajectory all-then-one prediction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(1):1070–1086, 2023.

[Xia *et al.*, 2022] Beihao Xia, Conghao Wong, Qinmu Peng, Wei Yuan, and Xinge You. Cscnet: Contextual semantic consistency network for trajectory prediction in crowded spaces. *Pattern Recognition*, 126:108552, 2022.

[Yamaguchi *et al.*, 2011] Kota Yamaguchi, Alexander C. Berg, Luis E. Ortiz, and Tamara L. Berg. Who are you with and where are you going? In *2011 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1345–1352, 2011.

[Yu *et al.*, 2020] Cunjun Yu, Xiao Ma, Jiawei Ren, Haiyu Zhao, and Shuai Yi. Spatio-temporal graph transformer networks for pedestrian trajectory prediction. In *Computer Vision – ECCV 2020*, pages 507–523, 2020.

[Yuan *et al.*, 2018] Quan Yuan, Haibo Zhou, Jinglin Li, Zhihan Liu, Fangchun Yang, and Xuemin Sherman Shen. Toward efficient content delivery for automated driving services: An edge computing solution. *IEEE Network*, 32(1):80–86, 2018.

[Zhou *et al.*, 2021] Xuan Zhou, Ruimin Ke, Hao Yang, and Chenxi Liu. When intelligent transportation systems sensing meets edge computing: Vision and challenges. *Applied Sciences*, 11(20):9680, 2021.