# Learning Pedestrian Group Representations for Multi-modal Trajectory Prediction

Inhwan Bae, Jin-Hwi Park, and Hae-Gon Jeon*

AI Graduate School, GIST, South Korea
{inhwanbae, jinhwipark}@gm.gist.ac.kr, haegonj@gist.ac.kr

**Abstract.** Modeling the dynamics of people walking is a problem of long-standing interest in computer vision. Many previous works involving pedestrian trajectory prediction define a particular set of individual actions to implicitly model group actions. In this paper, we present a novel architecture named GP-Graph which has collective group representations for effective pedestrian trajectory prediction in crowded environments, and is compatible with all types of existing approaches. A key idea of GP-Graph is to model both individual-wise and group-wise relations as graph representations. To do this, GP-Graph first learns to assign each pedestrian into the most likely behavior group. Using this assignment information, GP-Graph then forms both intra- and inter-group interactions as graphs, accounting for human-human relations within a group and group-group relations, respectively. To be specific, for the intra-group interaction, we mask pedestrian graph edges out of an associated group. We also propose group pooling&unpooling operations to represent a group with multiple pedestrians as one graph node. Lastly, GP-Graph infers a probability map for socially-acceptable future trajectories from the integrated features of both group interactions. Moreover, we introduce a group-level latent vector sampling to ensure collective inferences over a set of possible future trajectories. Extensive experiments are conducted to validate the effectiveness of our architecture, which demonstrates consistent performance improvements with publicly available benchmarks. Code is publicly available at https://github.com/inhwanbae/GPGraph.

**Keywords:** Pedestrian Trajectory Prediction, Group Representation

## 1 Introduction

Pedestrian trajectory prediction attempts to forecast the socially-acceptable future paths of people based on their past movement patterns. These behavior patterns often depend on each pedestrian's surrounding environments, as well as collaborative movement, mimicking a group leader, or collision avoidance. Collaborative movement, one of the most frequent patterns, occurs when several colleagues form a group and move together. Computational social scientists estimate that up to 70% of the people in a crowd will form groups [40,48]. They

* Corresponding author

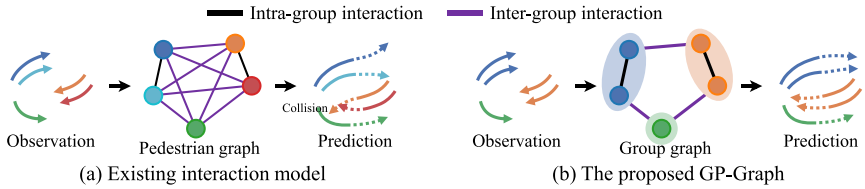



**Fig. 1.** Comparison of existing agent-agent interaction graphs and the proposed group-aware GP-Graph. To capture social interactions, (a) existing pedestrian trajectory prediction models each pedestrian on a graph node. Since the pedestrian graph is a complete graph, it is difficult to capture the group's movement because it becomes overly complex in a crowded scene. (b) GP-Graph is directly able to learn an intra-/inter-group interaction while keeping the agent-wise structure.

also gather surrounding information and have the same destination [40]. Such groups have characteristics that are distinguishable from those of individuals, maintain rather stable formations, and even provide important cues that can be used for future trajectory prediction [48,78].

Pioneering works in human trajectory forecasting model the group movement by assigning additional hand-crafted terms as energy potentials [41,66,47]. These works account for the presence of other group members and physics-based attractive forces, which are only valid between the same group members. In recent works, convolutional neural networks (CNNs) and graph neural networks (GNNs) show impressive progress modeling the social interactions, including traveling together and collision avoidance [1,17,39,2,54]. Nevertheless, trajectory prediction is still a challenging problem because of the complexity of implicitly learning individual and group behavior at once.

There are several attempts that explicitly encode the group coherence behaviors by assigning hidden states of LSTM with a summation of other agents' states, multiplied by a binary group indicator function [6]. However, existing studies have a critical problem when it comes to capturing the group interaction. Since their forecasting models focus more on individuals, the group features are shared at the individual node as illustrated in Fig. 1(a). Although this approach can conceptually capture group movement behavior, it is difficult for the learning-based methods to represent it because of the overwhelming number of edges for the individual interactions. And, this problem is increasingly difficult in crowded environments.

To address this issue, we propose a novel general architecture for pedestrian trajectory prediction: GrouP-Graph (GP-Graph). As illustrated in Fig. 1(b), our GP-Graph captures intra-(members in a group) and inter-group interactions by disentangling input pedestrian graphs. Specifically, our GP-Graph first learns to assign each pedestrian into the most likely behavior group. The group indices of each pedestrian are generated using a pairwise distance matrix. To make the indexing process end-to-end trainable, we introduce a straight-through group back-propagation trick inspired by the Straight-Through estimator [5,21,35]. Using the group information, GP-graph then transforms the input pedestrian graph into both intra- and inter-group interaction graphs. We construct the intra-

group graph by masking out edges of the input pedestrian graph for unassociated group members. For the inter-group graph, we propose group pooling&unpooling operations to represent a group with multiple members as one graph node. By applying these processes, GP-Graph architecture has three advantages: (1) It reduces the complexity of trajectory prediction which is caused by the different social behaviors of individuals, by modeling group interactions. (2) It alleviates inherent scene bias by considering the huge number of unseen pedestrian graph nodes between the training and test environments, as discussed in [8]. (3) It offers a graph augmentation effect with pedestrian node grouping.

Next, through weight sharing with baseline trajectory predictors, we force a hierarchy representation from both the input pedestrian graph and the disentangled interactions. This representation is used to infer a probability map for socially-acceptable future trajectories after passing through our group integration module. In addition, we introduce a group-level latent vector sampling to ensure collective inferences over a set of plausible future trajectories.

To the best of our knowledge, this is the first model that literally pools pedestrian colleagues into one group node to efficiently capture group motion behaviors, and learns pedestrian grouping in an end-to-end manner. Furthermore, GP-Graph has the best performance on various datasets among existing methods when unifying with GNN-based models, and it can be integrated with all types of trajectory prediction models, achieving consistent improvements. We also provide extensive ablation studies to analyze and evaluate our GP-Graph.

## 2 Related Works

### 2.1 Trajectory Prediction

Earlier works [18,42,38,66] model human motions in crowds using hand-crafted functions to describe attractive and repulsive forces. Since then, pedestrian trajectory prediction has been advanced by research interest in computer vision. Such research leverages the impressive capacity of CNNs which can capture social interactions between surrounding pedestrians. One pioneering work is Social-LSTM [1], which introduces a social pooling mechanism considering a neighbor's hidden state information inside a spatial grid. Much of the emphasis in subsequent research has been to add human-environment interactions from a surveillance view perspective [49,33,23,58,11,75,61,59,37,52]. Instead of taking environmental information into account, some methods directly share hidden states of agents between other interactive agents [17,64,50]. In particular, Social-GAN [17] takes the interactions via max-pooling in all neighborhood features in the scene, and Social-Attention [64] introduces an attention mechanism to impose a relative importance on neighbors and performs a weighted aggregation for the features.

In terms of graph notations, each pedestrian and their social relations can be represented as a node and an edge, respectively. When predicting pedestrian trajectories, graph representation is used to model social interactions with graph convolutional networks (GCNs) [22,39,59,2], graph attention networks (GATs) [63,19,23,32,54,3], and transformers [69,70,16]. Usually, these approaches infer future paths through recurrent estimations [1,17,50,74,9,26,16]

or extrapolations [39,2,54,31]. Other types of relevant research are based on probabilistic inferences for multi-modal trajectory prediction using Gaussian modeling [1,2,39,55,69,30,54,65], generative models [17,49,23,75,58,11,19], and a conditional variational autoencoder [27,29,20,50,36,9,60,26]. We note that these approaches focus only on learning implicit representations for group behaviors from agent-agent interactions.

### 2.2 Group-aware Representation

Contextual and spatial information can be derived from group-aware representations of agent dynamics. To accomplish this, one of the group-aware approaches is social grouping, which describes agents in groups that move differently than independent agents.

In early approaches [76,24,77], pedestrians can be divided into several groups based on behavior patterns. To represent the collective activities of agents in a supervised manner, a work in [41] exploits conditional random fields (CRF) to jointly predict the future trajectories of pedestrians and their group membership. Yamaguchi *et al*. [66] harness distance, speed, and overlap time to train a linear SVM to classify whether two pedestrians are in the same group or not. In contrast, a work in [14] proposes automatic detection for small groups of individuals using a bottom-up hierarchical clustering with speed and proximity features.

Group-aware predictors recognize the affiliations and relations of individual agents, and encode their proper reactions to moving groups. Several physics-based techniques represent group relations by adding attractive forces among group members [66,41,46,40,44,51,56]. Although a dominant learning paradigm [1,73,43,62,4] implicitly learns intra- and inter-group coherency, only two works in [6,12] explicitly define group information. To be specific, one [6] identifies pedestrians walking together in the crowd using a coherent filtering algorithm [77], and utilizes the group information in a social pooling layer to share their hidden states. Another work [12] proposes a generative adversarial model (GAN)-based trajectory model, jointly learning informative latent features for simultaneous pedestrian trajectory forecasting and group detection. These approaches only learn individual-level interactions within a group, but do not encode their affiliated groups and future paths at the same time. Unlike them, our GP-Graph aggregates a group-group relation via a novel group pooling in the proposed end-to-end trainable architecture without any supervision.

### 2.3 Graph Node Pooling

Pooling operations are used for features extracted from grid data, like images, as well as graph-structured data. However, there is no geographic proximity or order information in the graph nodes that existing pooling operations require. As alternative methods, three types of graph pooling are introduced: topology-based pooling [10,45], global pooling [15,72], and hierarchical pooling [7,13,68]. These approaches are designed for general graph structures. However, since human behavior prediction has time-variant and generative properties, it is no possible to leverage the advantages of these pooling operations for this task.

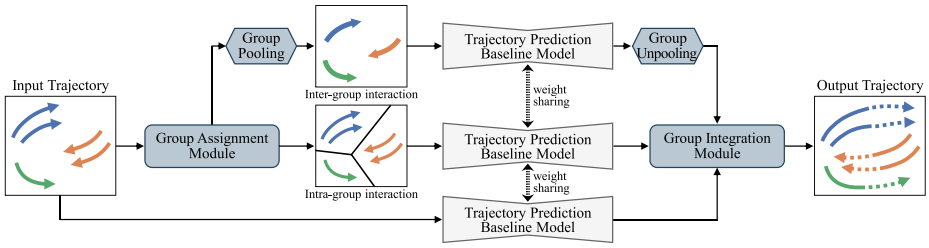



**Fig. 2.** An overview of our GP-Graph architecture. Starting with graph-structured trajectories for $N$ pedestrians, we first estimate grouping information with the Group Assignment Module. We then generate both intra-/inter-group interaction graphs by masking out unrelated nodes and by performing pedestrian group pooling. The weight-shared trajectory prediction model takes the three types of graphs and capture group-aware social interactions. Group pooling operators are then applied to encode agent-wise features from group-wise features, and then fed into the Group Integration Module to estimate the probability distribution for future trajectory prediction.

## 3 Proposed Method

In this work, we focus on how group awareness in crowds is formed for pedestrian trajectory prediction. We start with a definition of a pedestrian graph and trajectory prediction in Sec. 3.1. We then introduce our end-to-end learnable pedestrian group assignment technique in Sec. 3.2. Using group index information and our novel pedestrian group pooling&unpooling operations, we construct a group hierarchy representation of pedestrian graphs in Sec. 3.3. The overall architecture of our GP-Graph is illustrated in Fig. 2.

### 3.1 Problem Definition

Pedestrian trajectory prediction can be defined as a sequential inference task made observations for all agents in a scene. Suppose that $N$ is the number of pedestrians in a scene, the history trajectory of each pedestrian $n \in [1, ..., N]$ can be represented as $\boldsymbol{X}_n = \{(x_n^t, y_n^t) \,|\, t \in [1, ..., T_{obs}]\}$, where the $(x_n^t, y_n^t)$ is the 2D spatial coordinate of a pedestrian $n$ at specific time $t$. Similarly, the ground truth future trajectory of pedestrian $n$ can be defined as $\boldsymbol{Y}_n = \{(x_n^t, y_n^t) \,|\, t \in [T_{obs}+1, ..., T_{pred}]\}$.

The social interactions are modeled from the past trajectories of other pedestrians. In general, the pedestrian graph $\mathcal{G}_{ped} = (\mathcal{V}_{ped}, \mathcal{E}_{ped})$ refers to a set of pedestrian nodes $\mathcal{V}_{ped} = \{\boldsymbol{X}_n \,|\, n \in [1, ..., N]\}$ and edges on their pairwise social interaction $\mathcal{E}_{ped} = \{e_{i,j} \,|\, i, j \in [1, ..., N]\}$. The trajectory prediction process forecasts their future sequences based on their past trajectory and the social interaction as:

$$\widehat{\boldsymbol{Y}} = F_\theta \left(X, \mathcal{G}_{ped}\right) \tag{1}$$

where $\widehat{\boldsymbol{Y}} = \{\widehat{\boldsymbol{Y}}_n \,|\, n \in [1, ..., N]\}$ denotes the estimated future trajectories of all pedestrians in a scene, and $F_\theta(\,\cdot\,)$ is the trajectory generation network.

### 3.2 Learning the Trajectory Grouping Network

Our goal in this work is to encode powerful group-wise features beyond existing agent-wise social interaction aggregation models to achieve highly accurate human trajectory prediction. The group-wise features represent group members in input scenes as single nodes, making pedestrian graphs simpler. We use a U-Net architecture with pooling layers to encode the features on graphs. By reducing the number of nodes through the pooling layers in the U-Net, higher-level group-wise features can be obtained. After that, agent-wise features are recovered through unpooling operations.

Unlike conventional pooling&unpooling operators working on grid-structured data, like images, it is not feasible to apply them to graph-structured data. Some earlier works to handle this issue [7,13]. The works focus on capturing global information by removing relatively redundant nodes using a graph pooling, and restoring the original shapes by adding dummy nodes from a graph unpooling if needed. However, in pedestrian trajectory prediction, each node must keep its identity index information and describe the dynamic property of the group behavior in scenes. For that, we present pedestrian graph-oriented group pooling&unpooling methods. We note that it is the first work to exploit the pedestrian index itself as a group representation.

**Learning pedestrian grouping.** First of all, we estimate grouping information to which the pedestrian belongs using a Group Assignment Module. Using the history trajectory of each pedestrian, we measure the feature similarity among all pedestrian pairs based on their $L_2$ distance. With this pairwise distance, we pick out all pairs of pedestrians that are likely to be a colleague (affiliated with same group). The pairwise distance matrix $\boldsymbol{D}$ and a set of colleagues indices $\Upsilon$ are defined as:

$$\boldsymbol{D}_{i,j} = \|F_\phi(\boldsymbol{X}_i) - F_\phi(\boldsymbol{X}_j)\| \quad \text{for} \;\; i,j \in [1,...,N], \tag{2}$$

$$\Upsilon = \{\text{pair}(i,\, j) \,|\, i,j \in [1,...,N],\; i \neq j,\; \boldsymbol{D}_{i,j} \leq \pi\}, \tag{3}$$

where $F_\phi(\,\cdot\,)$ is a learnable convolutional layer and $\pi$ is a learnable thresholding parameter.

Next, using the pairwise colleague set $\Upsilon$, we arrange the colleague members in associated groups and assign their group index. We make a group index set $G$, which is formulated as follows:

$$G = \Big\{ G_k \,|\, G_k = \bigcup_{(i,j)\in\Upsilon} \{i,\, j\},\;\; G_a \cap G_b = \varnothing \;\; \text{for } a \neq b \Big\} \tag{4}$$

where $G_k$ denotes the $k$-th group and is the union of each pair set $(i, j)$. This information is used as important prior knowledge in the subsequent pedestrian group pooling and unpooling operators.

**Pedestrian group pooling.** Based on the group behavior property that group members gather surrounding information and share behavioral patterns, we group the pedestrian nodes, where the corresponding node's features are aggregated into

one node. The aggregated group features are then stacked for subsequent social interaction capturing modules (*i.e.* GNNs). Here, the most representative feature for each pedestrian node is selected via an average pooling. With the feature, we can model the group-wise graph structures, which have much fewer number of nodes than the input pedestrian graph, as will be demonstrated in Sec. 4.3. We define the pooled group-wise trajectory feature $\boldsymbol{Z}$ as follows:

$$\boldsymbol{Z} = \{\boldsymbol{Z}_k \,|\, k \in [1, ..., K]\}, \qquad \boldsymbol{Z}_k = \frac{1}{|G_k|} \sum_{i \in G_k} \boldsymbol{X}_i, \tag{5}$$

where $K$ is the total group numbers in $G$.

**Pedestrian group unpooling.** Next, we upscale the group-wise graph structures back to their original size by using an unpooling operation. This enables each pedestrian trajectory to be forecast with output agent-wise feature fusion information. In existing methods [7,13], zero vector nodes are appended into the group features during unpooling. The output of the convolution process on the zero vector nodes fails to exhibit the group properties. To alleviate this issue, we duplicate the group features and then assign them into nodes for all the relevant group members so that they have identical group behavior information. The pedestrian group unpooling operator can be formulated as follows:

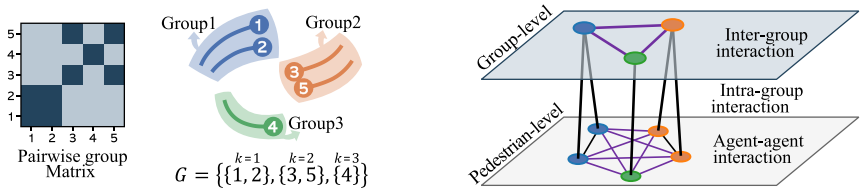
$$\overline{\boldsymbol{X}} = \{\overline{\boldsymbol{X}}_n \,|\, n \in [1, ..., N]\}, \qquad \overline{\boldsymbol{X}}_n = \boldsymbol{Z}_k \quad \text{where} \quad n \in G_k, \tag{6}$$

where $\overline{\boldsymbol{X}}$ is the agent-wise trajectory feature reconstructed from $Z$, having the same order of pedestrian indices as in $\boldsymbol{X}$.

**Straight-Through Group Estimator.** A major hurdle, when training the group assignment module in Eq. (4) which is a sampling function, is that index information is not treated as learnable parameters. Accordingly, the group index cannot be trained using standard backpropagation algorithms. The reason is why the existing methods utilize separate training steps from main trajectory prediction networks for the group detection task.

We tackle this problem by introducing a Straight-through (ST) trick, inspired by the biased path derivative estimators in [5,21,35]. Instead of making the discrete index set $G_k$ differentiable, we separate the forward pass and backward pass of the group assignment module in the training process. Our intuition for constructing the backward pass is that group members have similar features with closer pairwise distance between colleagues.

In the forward pass, we perform our group pooling over both pedestrian features and the group index from the input trajectory and estimated group assignment information, respectively. For the backward pass, we propose group-wise continuous relaxed features to approximate the group indexing process. We compute the probability that a pair of pedestrians belongs to the same group using the proposed differentiable binary thresholding function $\frac{1}{1+\exp(x-\pi)}$, and apply it on the pairwise distance matrix $\boldsymbol{D}$. We then measure the normalized probability $\boldsymbol{A}$ of the summation of all neighbors' probability. Lastly, we compute a new pedestrian trajectory feature $\boldsymbol{X}'$ by aggregating features between group




(a) Pedestrian group assignment (b) Pedestrian group hierarchy graph

**Fig. 3.** An illustration of our pedestrian group assignment method using a pairwise group probability matrix $A$. With a group index set $G$, a pedestrian group hierarchy is constructed based on three types of interaction graphs.

members through the matrix multiplication of $\boldsymbol{X}$ and $\boldsymbol{A}$ as follows:

$$\boldsymbol{A}_{i,j} = \frac{\frac{1}{1+\exp\left(\frac{\boldsymbol{D}_{i,j}-\pi}{\tau}\right)}}{\sum_{i=1}^{N}\left(\frac{1}{1+\exp\left(\frac{\boldsymbol{D}_{i,j}-\pi}{\tau}\right)}\right)} \quad \text{for} \quad i,j \in [1,...,N], \tag{7}$$

$$\boldsymbol{X}' = \langle\, \boldsymbol{X} - \boldsymbol{X}\boldsymbol{A} \,\rangle + \boldsymbol{X}\boldsymbol{A}, \tag{8}$$

where $\tau$ is the temperature of the sigmoid function and $\langle\cdot\rangle$ is the *detach* (in PyTorch) or *stop gradient* (in Tensorflow) function which prevents the backpropagation.

For further explanation of Eq. (8), we replace the input of pedestrian group pooling module $X$ with a new pedestrian trajectory feature $\boldsymbol{X}'$ in implementation. To be specific, we can remove $\boldsymbol{X}\boldsymbol{A}$ in the forward pass, allowing us to compute a loss for the trajectory feature $\boldsymbol{X}$. In contrast, due to the stop gradient $\langle\cdot\rangle$, the loss is only backpropagated to $\boldsymbol{X}\boldsymbol{A}$ in the backward pass. To this end, we can train both the convolutional layer $F_\phi$ and the learnable threshold parameter $\pi$ which are used for the computation of the pairwise distance matrix $\boldsymbol{D}$ and the construction of group index set $G$, respectively.

### 3.3 Pedestrian Group Hierarchy Architecture

Using the estimated pedestrian grouping information, we reconstruct the initial social interaction graph $\mathcal{G}_{ped}$ in an efficient form for pedestrian trajectory prediction. Instead of the existing complex and complete pedestrian graph, intra- and inter-group interaction graphs capture the group-ware social relation, as illustrated in Fig. 3.

**Intra-group interaction graph.** We design a pedestrian interaction graph that captures relations between members affiliated with the same group. The intra-group interaction graph $\mathcal{G}_{member} = (\mathcal{V}_{ped}, \mathcal{E}_{member})$ consists of a set of pedestrian nodes $\mathcal{V}_{ped}$ and edges on their pairwise social interaction of group members $\mathcal{E}_{member} = \{e_{i,j} \,|\, i,j \in [1,...,N], k \in [1,...,K], \{i,j\} \subset G_k\}$. Through this graph representation, pedestrian nodes can learn social norms of internal collision avoidance between group members while maintaining their own formations and on-going directions.

**Inter-group interaction graph.** Inter-group interactions (group-group relation) are indispensable to learn social norms between groups as well. To

take various group behaviors such as following a leading group, avoiding collisions and joining a new group, we create an inter-group interaction graph $\mathcal{G}_{group} = (\mathcal{V}_{group}, \mathcal{E}_{group})$. Here, nodes refer to each group's features $\mathcal{V}_{group} = \{\overline{\boldsymbol{X}}_k \,|\, k \in [1, ..., K]\}$ generated with our pedestrian group pooling operation, and edges mean the pairwise group-group interactions $\mathcal{E}_{group} = \{\bar{e}_{p,q} \,|\, p, q \in [1, ..., K]\}$.

**Group integration network.** We incorporate the social interactions as a form of group hierarchy into well-designed existing trajectory prediction baseline models in Fig. 3(b). Meaningful features can be extracted by feeding a different type of graph-structured data into the same baseline model. Here, the baseline models share their weights to reduce the amount of parameters while enriching the augmentation effect. Afterward, the output features from the baseline models are aggregated agent-wise, and are then used to predict the probability map of future trajectories using our group integration module. The generated output trajectory $\widehat{Y}$ with the group integration network $F_\psi$ is formulated as:

$$\widehat{Y} = F_\psi\big(\underbrace{F_\theta(X, \mathcal{G}_{ped})}_{\text{Agent-wise GNN}}, \underbrace{F_\theta(X, \mathcal{G}_{member})}_{\text{Intra-group GNN}}, \underbrace{F_\theta(\overline{X}, \mathcal{G}_{group})}_{\text{Inter-group GNN}}\big). \tag{9}$$

**Group-level latent vector sampling.** To infer the multi-modal future paths of pedestrians, an additional random latent vector is introduced with an input observation path. This latent vector becomes a factor, determining a person's choice of behavior patterns, such as acceleration/deceleration and turning to right/left. There are two ways to adopt this latent vector in trajectory generation: (1) Scene-level sampling [17] where everyone in the scene shares one latent vector, unifying the behavior patterns of all pedestrians in a scene (*e.g.*, all pedestrians are slow down); (2) Pedestrian-level sampling [50] that allocates the different latent vectors for each pedestrian, but forces the pedestrians to have different patterns, where the group behavior property is lost.

We propose a group-level latent vector sampling method as a compromise of the two ways. We use the group information estimated from the GP-Graph to share the latent vector between groups. If two people are not associated with the same group, an independent random noise is assigned as a latent vector. In this way, it is possible to sample a multi-modal trajectory, which is independent of other groups members and follows associated group behaviors. The effectiveness of the group-level sampling is visualized in Sec. 4.3.

### 3.4 Implementation Details

To validate the generality of our GP-Graph, we incorporate it into four state-of-the-art baselines: three different GNN-based baseline methods including STGCNN (GCN-based) [39], SGCN (GAT-based) [54] and STAR (Transformer-based) [69], and one non-GNN model, PECNet [36]. We simply replace their trajectory prediction parts with ours. We additionally embed our agent/intra-/inter-graphs on the baseline networks, and compute integrated output trajectories to obtain the group-aware prediction.

For our proposed modules, we initialize the learnable parameter $\pi$ as one, which cut the total number of nodes moderately down by half, with the group

pooling in the initial training step. Other learnable parameters such as $F_\theta$, $F_\phi$ and $F_\psi$ are randomly initialized. We set the hyperparameter $\tau$ to 0.1 to give the binary thresholding function a steep slope.

To train the GP-Graph architecture, we use the same training hyperparameters (*e.g.*, batch size, train epochs, learning rate, learning rate decay), loss functions, and optimizers of the baseline models. We note that we do not use additional group labels for an apple-to-apple comparison with the baseline models. Our group assignment module is trained to estimate effective groups for trajectory prediction in an unsupervised manner. Thanks to our powerful Straight-Through Group Estimator, it accomplish promising results over other supervised group detection networks [7] that require additional group labels.

## 4 Experiments

In this section, we conduct comprehensive experiments to verify how the grouping strategy contributes to pedestrian trajectory prediction. We first briefly describe our experimental setup (Sec. 4.1). We then provide comparison results with various baseline models for both group detection and trajectory prediction (Sec. 4.3 and Sec. 4.2). We lastly conduct an extensive ablation study to demonstrate the effect of each component of our method (Sec. 4.4).

### 4.1 Experimental Setup

**Datasets.** We evaluate the effectiveness of our GP-Graph by incorporating it into several baseline models and check the performance improvement on public datasets: ETH [42], UCY [28], Stanford Drone Dataset (SDD) [47], and the Grand Central Station (GCS) [67] datasets. The ETH & UCY datasets contain five unique scenes (ETH, Hotel, Univ, Zara1 and Zara2) with 1,536 pedestrians, and the official leave-one-out strategy is used to train and to validate the models. SDD consists of various types of objects with a birds-eye view, and GCS shows highly congested pedestrian walking scenes. We use the standard training and evaluation protocol [17,19,39,54,50,36] in which the first 3.2 seconds (8 frames) are observed and next 4.8 seconds (12 frames) are used for a ground truth trajectory. Additionally, two scenes (Seq-eth, Seq-hotel) of the ETH datasets provide ground-truth group labels. We use them to evaluate how accurately our GP-Graph groups individual pedestrians.

**Evaluation protocols.** For multi-modal human trajectory prediction, we follow a standard evaluation manner, in Social-GAN [17], generating 20 samples based on predicted probabilistic distributions, and then choosing the best sample to measure the evaluation metrics. We use same evaluation metrics of previous works [1,17,61,34] for future trajectory prediction. Average Displacement Error (ADE) computes the Euclidean distance between a prediction and ground-truth trajectory, while Final Displacement Error (FDE) computes the Euclidean distance between an end-point of prediction and ground-truth. Collision rate (COL) checks the percentage of test cases where the predicted trajectories of different agents run into collisions, and Temporal Correlation Coefficient (TCC) measures the Pearson correlation coefficient of motion patterns between a predicted and

| | STGCNN | | | | **GP-Graph - STGCNN** | | | | | SGCN | | | | **GP-Graph - SGCN** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ADE↓ | FDE↓ | COL↓ | TCC↑ | ADE↓ | FDE↓ | COL↓ | TCC↑ | Gain↑ | ADE↓ | FDE↓ | COL↓ | TCC↑ | ADE↓ | FDE↓ | COL↓ | TCC↑ | Gain↑ |
| ETH | 0.73 | 1.21 | 1.80 | 0.47 | **0.48** | **0.77** | **1.15** | **0.63** | 36.4% | 0.63 | 1.03 | 1.69 | 0.55 | **0.43** | **0.63** | **1.35** | **0.65** | 38.8% |
| HOTEL | 0.41 | 0.68 | 3.94 | 0.28 | **0.24** | **0.40** | **2.00** | **0.32** | 41.2% | 0.32 | 0.55 | 2.52 | 0.29 | **0.18** | **0.30** | **0.66** | **0.35** | 45.5% |
| UNIV | 0.49 | 0.91 | 9.69 | 0.63 | **0.29** | **0.47** | **7.54** | **0.77** | 48.4% | 0.37 | 0.70 | 6.85 | 0.69 | **0.24** | **0.42** | **5.52** | **0.80** | 40.0% |
| ZARA1 | 0.33 | 0.52 | 2.54 | 0.71 | **0.24** | **0.40** | **2.13** | **0.82** | 23.1% | 0.29 | 0.53 | 0.79 | 0.74 | **0.17** | **0.31** | **0.62** | **0.86** | 41.5% |
| ZARA2 | 0.30 | 0.48 | 7.15 | 0.39 | **0.23** | **0.40** | **3.80** | **0.49** | 16.7% | 0.25 | 0.45 | 2.23 | 0.49 | **0.15** | **0.29** | **1.44** | **0.56** | 35.6% |
| AVG | 0.45 | 0.76 | 5.02 | 0.50 | **0.29** | **0.49** | **3.32** | **0.60** | 35.5% | 0.37 | 0.65 | 2.82 | 0.55 | **0.23** | **0.39** | **1.92** | **0.64** | 40.0% |
| SDD | 20.8 | 33.2 | 6.79 | 0.47 | **10.6** | **20.5** | **4.36** | **0.67** | 38.3% | 25.0 | 41.5 | 4.45 | 0.57 | **15.7** | **32.5** | **2.59** | **0.60** | 21.7% |
| GCS | 14.7 | 23.9 | 3.92 | 0.70 | **11.5** | **19.3** | **1.24** | **0.73** | 19.2% | 11.2 | 20.7 | 1.45 | 0.78 | **7.8** | **13.7** | **0.67** | **0.79** | 33.8% |

| | STAR | | | | **GP-Graph - STAR** | | | | | PECNet | | | | **GP-Graph - PECNet** | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ADE↓ | FDE↓ | COL↓ | TCC↑ | ADE↓ | FDE↓ | COL↓ | TCC↑ | Gain↑ | ADE↓ | FDE↓ | COL↓ | TCC↑ | ADE↓ | FDE↓ | COL↓ | TCC↑ | Gain↑ |
| ETH | **0.36** | 0.65 | 1.46 | 0.72 | 0.37 | **0.58** | **0.88** | **0.77** | 11.0% | 0.64 | 1.13 | 3.08 | 0.58 | **0.56** | **0.82** | **2.38** | **0.59** | 27.3% |
| HOTEL | 0.17 | 0.36 | 1.51 | **0.32** | **0.16** | **0.24** | **1.46** | 0.31 | 32.2% | 0.22 | 0.38 | 5.69 | 0.33 | **0.18** | **0.26** | **3.45** | **0.34** | 32.1% |
| UNIV | **0.31** | 0.62 | 1.95 | 0.69 | **0.31** | **0.57** | **1.65** | **0.73** | 7.4% | 0.35 | 0.57 | 3.80 | 0.75 | **0.31** | **0.46** | **2.89** | **0.77** | 19.5% |
| ZARA1 | 0.26 | 0.55 | 1.55 | 0.73 | **0.24** | **0.44** | **1.39** | **0.82** | 20.3% | 0.25 | 0.45 | 2.99 | 0.80 | **0.23** | **0.40** | **2.57** | **0.82** | 11.7% |
| ZARA2 | 0.22 | 0.46 | 1.46 | **0.50** | **0.21** | **0.39** | **1.27** | 0.46 | 14.3% | 0.18 | 0.31 | 4.91 | 0.55 | **0.17** | **0.27** | **2.92** | **0.58** | 13.0% |
| AVG | **0.26** | 0.53 | 1.59 | 0.59 | **0.26** | **0.44** | **1.33** | **0.62** | 15.7% | 0.33 | 0.60 | 4.09 | 0.61 | **0.29** | **0.44** | **2.84** | **0.62** | 26.4% |
| SDD | 14.9 | 28.2 | 0.72 | 0.59 | **13.7** | **25.2** | **0.35** | **0.61** | 10.4% | 10.0 | 15.8 | **0.22** | 0.64 | **9.1** | **13.8** | 0.23 | **0.65** | 12.7% |
| GCS | 15.6 | 31.8 | 1.79 | **0.80** | **14.9** | **30.3** | **0.81** | **0.80** | 4.8% | 17.1 | 29.3 | 0.20 | 0.71 | **14.2** | **23.9** | **0.19** | **0.72** | 18.4% |

**Table 1.** Comparison between GP-Graph architecture and the vanilla agent-wise interaction graph for four state-of-the-art multi-modal trajectory prediction models, Social-STGCNN [39], SGCN [54], STAR [69] and PECNet [36]. The models are evaluated on the ETH [42], UCY [28], SDD [47] and GCS [67] datasets. Gain: performance improvement w.r.t FDE over the baseline models, Unit for ADE and FDE: meter, **Bold**: Best.

ground-truth trajectory. We use both ADE and FDE as accuracy measures, and both COL and TCC as reliability measures in our group-wise prediction. For the COL metric, we average a set of collision ratios over the 20 multi-modal samples.

For grouping measures, we use precision and recall values based on two popular metrics, proposed in prior works [6,12]: A group pair score (PW) measures the ratio between group pairs that disagree on their cluster membership, and all possible pairs in a scene. A Group-MITRE score (GM) is a ratio of the minimum number of links for group members and fake counterparts for pedestrians who are not affiliated with any group.

### 4.2 Quantitative Results

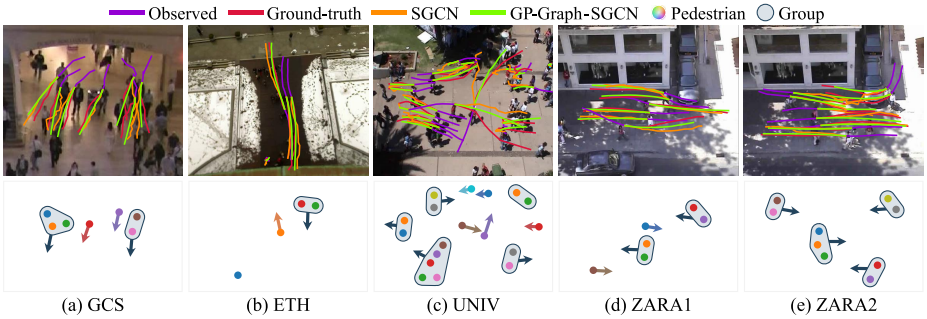
**Evaluation on trajectory prediction.** We first compare our GP-Graph with conventional agent-wise prediction models on the trajectory prediction benchmarks. As reported in Table 1, our GP-Graph achieves consistent performance improvements on all the baseline models. Additionally, our group-aware prediction also reduces the collision rate between agents, and shows analogous motion patterns with its ground truth by capturing the group movement behavior well. The results demonstrate that the trajectory prediction models benefit from the group-awareness cue of our group assignment module.

**Evaluation on group estimation.** We also compare the grouping ability of our GP-Graph with that of state-of-the-art models in Table 2. Our group assignment module trained in an unsupervised manner achieves superior results in the PW precision in both scenes, but shows relatively low recall values over the baseline models.

There are various group interaction scenarios in both scenes, and we found that our model sometimes fails to assign pedestrians into one large group when either

| | | Shao *et al.*[53] | Zanotto *et al.*[71] | Yamaguchi *et al.*[66] | Ge *et al.*[14] | Solera *et al.*[57] | Fernando *et al.*[12] | GP-Graph | GP-Graph+$\mathcal{S}$ |
|---|---|---|---|---|---|---|---|---|---|
| Seq-eth | PW↑ | 44.5 / **87.0** | 79.0 / 82.0 | 72.9 / 78.0 | 80.7 / 80.7 | 91.1 / 83.4 | 91.3 / 83.5 | **91.7** / 82.1 | 91.1 / 84.1 |
| | GM↑ | 69.3 / 68.2 | - / - | 60.6 / 76.4 | 87.0 / 84.2 | 91.3 / **94.2** | **92.5** / **94.2** | 86.9 / 86.8 | **92.5** / 91.3 |
| Seq-hotel | PW↑ | 51.5 / 90.4 | 81.0 / 91.0 | 83.7 / 93.9 | 88.9 / 89.3 | 89.1 / 91.9 | 90.2 / 93.1 | **91.5** / 80.1 | 90.4 / **93.3** |
| | GM↑ | 67.3 / 64.1 | - / - | 84.0 / 51.2 | 89.2 / 90.9 | 97.3 / **97.7** | **97.5** / **97.7** | 84.5 / 80.0 | 96.1 / 96.0 |

**Table 2.** Comparison of GP-Graph on SGCN with other state-of-the-art group detection models (Precision/Recall). For fair comparison, the evaluation results are directly referred from [6,12]. $\mathcal{S}$: Use a loss for supervision, **Bold**: Best, Underline: Second best.




(a) GCS (b) ETH (c) UNIV (d) ZARA1 (e) ZARA2

**Fig. 4.** (Top): Examples of pedestrian trajectory prediction results. (Bottom): Examples of group estimation results on ETH/UCY datasets [42,28].

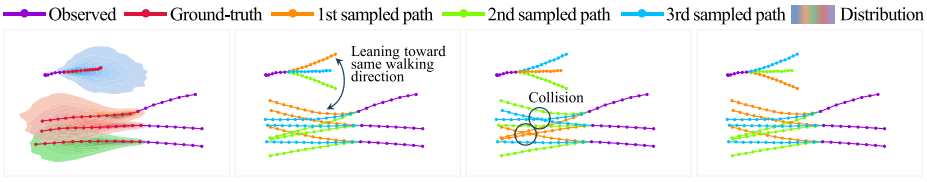
a person joins the group or the group splits into both sides to avoid a collision. In this situation, while forecasting agent-wise trajectories, it is advantageous to divide the group into sub-groups or singletons, letting them have different behavior patterns. Although false-negative group links sometimes occur during the group estimation because of this, it is not a big issue for trajectory prediction.

To measure the maximum capability of our group estimator, we additionally carry out an experiment with a supervision loss to reduce the false-negative group links. We use a binary cross-entropy loss between the distance matrix and the ground-truth group label. As shown in Table 2, the performance is comparable to the state-of-the art group estimation models with respect to the PW and GM metrics. This indicates that our learning trajectory grouping network can properly assign groups without needing complex clustering algorithms.

### 4.3 Qualitative Results

**Trajectory visualization.** In Fig. 4, we visualize some prediction results of GP-Graph and other methods. Since GP-Graph estimates the group-aware representations and captures both intra-/inter-group interactions, the predicted trajectories are closer to socially-acceptable trajectories and forms more stable behaviors between group members than those of the comparison models. Fig. 4 also shows the pedestrians forming a group with our group assignment module. GP-Graph uses movement patterns and proximity information to properly create a group node for pedestrians who will take the same behaviors and walking directions in the future. This simplifies complex pedestrian graphs and eliminates potential errors associated with the collision avoidance between colleagues.

**Group-level latent vector sampling.** To demonstrate the effectiveness of the group-level latent vector sampling strategy, we compare ours with two previous




(a) Output probability (b) Scene-level sampling (c) Pedestrian-level sampling (d) group-level sampling

**Fig. 5.** (a) Visualization of predicted trajectory distribution in ZARA1 scene. (b,c,d) Examples of three sampled trajectories with scene-level, pedestrian-level, and group-level latent vector sampling strategy.

| | ETH | HOTEL | UNIV | ZARA1 | ZARA2 | AVG |
|---|---|---|---|---|---|---|
| w/o Pool&Unpool | 1.03 / 1.69 / 0.55 | 0.55 / 2.52 / 0.29 | 0.70 / 6.85 / 0.69 | 0.53 / 1.79 / 0.74 | 0.45 / 2.23 / 0.49 | 0.65 / 3.02 / 0.55 |
| gPool&gUnpool [13] | 0.73 / 1.88 / 0.66 | 0.44 / 1.78 / **0.35** | 0.44 / 7.67 / 0.78 | 0.35 / 1.14 / 0.84 | 0.30 / 2.30 / 0.52 | 0.45 / 2.96 / 0.63 |
| SAGPool&gUnpool [25] | 0.77 / **1.15** / 0.63 | 0.40 / 2.00 / 0.32 | 0.47 / 7.54 / 0.77 | 0.40 / 2.13 / 0.82 | 0.40 / 3.80 / 0.49 | 0.49 / 3.32 / 0.60 |
| Group Pool&Unpool | 0.63 / 1.35 / 0.65 | 0.30 / 0.66 / **0.35** | **0.42** / **5.52** / **0.80** | **0.31** / **0.62** / **0.86** | **0.29** / **1.44** / **0.56** | **0.39** / **1.92** / **0.64** |
| +Oracle group label | **0.62** / 1.27 / **0.67** | **0.28** / **0.61** / **0.35** | - / - / - | - / - / - | - / - / - | - / - / - |

**Table 3.** Ablation study of various pooling&unpooling operations on SGCN [54] (FDE/COL/TCC). In the case of our Pedestrian Group Pooling&Unpooling, we additionally provide experimental results using the ground-truth group labels (Oracle). **Bold**: Best, Underline: Second best.

strategies: scene-level and pedestrian-level sampling in Fig. 5. Even though the probability maps of pedestrians are well predicted with the estimated group information (Fig. 5(a)), its limitation still remains. For example, all sampled trajectories in the probability distributions lean toward the same directions (Fig. 5(b)) or are scattered with different patterns even within group members, which leads to collisions between colleagues (Fig. 5(c)). Our GP-Graph with the proposed group-level sampling strategy predicts the collaborative walking trajectories of associated group members, which is independent of other groups (Fig. 5(d)).

### 4.4 Ablation Study

**Pooling&Unpooling.** To check the effectiveness of the proposed group pooling&unpooling layers, we compare it with different pooling methods including gPool [13] and SAGPool [25] with respect to FDE, COL and TCC. gPool proposes a top-$k$ pooling by employing a projection vector to compute a rank score for each node. SAGpool is similar to the gPool method, but encodes topology information in a self-attention manner. As shown in Table 3, for both gPool and SAGPool, pedestrian features are lost via the pooling operations on unimportant nodes. By contrast, our pooling approach focuses on group representations of the pedestrian graph structure because it is optimized to capture group-related patterns.

**Group hierarchy graph.** We examine each component of the group hierarchy graph in Table 4. Both intra-/inter-group interaction graphs show a noticeable performance improvement compared to the baseline models, and the inter-group graph with our group pooling operation has the most important role in performance improvement (variants 1 to 4). The best performances can be achieved when all three types of interaction graphs are used with a weight-shared baseline model, which takes full advantage of graph augmentations (variants 4 and 5).

| Varient ID | Components | | | | | | Performance | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | AW | MB | GP | WS | FG | GS | ETH | HOTEL | UNIV | ZARA1 | ZARA2 | AVG |
| 1 | - | ✓ | - | - | - | - | 0.45 / 0.74 | 0.26 / 0.48 | 0.39 / 0.66 | 0.28 / 0.48 | 0.23 / 0.41 | 0.32 / 0.55 |
| 2 | - | - | ✓ | - | - | - | 0.47 / 0.80 | **0.17** / 0.31 | 0.26 / 0.48 | 0.18 / 0.34 | 0.16 / **0.29** | 0.25 / 0.44 |
| 3 | - | ✓ | ✓ | ✓ | - | - | **0.43** / 0.69 | 0.20 / 0.37 | 0.25 / 0.47 | 0.19 / 0.35 | 0.17 / 0.32 | 0.25 / 0.44 |
| 4 | ✓ | ✓ | ✓ | - | - | - | 0.44 / 0.75 | 0.18 / **0.30** | **0.23** / 0.43 | 0.18 / 0.33 | 0.16 / **0.29** | 0.24 / 0.42 |
| 5 | ✓ | ✓ | ✓ | ✓ | - | - | **0.43** / **0.63** | 0.18 / **0.30** | 0.24 / **0.42** | **0.17** / **0.31** | **0.15** / **0.29** | **0.23** / **0.39** |
| 6 | ✓ | ✓ | ✓ | ✓ | ✓ | - | 0.55 / 0.87 | 0.24 / 0.31 | 0.42 / 0.82 | 0.30 / 0.56 | 0.22 / 0.35 | 0.35 / 0.58 |
| 7 | ✓ | ✓ | ✓ | ✓ | - | ✓ | **0.43** / **0.63** | 0.18 / **0.30** | 0.24 / **0.42** | **0.17** / **0.31** | **0.15** / **0.29** | **0.23** / **0.39** |

**Table 4.** Ablation study (ADE/FDE). AW, MB, GP, WS, FG and GS respectively denote agent-wise pedestrian graph, intra-group member graph, inter-group group graph, weight sharing among different interaction graph, fixed ratio node reduction of grouping and group-level latent vector sampling respectively. All tests are performed on SGCN. **Bold**: Best, Underline: Second best.

**Grouping method.** We introduce a learnable threshold parameter $\pi$ on the group assignment module in Eq. (2) because in practice the total number of groups in a scene can change according to the trajectory feature of the input pedestrian node. To highlight the importance of $\pi$, we test a fixed ratio group pooling with a node reduction ratio of 50%. As expected, the learnable threshold shows lower errors than the fixed ratio of group pooling (variants 5 and 6). This means that it is effective to guarantee the variability of group numbers, since the number can vary even when the same number of pedestrians exists in a scene.

Additionally, we report results for the group-level latent vector sampling strategy (variants 5 and 7). Since the ADE and FDE metrics are based on best-of-many strategies, there is no difference with respect to numerical performance. However, it allows each group to keep their own behavior patterns, and to represent independency between groups, as in Fig. 5.

## 5 Conclusion

In this paper, we present a GP-Graph architecture for learning group-aware motion representations. We model group behaviors in crowded scenes by proposing a group hierarchy graph using novel pedestrian group pooling&unpooling operations. We use them for our group assignment module and straight-forward group estimation trick. Based on the GP-Graph, we introduce a multi-modal trajectory prediction framework that can attend intra-/inter group interaction features to capture human-human interactions as well as group-group interactions. Experiments demonstrate that our method significantly improves performance on challenging pedestrian trajectory prediction datasets.

**Acknowledgement** This work is in part supported by the Institute of Information & communications Technology Planning & Evaluation (IITP) (No.2019-0-01842, Artificial Intelligence Graduate School Program (GIST), No.2021-0-02068, Artificial Intelligence Innovation Hub), the National Research Foundation of Korea (NRF) (No.2020R1C1C1012635) grant funded by the Korea government (MSIT), Vehicles AI Convergence Research & Development Program through the National IT Industry Promotion Agency of Korea (NIPA) funded by the Ministry of Science and ICT (No.S1602-20-1001), the GIST-MIT Collaboration grant and AI-based GIST Research Scientist Project funded by the GIST in 2022.

## References


1. Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., Savarese, S.: Social lstm: Human trajectory prediction in crowded spaces. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016) 2, 3, 4, 10
2. Bae, I., Jeon, H.G.: Disentangled multi-relational graph convolutional network for pedestrian trajectory prediction. Proceedings of the AAAI Conference on Artificial Intelligence (AAAI) (2021) 2, 3, 4
3. Bae, I., Park, J.H., Jeon, H.G.: Non-probability sampling network for stochastic human trajectory prediction. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2022) 3
4. Bartoli, F., Lisanti, G., Ballan, L., Del Bimbo, A.: Context-aware trajectory prediction. Pattern Recognition (PR) (2018) 4
5. Bengio, Y., Léonard, N., Courville, A.: Estimating or propagating gradients through stochastic neurons for conditional computation. arXiv preprint arXiv:1308.3432 (2013) 2, 7
6. Bisagno, N., Zhang, B., Conci, N.: Group lstm: Group trajectory prediction in crowded scenarios. In: Proceedings of European Conference on Computer Vision Workshop (ECCVW) (2018) 2, 4, 11, 12
7. Cangea, C., Velickovic, P., Jovanovic, N., Kipf, T., Lio', P.: Towards sparse hierarchical graph classifiers. arXiv preprint arXiv:1811.01287 (2018) 4, 6, 7, 10
8. Chen, G., Li, J., Lu, J., Zhou, J.: Human trajectory prediction via counterfactual analysis. In: Proceedings of International Conference on Computer Vision (ICCV) (2021) 3
9. Chen, G., Li, J., Zhou, N., Ren, L., Lu, J.: Personalized trajectory prediction via distribution discrimination. In: Proceedings of International Conference on Computer Vision (ICCV) (2021) 3, 4
10. Defferrard, M., Bresson, X., Vandergheynst, P.: Convolutional neural networks on graphs with fast localized spectral filtering. In: Proceedings of the Neural Information Processing Systems (NeurIPS) (2016) 4
11. Dendorfer, P., Elflein, S., Leal-Taixé, L.: Mg-gan: A multi-generator model preventing out-of-distribution samples in pedestrian trajectory prediction. In: Proceedings of International Conference on Computer Vision (ICCV) (2021) 3, 4
12. Fernando, T., Denman, S., Sridharan, S., Fookes, C.: Gd-gan: Generative adversarial networks for trajectory prediction and group detection in crowds. In: Proceedings of Asian Conference on Computer Vision (ACCV) (2018) 4, 11, 12
13. Gao, H., Ji, S.: Graph u-nets. In: Proceedings of the International Conference on Machine Learning (ICML) (2019) 4, 6, 7, 13
14. Ge, W., Collins, R.T., Ruback, R.B.: Vision-based analysis of small groups in pedestrian crowds. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) (2012) 4, 12
15. Gilmer, J., Schoenholz, S.S., Riley, P.F., Vinyals, O., Dahl, G.E.: Neural message passing for quantum chemistry. In: Proceedings of the International Conference on Machine Learning (ICML) (2017) 4
16. Gu, T., Chen, G., Li, J., Lin, C., Rao, Y., Zhou, J., Lu, J.: Stochastic trajectory prediction via motion indeterminacy diffusion. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2022) 3
17. Gupta, A., Johnson, J., Fei-Fei, L., Savarese, S., Alahi, A.: Social gan: Socially acceptable trajectories with generative adversarial networks. In: Proceedings of

IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2018) 2, 3, 4, 9, 10
18. Helbing, D., Molnar, P.: Social force model for pedestrian dynamics. Physical review E **51**(5), 4282 (1995) 3
19. Huang, Y., Bi, H., Li, Z., Mao, T., Wang, Z.: Stgat: Modeling spatial-temporal interactions for human trajectory prediction. In: Proceedings of International Conference on Computer Vision (ICCV) (2019) 3, 4, 10
20. Ivanovic, B., Pavone, M.: The trajectron: Probabilistic multi-agent trajectory modeling with dynamic spatiotemporal graphs. In: Proceedings of International Conference on Computer Vision (ICCV) (2019) 4
21. Jang, E., Gu, S., Poole, B.: Categorical reparameterization with gumbel-softmax. International Conference on Learning Representations (ICLR) (2017) 2, 7
22. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. In: International Conference on Learning Representations (ICLR) (2017) 3
23. Kosaraju, V., Sadeghian, A., Martín-Martín, R., Reid, I., Rezatofighi, H., Savarese, S.: Social-bigat: Multimodal trajectory forecasting using bicycle-gan and graph attention networks. In: Proceedings of the Neural Information Processing Systems (NeurIPS) (2019) 3, 4
24. Lawal, I.A., Poiesi, F., Anguita, D., Cavallaro, A.: Support vector motion clustering. IEEE Transactions on Circuits and Systems for Video Technology (TCSVT) (2017) 4
25. Lee, J., Lee, I., Kang, J.: Self-attention graph pooling. In: Proceedings of the International Conference on Machine Learning (ICML) (2019) 13
26. Lee, M., Sohn, S.S., Moon, S., Yoon, S., Kapadia, M., Pavlovic, V.: Muse-vae: Multi-scale vae for environment-aware long term trajectory prediction. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2022) 3, 4
27. Lee, N., Choi, W., Vernaza, P., Choy, C.B., Torr, P.H.S., Chandraker, M.: Desire: Distant future prediction in dynamic scenes with interacting agents. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2017) 4
28. Lerner, A., Chrysanthou, Y., Lischinski, D.: Crowds by example. Computer Graphics Forum **26**(3), 655–664 (2007) 10, 11, 12
29. Li, J., Ma, H., Tomizuka, M.: Conditional generative neural system for probabilistic trajectory prediction. Proceedings of IEEE International Conference on Intelligent Robots and Systems (IROS) (2019) 4
30. Li, J., Yang, F., Tomizuka, M., Choi, C.: Evolvegraph: Multi-agent trajectory prediction with dynamic relational reasoning. In: Proceedings of the Neural Information Processing Systems (NeurIPS) (2020) 4
31. Li, S., Zhou, Y., Yi, J., Gall, J.: Spatial-temporal consistency network for low-latency trajectory forecasting. In: Proceedings of International Conference on Computer Vision (ICCV) (2021) 4
32. Liang, J., Jiang, L., Murphy, K., Yu, T., Hauptmann, A.: The garden of forking paths: Towards multi-future trajectory prediction. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2020) 3
33. Liang, J., Jiang, L., Niebles, J.C., Hauptmann, A.G., Fei-Fei, L.: Peeking into the future: Predicting future person activities and locations in videos. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019) 3

34. Liu, Y., Yan, Q., Alahi, A.: Social nce: Contrastive learning of socially-aware motion representations. In: Proceedings of International Conference on Computer Vision (ICCV) (2021) 10
35. Maddison, C.J., Mnih, A., Teh, Y.W.: The concrete distribution: A continuous relaxation of discrete random variables. International Conference on Learning Representations (ICLR) (2017) 2, 7
36. Mangalam, K., Girase, H., Agarwal, S., Lee, K.H., Adeli, E., Malik, J., Gaidon, A.: It is not the journey but the destination: Endpoint conditioned trajectory prediction. In: Proceedings of European Conference on Computer Vision (ECCV) (2020) 4, 9, 10, 11
37. Marchetti, F., Becattini, F., Seidenari, L., Bimbo, A.D.: Mantra: Memory augmented networks for multiple trajectory prediction. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2020) 3
38. Mehran, R., Oyama, A., Shah, M.: Abnormal crowd behavior detection using social force model. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2009) 3
39. Mohamed, A., Qian, K., Elhoseiny, M., Claudel, C.: Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2020) 2, 3, 4, 9, 10, 11
40. Moussaïd, M., Perozo, N., Garnier, S., Helbing, D., Theraulaz, G.: The walking behaviour of pedestrian social groups and its impact on crowd dynamics. Public Library of Science One (2010) 1, 2, 4
41. Pellegrini, S., Ess, A., Gool, L.V.: Improving data association by joint modeling of pedestrian trajectories and groupings. In: Proceedings of European Conference on Computer Vision (ECCV) (2010) 2, 4
42. Pellegrini, S., Ess, A., Schindler, K., Van Gool, L.: You'll never walk alone: Modeling social behavior for multi-target tracking. In: Proceedings of International Conference on Computer Vision (ICCV) (2009) 3, 10, 11, 12
43. Pfeiffer, M., Paolo, G., Sommer, H., Nieto, J.I., Siegwart, R.Y., Cadena, C.: A data-driven model for interaction-aware pedestrian motion prediction in object cluttered environments. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA) (2018) 4
44. Qiu, F., Hu, X.: Modeling group structures in pedestrian crowd simulation. Simulation Modelling Practice and Theory (2010) 4
45. Rhee, S., Seo, S., Kim, S.: Hybrid approach of relation network and localized graph convolutional filtering for breast cancer subtype classification. In: Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligencev (IJCAI) (2018) 4
46. Robicquet, A., Sadeghian, A., Alahi, A., Savarese, S.: Learning social etiquette: Human trajectory understanding in crowded scenes. In: Proceedings of European Conference on Computer Vision (ECCV) (2010) 4
47. Robicquet, A., Sadeghian, A., Alahi, A., Savarese, S.: Learning social etiquette: Human trajectory understanding in crowded scenes. In: Proceedings of European Conference on Computer Vision (ECCV) (2016) 2, 10, 11
48. Rudenko, A., Palmieri, L., Lilienthal, A.J., Arras, K.O.: Human motion prediction under social grouping constraints. In: Proceedings of IEEE International Conference on Intelligent Robots and Systems (IROS) (2018) 1, 2
49. Sadeghian, A., Kosaraju, V., Sadeghian, A., Hirose, N., Rezatofighi, H., Savarese, S.: Sophie: An attentive gan for predicting paths compliant to social and physical

constraints. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019) 3, 4
50. Salzmann, T., Ivanovic, B., Chakravarty, P., Pavone, M.: Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data. In: Proceedings of European Conference on Computer Vision (ECCV) (2020) 3, 4, 9, 10
51. Seitz, M., Köster, G., Pfaffinger, A.: Pedestrian group behavior in a cellular automaton. In: Pedestrian and Evacuation Dynamics (2012) 4
52. Shafiee, N., Padir, T., Elhamifar, E.: Introvert: Human trajectory prediction via conditional 3d attention. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2021) 3
53. Shao, J., Loy, C.C., Wang, X.: Scene-independent group profiling in crowd. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2014) 12
54. Shi, L., Wang, L., Long, C., Zhou, S., Zhou, M., Niu, Z., Hua, G.: Sgcn: Sparse graph convolution network for pedestrian trajectory prediction. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2021) 2, 3, 4, 9, 10, 11, 13
55. Shi, X., Shao, X., Fan, Z., Jiang, R., Zhang, H., Guo, Z., Wu, G., Yuan, W., Shibasaki, R.: Multimodal interaction-aware trajectory prediction in crowded space. In: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI) (2020) 4
56. Singh, H., Arter, R., Dodd, L., Langston, P., Lester, E., Drury, J.: Modelling subgroup behaviour in crowd dynamics dem simulation. Applied Mathematical Modelling (2009) 4
57. Solera, F., Calderara, S., Cucchiara, R.: Socially constrained structural learning for groups detection in crowd. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) (2016) 12
58. Sun, H., Zhao, Z., He, Z.: Reciprocal learning networks for human trajectory prediction. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2020) 3, 4
59. Sun, J., Jiang, Q., Lu, C.: Recursive social behavior graph for trajectory prediction. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2020) 3
60. Sun, J., Li, Y., Fang, H.S., Lu, C.: Three steps to multimodal trajectory prediction: Modality clustering, classification and synthesis. In: Proceedings of International Conference on Computer Vision (ICCV) (2021) 4
61. Tao, C., Jiang, Q., Duan, L.: Dynamic and static context-aware lstm for multi-agent motion prediction. In: Proceedings of European Conference on Computer Vision (ECCV) (2020) 3, 10
62. Varshneya, D., Srinivasaraghavan, G.: Human trajectory prediction using spatially aware deep attention models. arXiv preprint arXiv:1705.09436 (2017) 4
63. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P., Bengio, Y.: Graph attention networks. In: International Conference on Learning Representations (ICLR) (2018) 3
64. Vemula, A., Muelling, K., Oh, J.: Social attention: Modeling attention in human crowds. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA) (2018) 3
65. Xu, Y., Wang, L., Wang, Y., Fu, Y.: Adaptive trajectory prediction via transferable gnn. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2022) 4

66. Yamaguchi, K., Berg, A.C., Ortiz, L.E., Berg, T.L.: Who are you with and where are you going? In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2011) 2, 3, 4, 12
67. Yi, S., Li, H., Wang, X.: Understanding pedestrian behaviors from stationary crowd groups. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2015) 10, 11
68. Ying, Z., You, J., Morris, C., Ren, X., Hamilton, W., Leskovec, J.: Hierarchical graph representation learning with differentiable pooling. In: Proceedings of the Neural Information Processing Systems (NeurIPS) (2018) 4
69. Yu, C., Ma, X., Ren, J., Zhao, H., Yi, S.: Spatio-temporal graph transformer networks for pedestrian trajectory prediction. In: Proceedings of European Conference on Computer Vision (ECCV) (2020) 3, 4, 9, 11
70. Yuan, Y., Weng, X., Ou, Y., Kitani, K.: Agentformer: Agent-aware transformers for socio-temporal multi-agent forecasting. In: Proceedings of International Conference on Computer Vision (ICCV) (2021) 3
71. Zanotto, M., Bazzani, L., Cristani, M., Murino, V.: Online bayesian nonparametrics for group detection. In: Proceedings of British Machine Vision Conference (BMVC) (2012) 12
72. Zhang, M., Cui, Z., Neumann, M., Chen, Y.: An end-to-end deep learning architecture for graph classification. In: Proceedings of the AAAI Conference on Artificial Intelligence (AAAI) (2018) 4
73. Zhang, P., Ouyang, W., Zhang, P., Xue, J., Zheng, N.: Sr-lstm: State refinement for lstm towards pedestrian trajectory prediction. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019) 4
74. Zhao, H., Wildes, R.P.: Where are you heading? dynamic trajectory prediction with expert goal examples. In: Proceedings of International Conference on Computer Vision (ICCV) (2021) 3
75. Zhao, T., Xu, Y., Monfort, M., Choi, W., Baker, C., Zhao, Y., Wang, Y., Wu, Y.N.: Multi-agent tensor fusion for contextual trajectory prediction. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2019) 3, 4
76. Zhong, J., Cai, W., Luo, L., Yin, H.: Learning behavior patterns from video: A data-driven framework for agent-based crowd modeling. In: Proceedings of the 2015 International Conference on Autonomous Agents and Multiagent Systems (AAMAS) (2015) 4
77. Zhou, B., Tang, X., Wang, X.: Coherent filtering: Detecting coherent motions from crowd clutters. In: Proceedings of European Conference on Computer Vision (ECCV) (2012) 4
78. Zhou, B., Wang, X., Tang, X.: Understanding collective crowd behaviors: Learning a mixture model of dynamic pedestrian-agents. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2012) 2