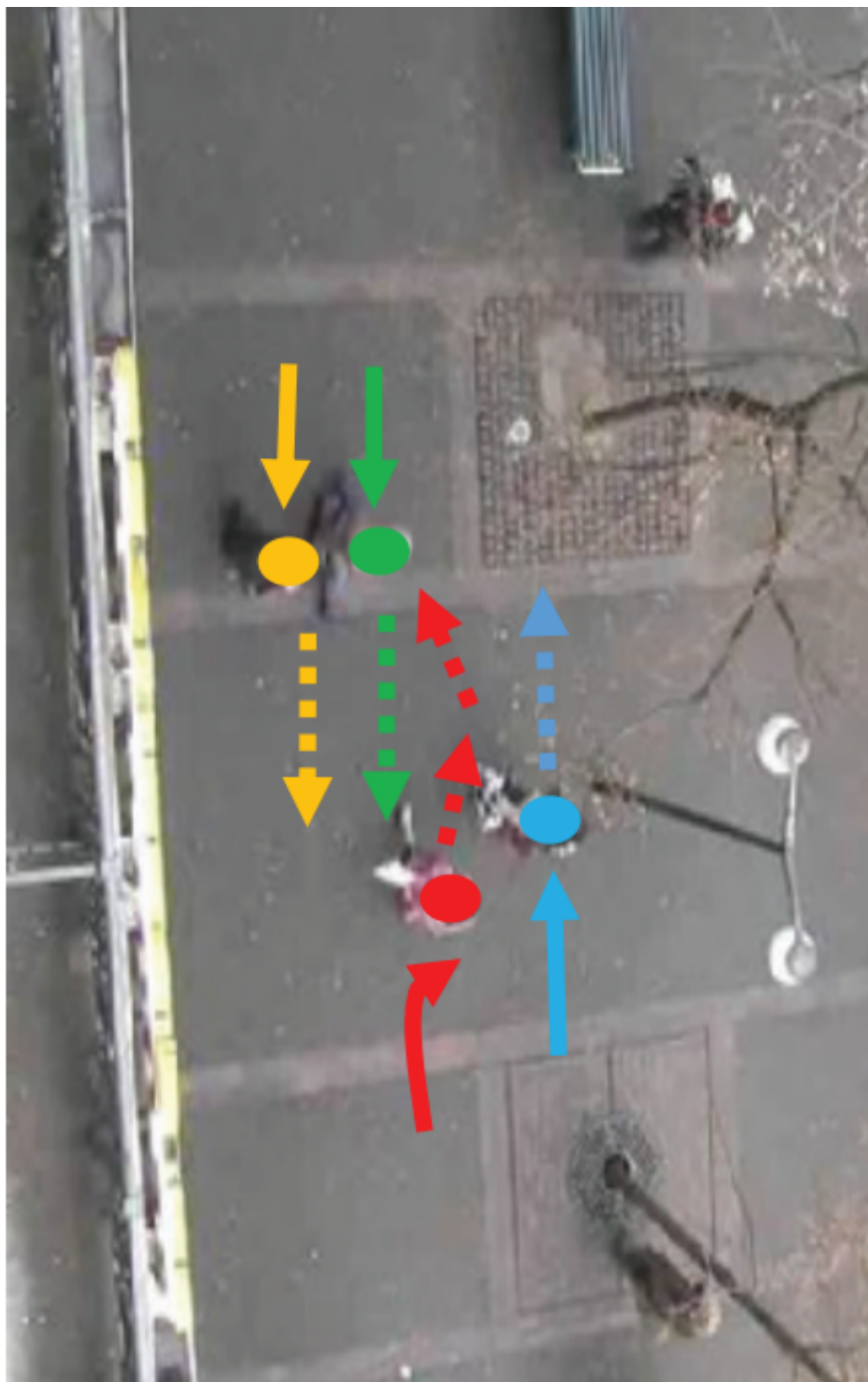


Sparse Graph Convolution Network for Pedestrian Trajectory Prediction

动机

从前基于密集无向交互活动 (dense undirected interaction) 的工作，因为对无关交互建模、缺乏对轨迹趋势的跟踪导致预测的结果和实际相差甚远，具体而言：



在这个场景下，只有红色行人需要改变自己的轨迹以避免撞到绿色的人，而其他行人的轨迹不会发生变动。

1. dense undirected graph:



很显然，有多余的边。因为蓝色和黄色并不会相互影响。

2. sparse undirected graph;



这张图语义有问题，因为是绿色影响了红色，绿色并没有变道。

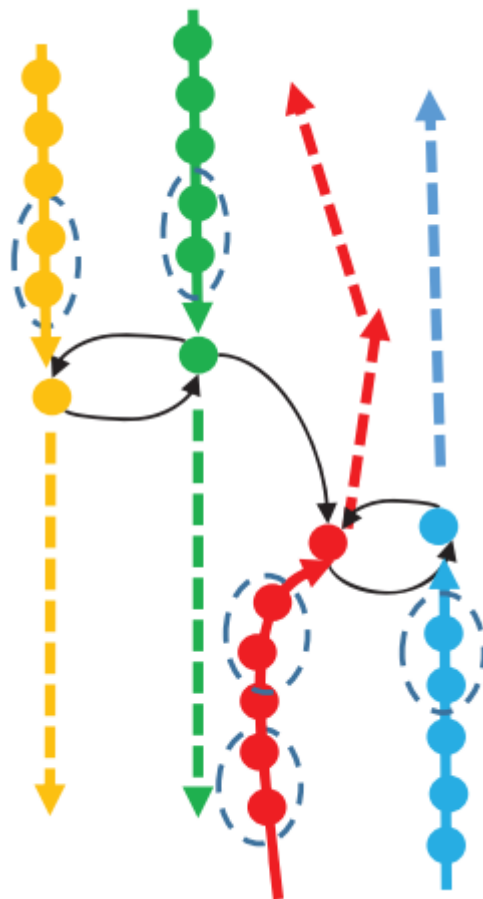
3. sparse directed graph:



这张图描述了行人相互影响的状态。

因此，选择sparse directed graph来描述行人间的相互影响

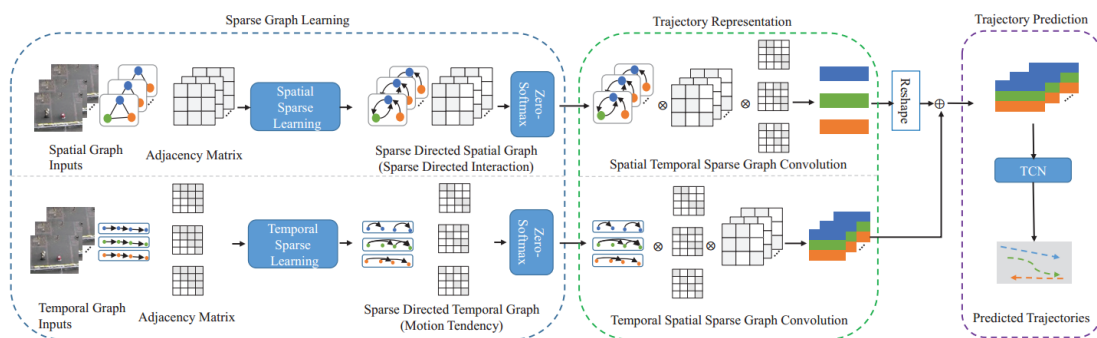
另外，由于传统的预测方式倾向于通过“绕路”来避免碰撞，从而会预测出与真实结果大相径庭的轨迹。所以引入了motion tendency的概念，用于更精确地预测轨迹。



网络结构

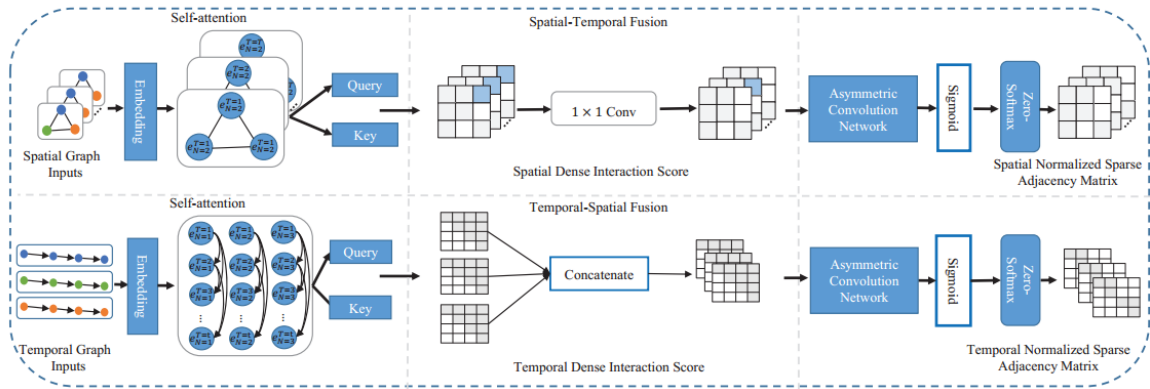
概览

网络有两个并列的部分——Spatial Graph和Temporal Graph，分别用来表示前面提到的sparse directed graph和motion tendency.



1. 输入用邻接矩阵表示的dense undirected graph
2. Sparse Learning模块会为图标出边的方向以及相应的权重
3. 经过Zero-softmax之后得到最终的两张图
4. Graph Convolution Networks和Time Convolution Networks处理后得到相应的轨迹预测结果。

Sparse Learning



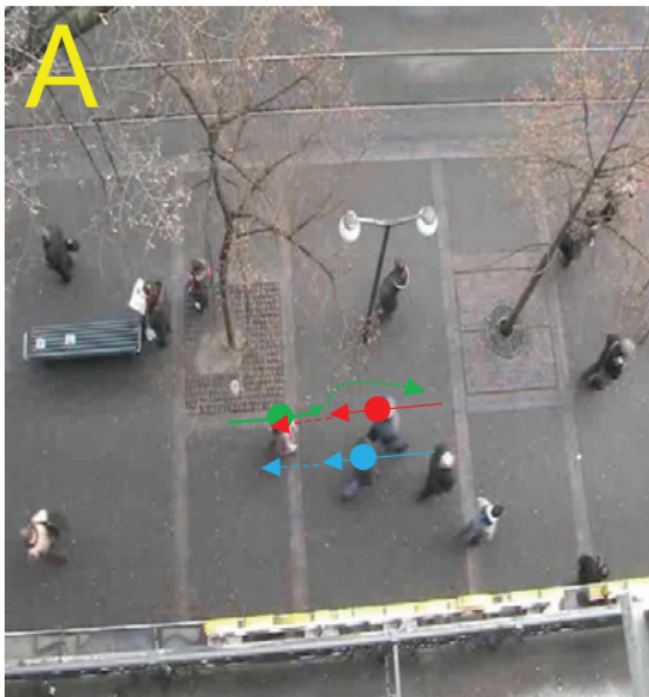
1. self-attention模块会为两张图上的点分别生成不同的权重。
2. Spatial-Temporal Fusion部分，分别使用 1×1 卷积和拼接操作为图的边标记权重。
3. 经过非对称卷积网络和激活层输出结果。

成果

Model	Year	ETH	HOTEL	UNIV	ZARA1	ZARA2	AVG
Vanilla LSTM [1]	2016	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 LSTM [1]	2016	1.09/2.35	0.79/1.76	0.67/1.40	0.47/1.00	0.56/1.17	0.72/1.54
SGAN [11]	2018	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 [37]	2019	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 [26]	2019	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 [23]	2019	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 [23]	2019	0.69/1.29	0.49/1.01	0.55/1.32	0.30/0.62	0.36/0.75	0.48/1.00
Social-STGCNN [32]	2020	0.64/1.11	0.49/0.85	0.44/0.79	0.34/ 0.53	0.30/0.48	0.44/0.75
RSBG w/o context [38]	2020	0.80/1.53	0.33/0.64	0.59/1.25	0.40/0.86	0.30/0.65	0.48/0.99
STAR [46]	2020	0.56/1.11	0.26/0.50	0.52/1.15	0.41/0.90	0.31/0.71	0.41/0.87
SGCN (Ours)	-	0.63/ 1.03	0.32/0.55	0.37/0.70	0.29/0.53	0.25/0.45	0.37/0.65

Table 1. Comparison with the baselines approach on the public benchmark dataset ETH and UCY for ADE/FDE. All approaches input 8 frames and output 12 frames. Our SGCN significantly outperform the comparison state-of-the-art works. The lower the better.

直观效果

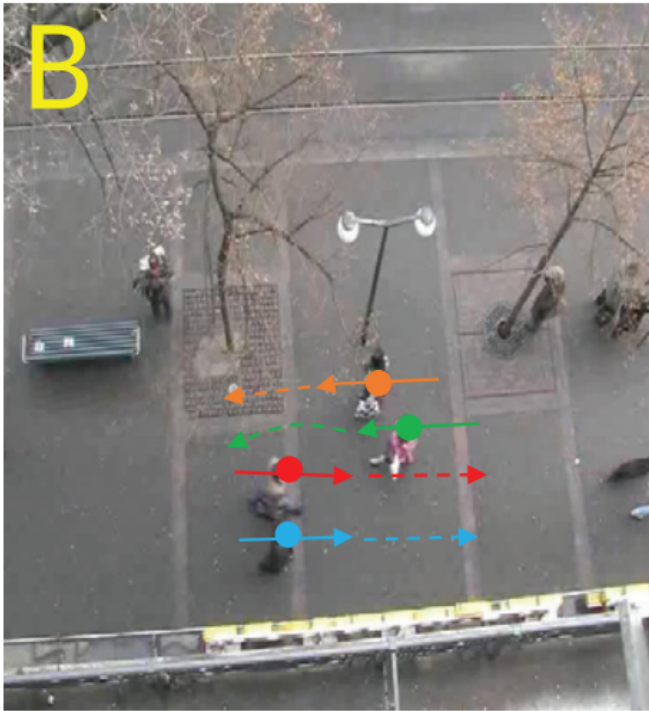


a1



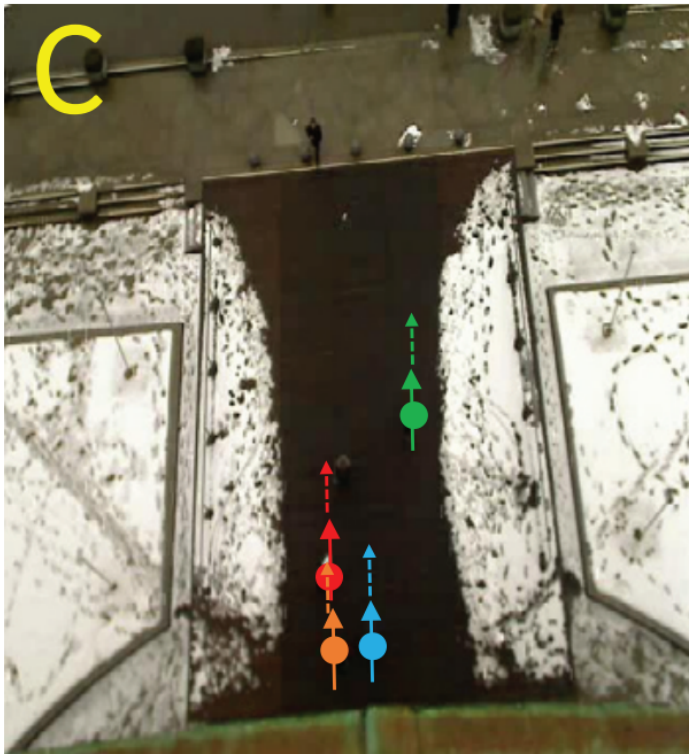
a2





b1

b2



c1

c2