Supplemental Materials for "SocialCircle: Learning the Angle-based Social Interaction Representation for Pedestrian Trajectory Prediction"

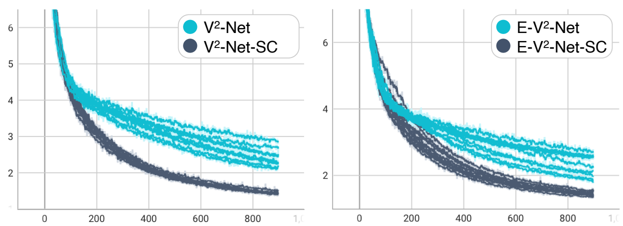

Figure 1. Loss curves (ℓ_2 loss at different training epochs) of different models at different training runs on NBA dataset. Curves are smoothed with the decay factor $= 0.8$.

A. Explanation of Attention Scores

We introduce the *Attention Scores* to quantitatively analyze how each SocialCircle partition relatively contributes to the final predicted trajectories. For the target agent i, it is defined as the squared sum of each $\mathbf{f}^i_{\theta_n} \in \mathbb{R}^{d_{\text{sc}}}$. Formally,

$$\text{AttentionScore}(i, n) = \sum_{j=1}^{d_{\text{sc}}} \left(\mathbf{f}^i_{\theta_n}[j]\right)^2. \tag{1}$$

Here, $\mathbf{f}[j]$ indicates the j-th element in the vector $\mathbf{f}$.

The attention score evaluates the contribution of different partitions to the subsequent prediction network at the **feature level**, meaning that a partition with more neighbors may not directly lead to a higher score. It is obtained through the combined effect of multiple layers together during the training process, including the embedding layers g_{embed}, the fuse layer $\{\mathbf{W}_{\text{fuse}}, \mathbf{b}_{\text{fuse}}\}$, as well as the backbone prediction model B_{pred}. Thus, we choose this item to analyze how the SocialCircle contributes to the whole prediction model only *qualitatively*.

B. Validation on NBA SportVU Dataset

Due to the page limitations, we only report SocialCircle models' performances on ETH-UCY and SDD. In this section, we further validate their performance in handling social interactions using the **NBA SportVU Dataset**.

Models	ADE (4.0s)	FDE (@2.0s)	FDE (@4.0s)
Social-LSTM [1]	1.79	1.53	3.16
S-GAN [2]	1.62	1.36	2.51
Social-STGCNN [6]	1.59	0.99	2.37
STAR [11]	1.26	1.28	2.04
PECNet [5]	1.83	1.69	3.41
NMMP [3]	1.33	1.11	2.05
GroupNet+NMMP [9]	1.25	1.08	1.80
GroupNet+CVAE [9]	**1.13**	**0.95**	1.69
MemoNet [10]	1.25	N/A	**1.47**
V^2-Net* [7]	1.28	0.96	1.68
V^2-Net-SC	1.22	**0.92**	**1.51**
E-V^2-Net* [8]	1.26	**0.93**	1.64
E-V^2-Net-SC	**1.18**	**0.90**	**1.46**

Table 1. Comparisons on NBA with *best-of-20* in meters. Lower ADE and FDE indicate better prediction performance. Models with "*" are reproduced under the same training settings.

B.1. Dataset Configurations

The **NBA SportVU Dataset** [4] (short for **NBA** dataset) is made up of a large number of real-world trajectories of ten players plus a ball captured by the SportVU tracking system during several NBA games. The complex interactions between different players will pose significant challenges for trajectory prediction. Positions of all players and balls are labeled in foot (1 foot = 0.3048 meter).

Following the settings of [9, 10], we predict future $t_f = 10$ frames of trajectories based on the past $t_h = 5$ frames of observations. The sample interval between two frames is still set to $\Delta t = 0.4$s. Frames where the basketball is not on the court will be ignored. We randomly sample about 50K prediction cases (*i.e.*, 50K trajectories) from multiple games to validate models. Among these cases, 65% (about 32,500 samples) will be used for training, 25% (about 12,500 samples) for testing, and the remaining 10% for validation.

B.2. Baselines

We choose Social-LSTM [1], S-GAN [2], Social-STGCNN [6], STAR [11], PECNet [5], NMMP [3], GroupNet+NMMP [9], GroupNet+CVAE [9], MemoNet [10], V^2-

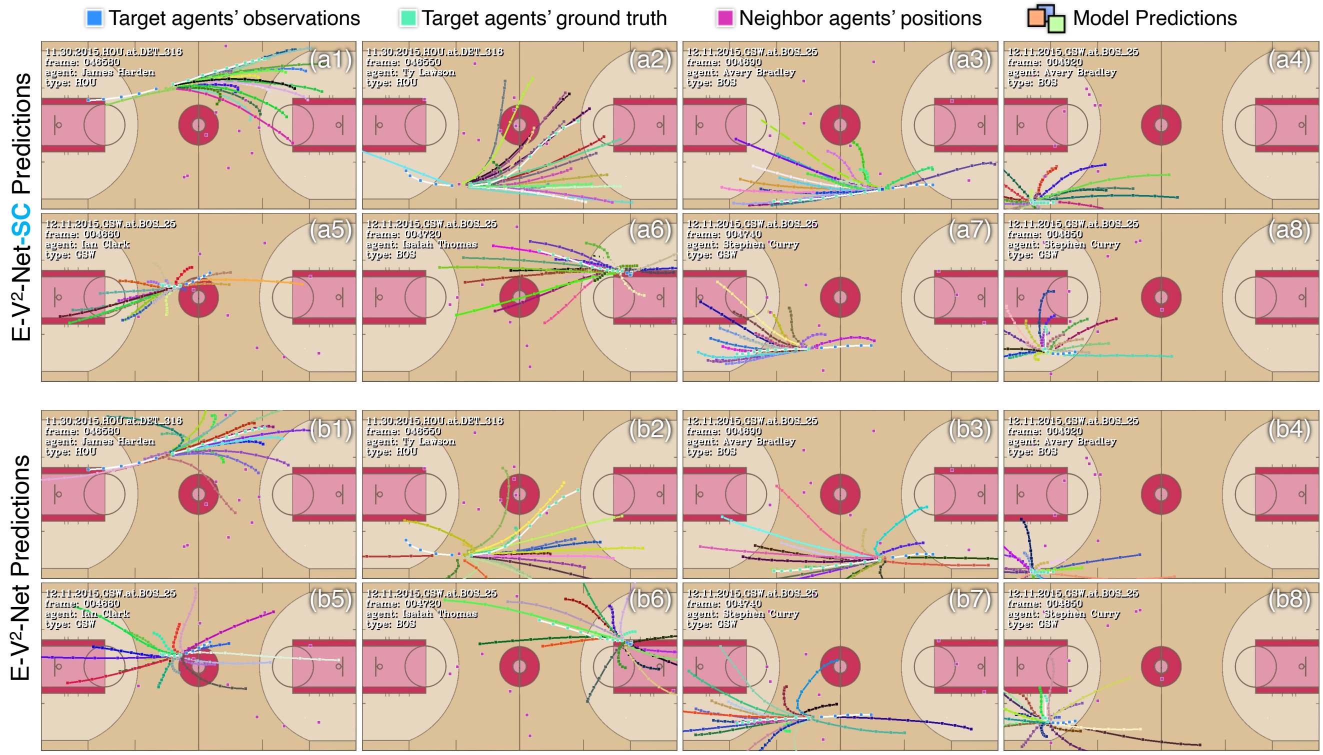

Figure 2. Visualized predicted trajectories provided by SocialCircle model E-V^2-Net-SC (subfigures (a1) to (a8)) and the original E-V^2-Net (subfigures (b1) to (b8)) on several NBA prediction scenes. Each sample includes 20 randomly generated trajectories.

Net* [7], and E-V^2-Net* [8] as our baselines on NBA.

B.3. Metrics

Except for ADE and FDE (minADE$_{20}$ and minFDE$_{20}$), following [9], we use the FDE-at-t-moment as a new metric to measure prediction performance. In detail, under the setting of $(t_h, t_f) = (5, 10)$ with sample interval $\Delta t = 0.4$s, the newly add metric FDE-at-5th-moment (minFDE@2.0s$_{20}$, short for FDE@2.0s) is defined as

$$\text{minFDE}_{20}(t) = \min_k \left\| \mathbf{p}_t^i - \hat{\mathbf{p}}_{k\,t}^{\,i} \right\|_2 , \tag{2}$$

$$\text{FDE@2.0s} = \text{minFDE}_{20}(t = t_h + 5). \tag{3}$$

The original FDE can be treated as FDE@4.0s, *i.e.*,

$$\text{FDE@4.0s} = \text{minFDE}_{20}(t = t_h + 10). \tag{4}$$

B.4. Quantitative Analyses

Comparisons to State-of-the-Art Methods. As shown in Tab. 1, the SocialCircle model E-V^2-Net-SC has achieved competitive results. Compared with the GroupNet+CVAE that obtains the best ADE, E-V^2-Net-SC's ADE is not as well as that model (about 4.42% worse ADE), but its FDEs (both at 2.0s and 4.0s) are better than those for about 5.26% and 13.60%. In addition, even though the FDE@4.0s of MemoNet and E-V^2-Net-SC are at the same level (less than 1% differences), E-V^2-Net-SC outperforms the other for about 5.60% ADE. Although the original E-V^2-Net performs not as well as these newly published methods, the proposed SocialCircle makes it available to achieve competitive results.

Ablation Studies. We validate SocialCircle on two backbone models, V^2-Net and E-V^2-Net, and report their corresponding SocialCircle models' performance in Tab. 1. With the help of the proposed SocialCircle, both these models have achieved considerable quantitative performance gains. In detail, compared with the basic V^2-Net, V^2-Net-SC has achieved the 4.68% better ADE and the 10.11% better FDE (@4.0s). The E-V^2-Net-SC also outperforms E-V^2-Net for about 6.34% ADE and 10.97% FDE (@4.0s). These results indicate the quantitative effectiveness of the proposed SocialCircle for handling prediction cases with complex social interactions in NBA dataset.

B.5. Qualitative Analyses

Analyses of the Training Process. We visualize the loss (ℓ_2 loss) curves of V^2-Net, E-V^2-Net, and their SocialCircle models at multiple training runs on NBA dataset in Fig. 1. All these models are trained under the same settings. It shows that the loss values drop faster and finally become lower by introducing SocialCircle to baseline models. In addition, their loss values become more stable across different training runs compared to the original model. We can infer that the proposed SocialCircle may also play a normalization factor, thus reducing the influence of randomized training factors (such as the shuffle operation at each training epoch and the randomly sampled noise vectors to generate multiple predictions).

Visualizations of Social Behaviors. We visualize trajectories forecasted by the SocialCircle model E-V^2-Net-SC and the original E-V^2-Net in several NBA scenes in Fig. 2. These models do not take into account agents' categories (*i.e.*, players with different teams or basketball) when forecasting trajectories. For prediction scenes with different distributions of neighbor players, E-V^2-Net-SC's predictions present better interactive trends.

Comparing Fig. 2 (a1 to a4) and (b1 to b4), several trajectories predicted by the non-SocialCircle model (b1 to b4) have gone out of the court, while there are rarely these cases in the predictions of SocialCircle model (a1 to a4). It shows that SocialCircle models could learn players' different behavior patterns according to the SocialCircle, even though they do not know where the borders of the court are, thus making their predictions in line with the scene context.

In addition, the game-related interaction is a class of interactions specific to the NBA dataset, such as players carrying the ball on offense, switching from offense to defense, and many other interactive behaviors. Comparing Fig. 2 (a5 to a8) and (b5 to b8), we can see that SocialCircle could also better describe these interactive behaviors. For example, agent "Isaiah Thomas" moves from a complete standstill to start moving from the free throw lane during the observation period in case (a6). According to other players' status, the SocialCircle model finally provides predictions that seem like running to the frontcourt to start the offense. Unlike predictions shown in Fig. 2 (a6), trajectories predicted by the non-SocialCircle model appear very confusing, including both aggressive and defensive. Other game-interactive cases, like scoring in various ways in case (a7) and the flexible movements in case (a8), present similar trends, which indicates SocialCircle's capability to handle various social-interactive behaviors in different prediction scenes.

C. Further Discussions on Limitations

As mentioned in the "Limitations" section, neighbor agents' movement directions have not been considered in the proposed SocialCircle. This section further discusses whether the movement direction factor should be considered as one of the SocialCircle meta components.

Variations	V D R mR	ADE/FDE	Drop (%)
E-V^2-Net*	× × × ×	6.73/10.75	2.9%/3.8%
E-V^2-Net-SC	✓ ✓ ✓ ×	6.54/10.36	(base)
E-V^2-Net-SC-4f	✓ ✓ ✓ ✓	6.84/10.94	4.6%/5.6%

Table 2. Ablation studies on validating the movement direction ("mR") factor on SDD. "V", "D", and "R" represent current velocity, distance, and direction factors. Values in "Drop" are the percentage matrices drop compared to the base model.

C.1. Limitation Analysis

As shown in Fig. 3, we conducted another toy experiment to show models' responses to the manual agent with different movement directions. In all 3-factor cases (a2) to (a5), the SocialCircle model forecasts almost the same trajectories (except for the noise factor for random generation). It is worth noting that the predictions in case (a3) are relatively "dangerous", for there might be potential collisions with the manual neighbor.

From the point of view of network training, we can simply understand that the whole prediction network forecasts an "average" trajectory to satisfy all these training samples with the same SocialCircle but move in different directions. As a result, it may predict trajectories with avoidances for the neighbors that may not collide with the target agent (like Fig. 3 (a5)), or may still collide with others (like Fig. 3 (a3)).

It should be noted that these extreme cases in the toy experiments are rarely seen in real-world prediction scenarios. In most ETH-UCY and SDD scenes, SocialCircle models still work as expected. Nevertheless, these few uncovered collision cases still indicate their limitations, although they have achieved better quantitative performance.

C.2. The Movement Direction Factor.

Following the "lite-rules" assumption, we attempt to add the movement direction factor to provide detailed interactive information. It is defined as the average of each neighbor's moving direction located in some partition. Formally,

$$\mathbf{f}^i_{\text{mdir}}\left(\theta_n\right) = \frac{1}{|\mathbf{N}^i(\theta_n)|} \sum_{j \in \mathbf{N}^i(\theta_n)} \text{atan2}\left(f_{2\text{D}}\left(\mathbf{p}^j_{t_h} - \mathbf{p}^j_1\right)\right). \tag{5}$$

The corresponding 4-factor SocialCircle meta vector is

$$\mathbf{f}^i_{\text{meta}}\left(\theta_n\right) = \left(\mathbf{f}^i_{\text{vel}}\left(\theta_n\right), \mathbf{f}^i_{\text{dis}}\left(\theta_n\right), \mathbf{f}^i_{\text{dir}}\left(\theta_n\right), \mathbf{f}^i_{\text{mdir}}\left(\theta_n\right)\right)^\top. \tag{6}$$

C.3. Ablation Studies and Visualized Analyses of the Movement Direction Factor

Quantitative Analyses. We run experiments to quantitatively validate the usefulness of this movement direction factor on SDD, and their results are reported in Tab. 2.

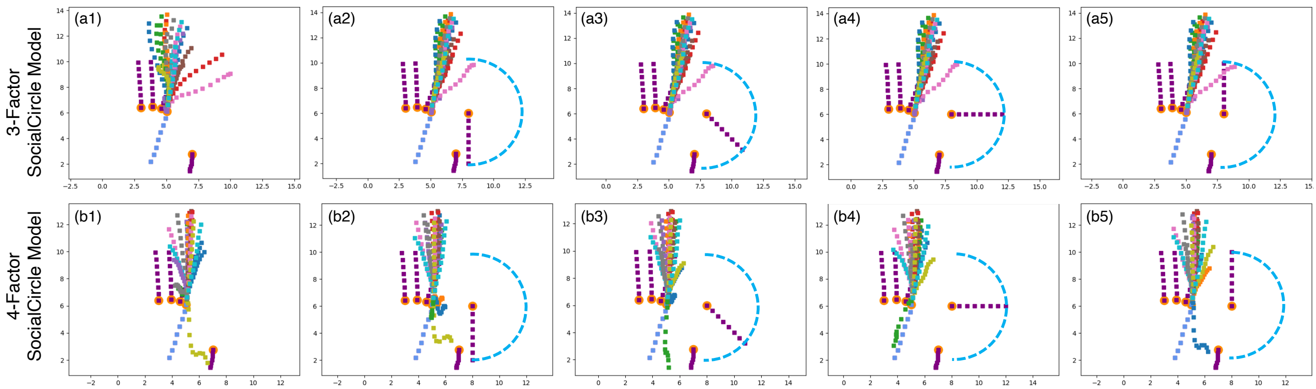

Figure 3. Visualized E-V^2-Net-SC predictions with manual neighbors with different movement directions. In this toy experiment, we set $d_m = 2.97$ and $v_m = 4.00$. (a1) to (a5) are predictions provided by the **3-factor** SocialCircle model, and (b1) to (b5) are predictions by **4-factor** model. Cases (a1) and (a5) are their original predictions without any given manual neighbors.

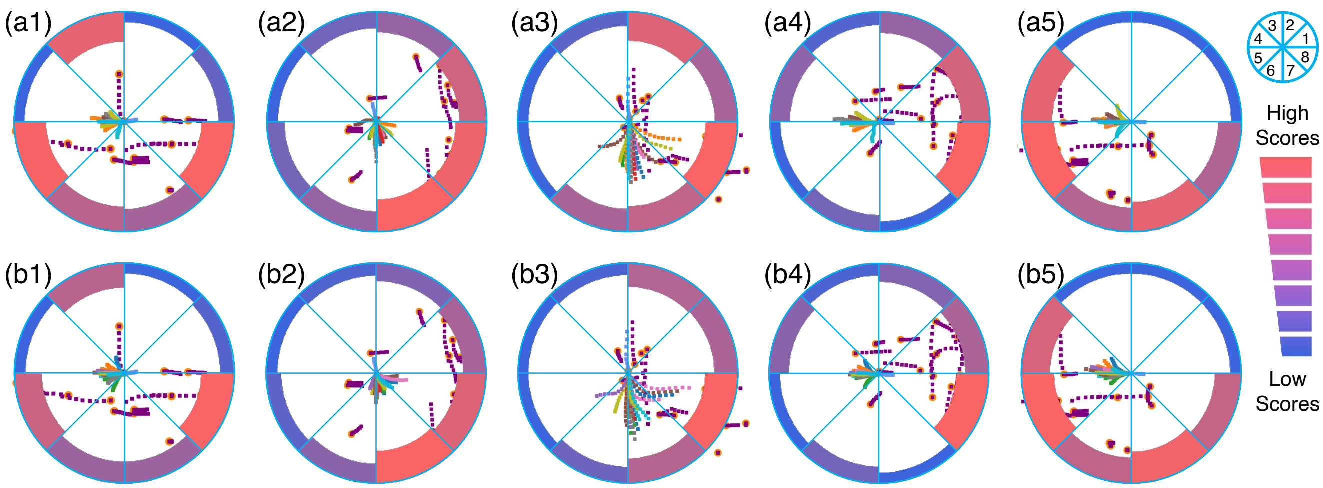

Figure 4. Visualized predicted trajectories and their corresponding attention scores in several real-world prediction cases (SDD-little0) provided by the **4-factor** E-V^2-Net-SC (a1) to (a5) and the **3-factor** E-V^2-Net-SC (b1) to (b5).

By adding this additional factor, the E-V^2-Net-SC-4f's performance drops significantly. Compared to the 3-factor E-V^2-Net-SC, it has 4.6% worse ADE and 5.6% worse FDE. Especially, its performance is even worse than the non-SocialCircle-model E-V^2-Net, which means that just adding such a simple new factor prevents other factors from expressing their contributions.

We infer that the movement direction factor brings more complex constraints to each prediction case, thus making the training process more difficult while reducing the model's generalization capability. In detail, the current three factors (velocity, distance, direction) are relatively "weak" rules to describe social interactions. Thus, the obtained SocialCircles could be similar even in different prediction cases. On the contrary, the movement direction factor varies from 0 to 2π for each neighbor in each partition, which brings extra "complexity" for each interactive case, thus further increasing the difficulty of model training in the case of the same network structure and training data.

Validation of Moving Directions. In Fig. 3 (b1) to (b5), we visualize the predicted trajectories provided by the 4-factor E-V^2-Net-SC corresponding to cases (a1) to (a5). We can easily see that predictions in cases (b2) to (b5) are different due to the various moving directions of the given manual neighbor. However, trajectories forecasted by the 4-factor model are far worse than those predicted by the 3-factor model. In detail, several randomly generated trajectories are distributed "messily" around the target agent, which could be caused by the "misleading" of 4-factor SocialCircle on predicted trajectories at different spatial positions. In other words, the newly added movement direction factor may prevent the backbone prediction model from exhibiting its original prediction performance.

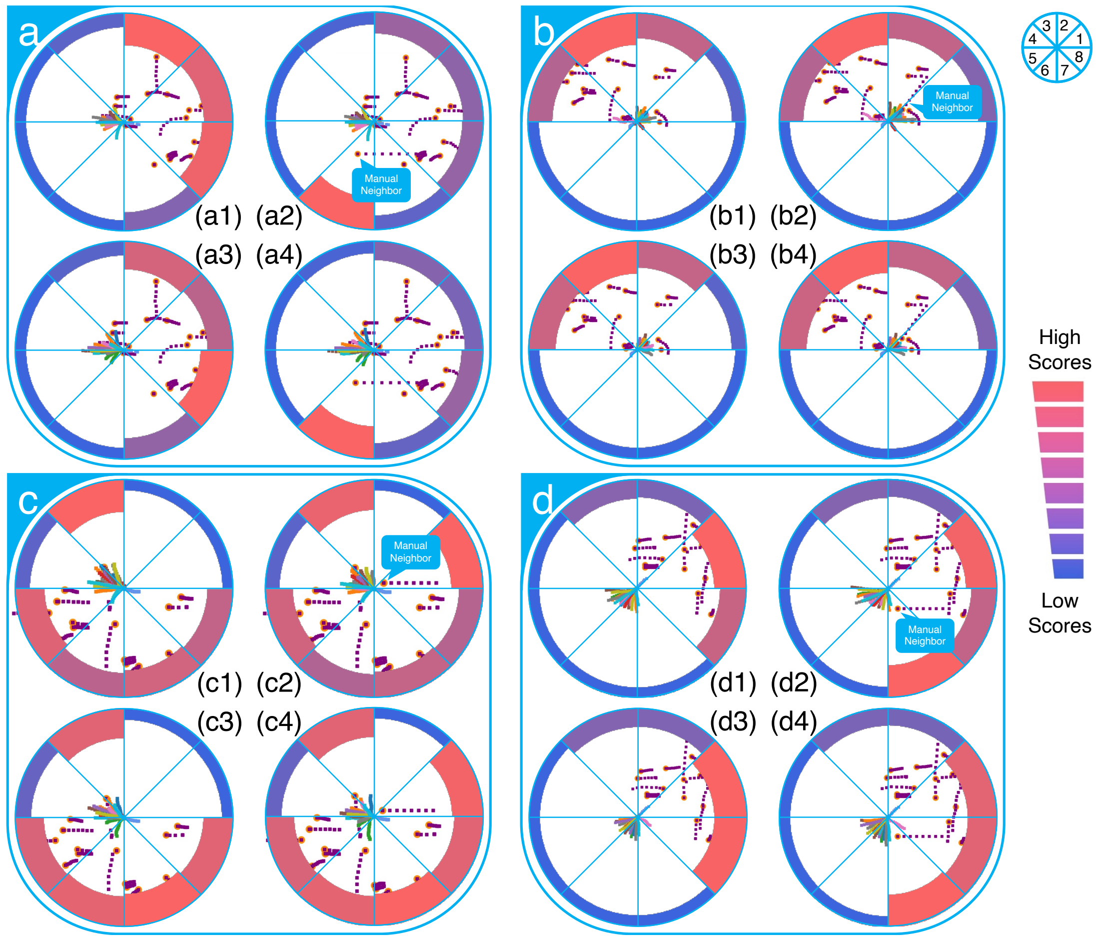

Figure 5. Visualized predicted trajectories and the corresponding attention scores of several real-world cases by adding additional manual neighbors. For each case $x \in \{$a, b, c, d$\}$, subfigure (x1) is the **4-factor** model's prediction, and (x3) is the **3-factor** model's prediction. subfigures (x2) and (x4) are obtained by adding manual neighbors to cases (x1) and (x3), respectively.

Moving Directions and Attention Scores. We visualize predictions of both 3-factor and 4-factor SocialCircle models on more real-world scenes in Fig. 4 and toy prediction cases with manual neighbors in Fig. 5. Comparing Fig. 4 (a1) and (b1), it shows that more SocialCircle partitions have been paid attention to (red colored partitions) in the 4-factor model in (a1) than (b1). Cases $\{$(a2), (b2)$\}$ and $\{$(a3), (b3)$\}$ also show similar trends. It means that more partitions or neighbors (*i.e.*, more "rules") are considered simultaneously to make final predictions for the 4-factor SocialCircle model. In addition, predictions provided by the 4-factor SocialCircle could hardly handle interactive behaviors in complex social interaction cases. For example, predictions in partitions 7 and 8 in Fig. 4 (b3) show strong avoidance trends to the coming neighbor. In contrast, predictions in the same partitions in (a3) have almost no responses to the potential collision. More visualized toy results with manual neighbors on real-world scenes are available in Fig. 4.

C.4. Summary of the Movement Direction Factor

The 3-factor SocialCircle (velocity, distance, direction) could not reflect neighbor agents' moving directions when modeling social interactions and forecasting trajectories. It takes an "average" way to handle neighbors with different movement directions, which means that its forecasted trajectories may not fit the interaction context well in some "extreme" interaction cases (like Fig. 3 (a3)).

We try to address this limitation by adding the new movement direction factor to the SocialCircle meta components. However, the newly added factor may lead to a performance drop. As we can see from the visualized predictions and attention scores, it is most likely due to adding too many constraints to the interaction cases, which reduces the model's ability to generalize across different complex prediction scenarios. Although the new factor could help to represent better interactive behaviors in some specific cases, degrading the original performance of the prediction model is something we do not expect. Therefore, the movement direction factor is deprecated in the SocialCircle. The currently proposed SocialCircle is a compromise that devotes itself to describing interactive behaviors through as few rules as possible while maximizing its usability in different trajectory prediction scenes. We will further investigate this limitation in our subsequent work.

D. Visualized Validation of the Number of SocialCircle Partitions

The proposed SocialCircle represent agents' social behaviors in an angle-based way by partitioning the cyclic representation. We have quantitatively analyzed how the number of partitions N_θ affects models' performance in "The Number of SocialCircle Partitions" in Sec. 4.3. Its results indicate that models with the larger N_θ (8 in our experiments) perform the best, and the smaller N_θ (like $N_\theta = 1$) could even lead to a performance drop compared to the non-SocialCircle model. Note that due to our settings of predicting trajectories based on 8 historical observed frames, the maximum number of partitions is set to 8 to prevent unnecessary zero-paddings in trajectories' representations from pulling down the performance of the original backbone trajectory prediction network. In this section, we further validate how the number of partitions contributes to the prediction model qualitatively by visualizing the SocialCircle.

Visualized SocialCircle with Different N_θ. Fig. 6 provides the visualized attention scores on different prediction cases in SDD-little0 with the $N_\theta = 4$ (subfigures (a1) to (a5)) and the $N_\theta = 8$ ((b1) to (b5)) E-V^2-Net-SC models. These two models are trained and validated under the same condition except for the N_θ.

Comparing Fig. 6 (a3) and (b3), the 8-partition model provides trajectories with different social behaviors for $\theta \in [1.5\pi, 2\pi)$, *i.e.*, partitions 7 and 8. In detail, predictions in partition-8 mostly try to avoid the right-coming neighbor, while predictions in partition-7 mostly walk as normal cases. For the 4-partition models' predictions in Fig. 6 (a3), predictions within the whole partition-4 all present the avoidance tendance, even though some predicted trajectories are far away from the existing neighbors. Similar cases also appear in cases (a2, partition-4) v.s. (b2, partitions 7 and 8) and cases (a5, partition-3) v.s. (b5, partitions 5 and 6). All these comparisons point out that a smaller number of SocialCircle partitions may lead to a coarser recognition and modeling of social behaviors, thus further causing misleading shifts in the predicted trajectories.

We also add manual neighbors to real-world prediction cases on SDD-little0 to validate both $N_\theta = 4$ and $N_\theta = 8$ E-V^2-Net-SC models' responses. As shown in Fig. 7, $N_\theta = 8$ model presents better spatial resolutions for handling social interactions. For example, compared to the $N_\theta = 4$ case (c2, partition-1), the corresponding $N_\theta = 8$ partition (c4, partition-2) has been less affected due to the manual neighbor. As a result, predictions in 8-partitions cases $\{$(c4, partition-3), (c4, partition-4)$\}$ show different interactive trends. These results indicate that 8-partition SocialCircle models have better angular resolution to model potential social interactions as well as quantify their roles in modifying forecast results.

References

[1] Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. Social lstm: Human trajectory prediction in crowded spaces. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 961–971, 2016. 1

[2] Agrim Gupta, Justin Johnson, Li Fei-Fei, Silvio Savarese, and Alexandre Alahi. Social gan: Socially acceptable trajectories with generative adversarial networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2255–2264, 2018. 1

[3] Yue Hu, Siheng Chen, Ya Zhang, and Xiao Gu. Collaborative motion prediction via neural motion message passing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6319–6328, 2020. 1

[4] Kostya Linou, Dzmitryi Linou, and Martijn de Boer. Nba player movements. https://github.com/linouk23/NBA-Player-Movements, 2016. 1

[5] Karttikeya Mangalam, Harshayu Girase, Shreyas Agarwal, Kuan-Hui Lee, Ehsan Adeli, Jitendra Malik, and Adrien Gaidon. It is not the journey but the destination: Endpoint conditioned trajectory prediction. In *European Conference on Computer Vision*, pages 759–776, 2020. 1

[6] Abduallah Mohamed, Kun Qian, Mohamed Elhoseiny, and Christian Claudel. Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14424–14432, 2020. 1

[7] Conghao Wong, Beihao Xia, Ziming Hong, Qinmu Peng, Wei Yuan, Qiong Cao, Yibo Yang, and Xinge You. View vertically: A hierarchical network for trajectory prediction via fourier spectrums. In *European Conference on Computer Vision*, pages 682–700. Springer, 2022. 1, 2

[8] Conghao Wong, Beihao Xia, Qinmu Peng, and Xinge You. Another vertical view: A hierarchical network for heterogeneous trajectory prediction via spectrums. *arXiv preprint arXiv:2304.05106*, 2023. 1, 2

[9] Chenxin Xu, Maosen Li, Zhenyang Ni, Ya Zhang, and Siheng Chen. Groupnet: Multiscale hypergraph neural networks for trajectory prediction with relational reasoning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6498–6507, June 2022. 1, 2

[10] Chenxin Xu, Weibo Mao, Wenjun Zhang, and Siheng Chen. Remember intentions: Retrospective-memory-based trajectory prediction. In *Proceedings of the IEEE/CVF Conference*

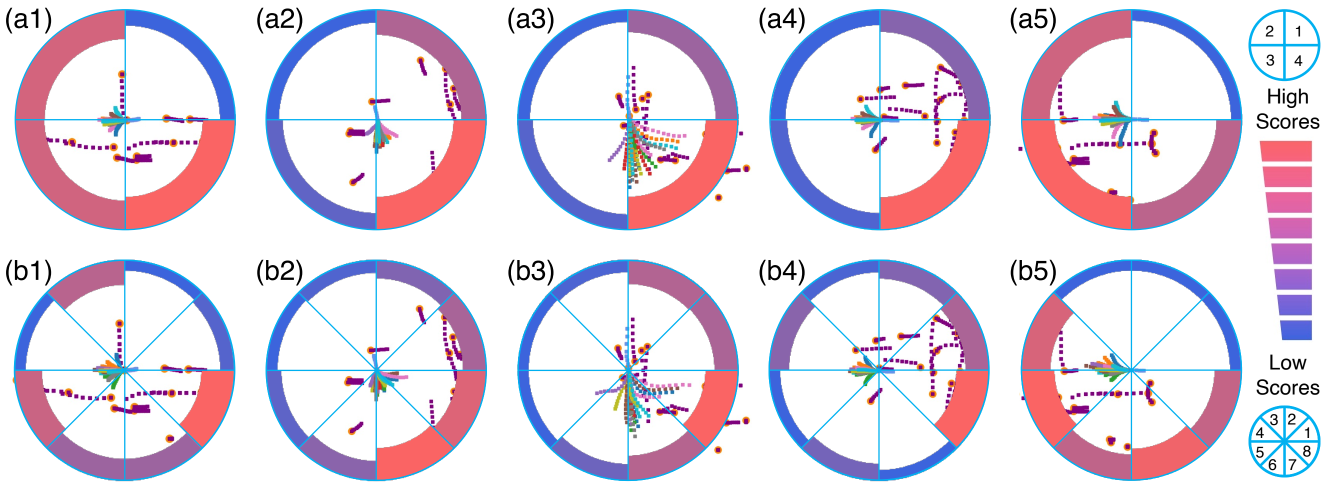

Figure 6. Visualized predicted trajectories and the corresponding attention scores of several real-world prediction cases on SDD-little0 provided by the **4-partition** E-V^2-Net-SC (a1) to (a5) and the **8-partition** E-V^2-Net-SC (b1) to (b5).

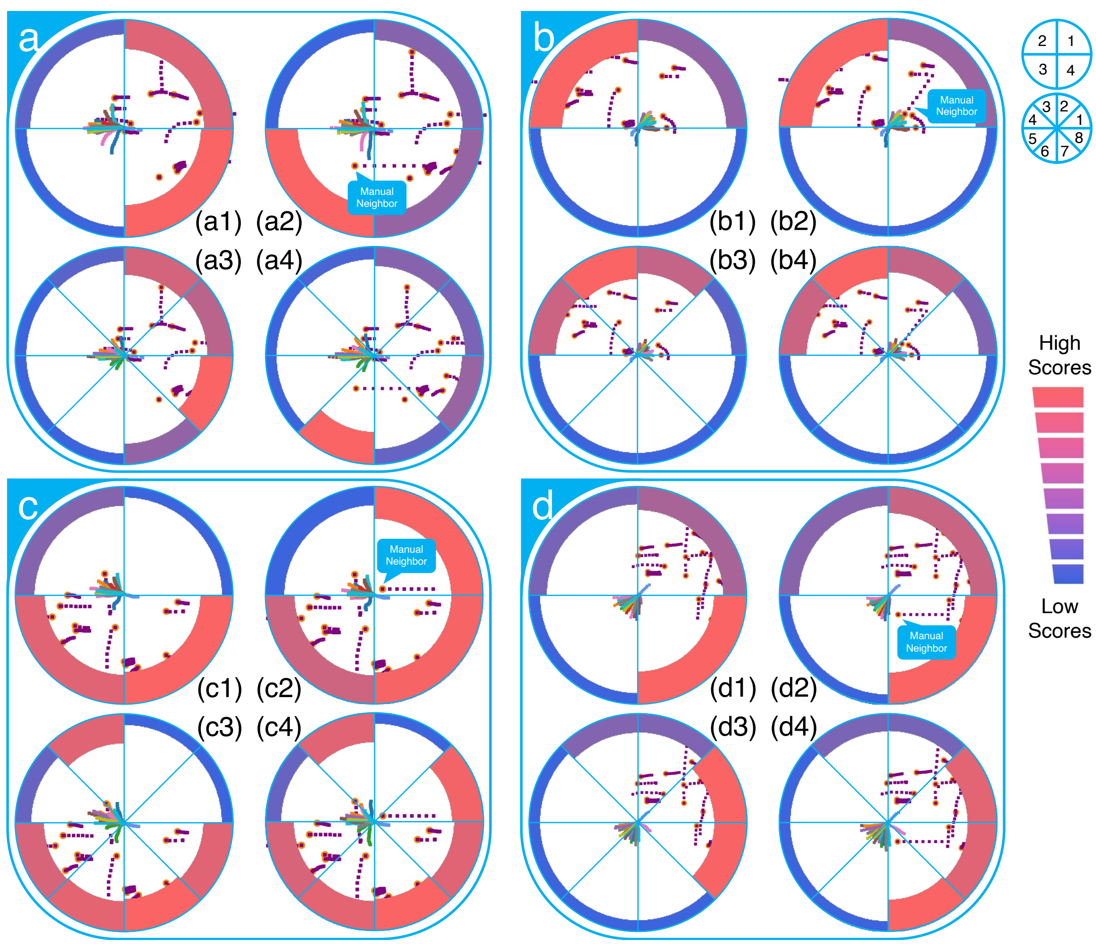

Figure 7. Visualized predicted trajectories and the corresponding attention scores of several real-world cases by adding additional manual neighbors. For each case $x \in \{\text{a, b, c, d}\}$, subfigure $(x1)$ is the **4-partition** ($N_\theta = 4$) model's prediction, and $(x3)$ is **8-partition** ($N_\theta = 8$) prediction. subfigures $(x2)$ and $(x4)$ are obtained by adding manual neighbors to cases $(x1)$ and $(x3)$, respectively.

on Computer Vision and Pattern Recognition (CVPR), pages 6488–6497, June 2022. 1

[11] Cunjun Yu, Xiao Ma, Jiawei Ren, Haiyu Zhao, and Shuai Yi. Spatio-temporal graph transformer networks for pedestrian trajectory prediction. In *European Conference on Computer Vision*, pages 507–523. Springer, 2020. 1