# SocialMOIF: Multi-Order Intention Fusion for Pedestrian Trajectory Prediction

## Supplementary Material

## 1. Definition of Metrics

The optimal prediction is utilized to calculate the average displacement error (ADE) and the final displacement error (FDE), which serve as the main metrics [1]. Additionally, the negative log likelihood (NLL) estimated from the test set is reported [16].

(1) Average Displacement Error (ADE): the average Euclidean distance between the ground truth coordinates and the predicted coordinates at all the predicted moments.

$$ADE = \frac{\sum_{i=1}^{(N_n+1)} \sum_{t=T_H+1}^{T_P} \left\| \hat{g}_i^t - \hat{g}_{igt}^t \right\|}{(N_n+1) \times T_P} \tag{1}$$

where $g_{igt}^t$ is the true position of $i$ at moment $t$ of the future prediction.

(2) Final Displacement Error (FDE): the Euclidean distance between the predicted point and the true point on the ground at the final prediction moment $T_P$.

$$FDE = \frac{\sum_{i=1}^{(N_n+1)} \left\| \hat{g}_i^{T_P} - \hat{g}_{igt}^{T_P} \right\|}{(N_n+1)} \tag{2}$$

(3) Negative Log-Likelihood (NLL): how well the predicted trajectory matches the true trajectory for the entire prediction period for each agent.

$$NLL = -\sum_{t=T_H+1}^{T_p} \log p\left(\hat{g}_i^t \mid U_i, S_i\right) \tag{3}$$

## 2. Implementation Details of Training

The Adam optimizer, which is an extension of stochastic gradient descent, integrates both momentum and adaptive learning rate strategies. This optimizer was employed during the training phase of the SocialMOIF to enhance both the speed and stability of training. The details of the hyperparameters were reported in Table 1.

Table 1. Performing hyperparameters for each scenario. Due to the different characteristics of each scenario, there were differences in the corresponding hyperparameters.

| | | Epoch | Learning Rate | Batch Size |
|---|---|---|---|---|
| **ETH/UCY** | Eth | 1000 | 1e-4 | 32 |
| | Hotel | 1000 | 7e-4 | 32 |
| | Univ | 1000 | 1e-3 | 32 |
| | Zara01 | 1000 | 3e-4 | 32 |
| | Zara02 | 1000 | 5e-4 | 32 |
| **SDD** | SDD | 1000 | 7e-4 | 32 |
| **NBA** | Rebound | 2000 | 1e-4 | 32 |
| | Scores | 2000 | 1e-4 | 32 |
| **Nuscenes** | Nuscenes | 3000 | 1e-4 | 32 |

## 3. Baselines in the Comparison Experiments

We conducted a evaluation of our SocialMOIF model against several state-of-the-art prediction approaches, which were succinctly outlined below.

- FLEAM [1]: It used an image semantic segmentation algorithm to obtain multi-category obstacle information and designed an end-to-end fully convolutional LSTM encoder-decoder with an attention mechanism to overcome the shortcomings of LSTM.
- ST-MR [3]: This paper aimed to forecast multiple paths based on a historical trajectory by modeling multi-scale graph-based spatial transformers combined with a trajectory smoothing algorithm named "Memory Replay" utilizing a memory graph.
- MTN [17]: It proposed an efficient multimodal transformer network that aggregated the trajectory and ego-vehicle speed variations at a coarse granularity and interacted with the optical flow at a fine-grained level to fill the vacancy of highly dynamic motion.
- MANTRA [6]: It incorporated scene knowledge in the decoding state by learning a CNN on top of semantic scene maps. Memory growth was limited by learning a writing controller based on the predictive capability of existing embeddings.
- Agent-former [18]: This paper proposed a new Transformer, termed AgentFormer, that simultaneously modeled the time and social dimensions. The model leveraged a sequence representation of multi-agent trajectories by flattening trajectory features across time and agents.
- Bitrap [16]: It presented BiTraP, a goal-conditioned bi-directional multi-modal trajectory prediction method based on the CVAE. BiTraP estimated the goal of trajectories and introduced a novel bidirectional decoder to improve longer-term trajectory prediction accuracy.
- MemoNet [14]: This paper imitated the mechanism of retrospective memory in neuropsychology and proposed MemoNet, an instance-based approach that predicted the movement intentions of agents by looking for similar scenarios in the training data.
- EqMotion [15]: To achieve motion equivariance, this paper proposed an equivariant geometric feature learning module to learn a Euclidean transformable feature through dedicated designs of equivariant operations.
- V2-Net [10]: It propose a hierarchical network V2 -Net,

Table 2. Comparison of the Negative Log-Likelihood estimation results of related methods on ETH/UCY, NBA, SDD, and NuScenes datasets. High probabilities meaned low NLL.

| | Eth | Hotel | Univ | Zara01 | Zara02 | Rebound | Scores | SDD | Nuscene |
|---|---|---|---|---|---|---|---|---|---|
| **MANTRA [6]** | 3.36 | 0.67 | 0.73 | 0.98 | 0.06 | 3.06 | 3.24 | 3.36 | 4.03 |
| **PECNet [4]** | 3.28 | 0.16 | 0.86 | 0.66 | 0.13 | 2.87 | 3.03 | 3.28 | 3.92 |
| **Trajection++ [8]** | 2.26 | −0.52 | 0.36 | 0.12 | −0.92 | 2.56 | 2.81 | 2.26 | 3.24 |
| **BiTraP [16]** | 3.69 | 0.49 | 0.79 | 0.56 | −0.58 | 3.21 | 3.54 | 3.69 | 3.74 |
| **SGNet [9]** | 2.65 | 0.94 | 1.41 | 0.24 | −0.69 | 3.41 | 2.95 | 2.65 | 2.97 |
| **GroupNet+CVAE [13]** | 1.36 | −1.23 | 0.66 | −0.58 | −2.74 | 1.98 | 1.82 | −0.24 | 1.96 |
| **Ours** | **0.82** | **−1.84** | **−0.32** | **−1.21** | **−2.97** | **1.71** | **0.94** | **−1.61** | **1.24** |

which contains two sub-networks, to hierarchically model and predict agents' trajectories with trajectory spectrums.

- E-V2-Net [11]:This paper brought a new "view" for trajectory prediction to model and forecast trajectories hierarchically according to different frequency portions from the spectral domain and learned to forecast trajectories by considering their frequency responses.
- SocialCircle [12]:Inspired by marine animals that localize the positions of their companions underwater through echoes, it built a new anglebased trainable social interaction representation, named SocialCircle, for continuously reflecting the context of social interactions at different angular orientations relative to the target agent.
- SGNet [9]: It presented a recurrent network called Stepwise Goal-Driven Network. It incorporated an encoder that captured historical information, a stepwise goal estimator that predicted successive goals into the future, and a decoder that predicted future trajectories.
- GroupNet [13]: It proposed a trainable multiscale hypergraph to capture both pair-wise and group-wise interactions at multiple group sizes. From the aspect of interaction representation learning, it proposed a three-element format that could be learned end-to-end and explicitly reasoned some relational factors including the interaction strength and category.
- SHENet [7]: This paper proposed to forecast a person's future trajectory by learning from implicit scene regularities. It categorized scene history information into two types: historical group trajectory and individual-surroundings interaction.
- MID [2]: It presented a new framework that formulated the trajectory prediction task as a reverse process of motion indeterminacy diffusion. It adjusted the length of the chain to control the degree of indeterminacy and balanced the diversity and determinacy of the predictions.
- Y-Net [5]: It modeled epistemic uncertainty through multimodality in long-term goals and aleatoric uncertainty through multimodality in waypoints and paths.
- NSP-SFM [19]: This paper proposed Neural Social Physics combining both methodologies based on a new Neural Differential Equation model. It was a deep neural network within which it used an explicit physics model with learnable parameters.
- PECNet [4]: This paper presented Predicted Endpoint Conditioned Network for flexible human trajectory prediction. PECNet inferred distant trajectory endpoints to assist in long-range multi-modal trajectory prediction.
- Trajectiopn++ [8]: It presented Trajectron++, a modular, graph-structured recurrent model that forecasted the trajectories of a general number of diverse agents while incorporating agent dynamics and heterogeneous data.

## 4. Comparison Experiments on NLL Metrics

The trajectory distribution approximator captured the distribution of latent variables while continuously updated it through RNN joint predictions. The results in Table. 2 indicated that the proposed method effectively captured dynamic changes in agents. The findings demonstrated superior prediction quality across all scenarios in each dataset, showing higher probabilities (lower NLL) on Ground Truth trajectories.

## 5. Experimental Analysis on Directional Component A

Experiments were conducted on NBA dataset using three methods with similar loss function architectures. As shown in Table 3, the inclusion of Component **A** led to reductions in both ADE and FDE metrics. This suggested that the directional component in the loss function was beneficial not only for the method proposed in this study but also significantly enhanced other models. This outcome further supported the plug-and-play characteristic of Component **A**.

## 6. Qualitative Analysis on NBA Dataset

Fig. 1 visualized the trajectory predictions of the proposed method on the Rebound and Scores subsets. In (a1), (a4), and (b5), the target agents moved along the court boundaries, and SocialMOIF did not predict trajectories beyond these boundaries. This demonstrated that SocialMOIF not only learned complex interaction patterns between agents but also understood the existence of boundary constraints. In (a2), (b4), and (b8), the target agents made sudden turning decisions to counter opponents, and SocialMOIF suc-

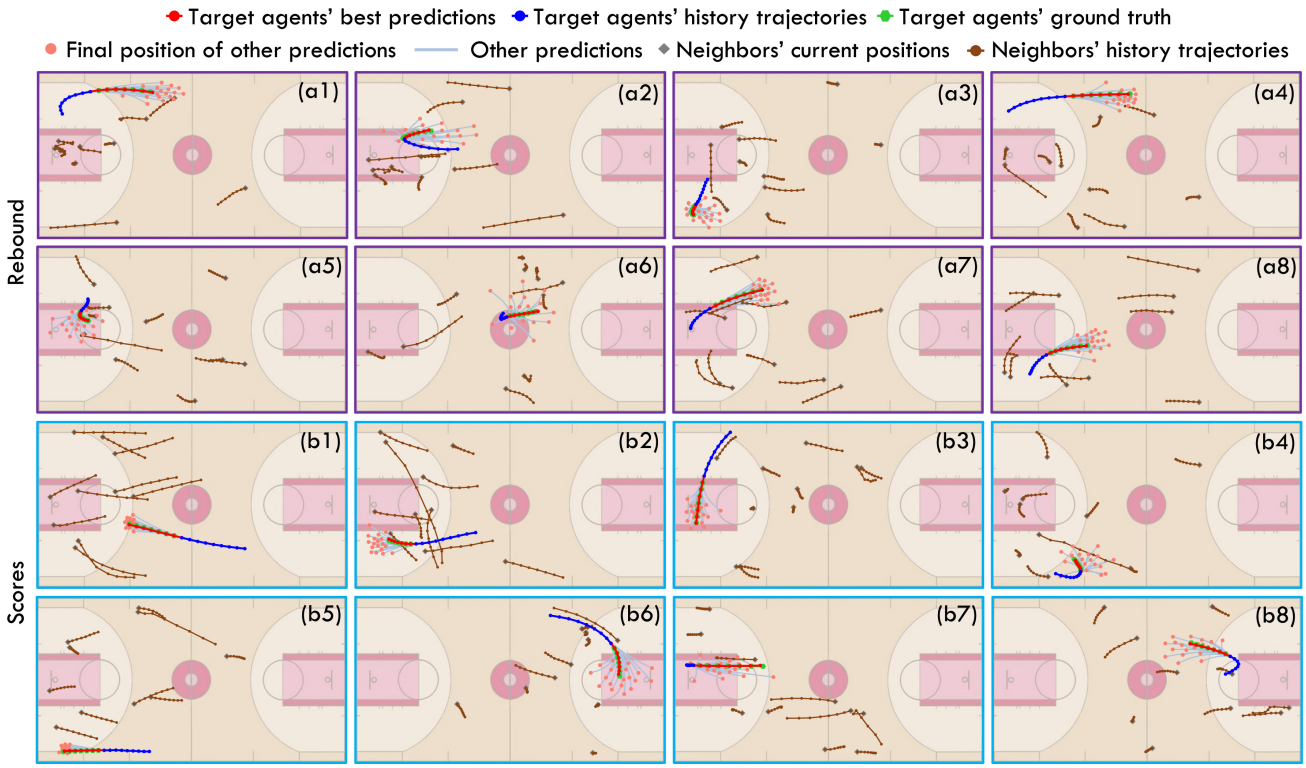


Figure 1. Visualization of predictions in NBA datasets(Rebound and Scores). Each target agent included 20 randomly generated trajectories. The best predictions were shown in **red**. Basketball was also treated as neighbors.

Table 3. Comparison experiments of Component A with other methods.

| | A | Rebound | Scores |
|---|---|---|---|
| **BiTraP** [16] | | 0.86/1.88 | 0.76/1.55 |
| | √ | **0.84/1.79** | **0.72/1.49** |
| **SGNet** [9] | | 0.80/1.78 | 0.71/1.45 |
| | √ | **0.78/1.72** | **0.66/1.39** |
| **GroupNet+CVAE** [13] | | 0.63/0.84 | 0.46/0.76 |
| | √ | **0.58/0.78** | **0.44/0.72** |

cessfully captured these adversarial strategies, accurately predicting their future trajectories. In (a3), (a5), (a6), and (b7), the target agents exhibited relatively conservative movement behaviors. In addition to providing accurate predictions, SocialMOIF offered decision-making options in multiple directions, showcasing its versatility in handling diverse scenarios.

## 7. Limitation and Future Work

SocialMOIF revisits social interactions among agents from a fresh perspective. While this method provides valuable insights, relying solely on historical trajectory data poses challenges in fully capturing the agents' intentions. To address this limitation, future work will enhance the model by integrating heterogeneous data, enabling the network to learn richer semantic information and better understand complex social dynamics.

## References


[1] Kai Chen, Xiao Song, Haitao Yuan, and Xiaoxiang Ren. Fully convolutional encoder-decoder with an attention mechanism for practical pedestrian trajectory prediction. *IEEE Transactions on Intelligent Transportation Systems*, 23(11): 20046–20060, 2022. 1

[2] Tianpei Gu, Guangyi Chen, Junlong Li, Chunze Lin, Yongming Rao, Jie Zhou, and Jiwen Lu. Stochastic trajectory prediction via motion indeterminacy diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 17113–17122, 2022. 2

[3] Lihuan Li, Maurice Pagnucco, and Yang Song. Graph-based spatial transformer with memory replay for multi-future pedestrian trajectory prediction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2221–2231, 2022. 1

[4] 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 *Computer Vision–*

*ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*, pages 759–776. Springer, 2020. 2

[5] Karttikeya Mangalam, Yang An, Harshayu Girase, and Jitendra Malik. From goals, waypoints & paths to long term human trajectory forecasting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15233–15242, 2021. 2

[6] Francesco Marchetti, Federico Becattini, Lorenzo Seidenari, and Alberto Del Bimbo. Mantra: Memory augmented networks for multiple trajectory prediction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7143–7152, 2020. 1, 2

[7] Mancheng Meng, Ziyan Wu, Terrence Chen, Xiran Cai, Xiang Zhou, Fan Yang, and Dinggang Shen. Forecasting human trajectory from scene history. *arXiv preprint arXiv:2210.08732*, 2022. 2

[8] Tim Salzmann, Boris Ivanovic, Punarjay Chakravarty, and Marco Pavone. Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVIII 16*, pages 683–700. Springer, 2020. 2

[9] Chuhua Wang, Yuchen Wang, Mingze Xu, and David J Crandall. Stepwise goal-driven networks for trajectory prediction. *IEEE Robotics and Automation Letters*, 7(2):2716–2723, 2022. 2, 3

[10] 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, 2022. 1

[11] 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. 2

[12] Conghao Wong, Beihao Xia, Ziqian Zou, Yulong Wang, and Xinge You. Socialcircle: Learning the angle-based social interaction representation for pedestrian trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19005–19015, 2024. 2

[13] 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*, pages 6498–6507, 2022. 2, 3

[14] Chenxin Xu, Weibo Mao, Wenjun Zhang, and Siheng Chen. Remember intentions: Retrospective-memory-based trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6488–6497, 2022. 1

[15] Chenxin Xu, Robby T Tan, Yuhong Tan, Siheng Chen, Yu Guang Wang, Xinchao Wang, and Yanfeng Wang. Eqmotion: Equivariant multi-agent motion prediction with invariant interaction reasoning. *arXiv preprint arXiv:2303.10876*, 2023. 1

[16] Yu Yao, Ella Atkins, Matthew Johnson-Roberson, Ram Vasudevan, and Xiaoxiao Du. Bitrap: Bi-directional pedestrian trajectory prediction with multi-modal goal estimation. *IEEE Robotics and Automation Letters*, 6(2):1463–1470, 2021. 1, 2, 3

[17] Ziyi Yin, Ruijin Liu, Zhiliang Xiong, and Zejian Yuan. Multimodal transformer networks for pedestrian trajectory prediction. In *IJCAI*, pages 1259–1265, 2021. 1

[18] Ye Yuan, Xinshuo Weng, Yanglan Ou, and Kris M Kitani. Agentformer: Agent-aware transformers for socio-temporal multi-agent forecasting. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9813–9823, 2021. 1

[19] Jiangbei Yue, Dinesh Manocha, and He Wang. Human trajectory prediction via neural social physics. In *European conference on computer vision*, pages 376–394. Springer, 2022. 2