Trajectory Unified Transformer for Pedestrian Trajectory Prediction —Supplementary Materials—

Liushuai Shi[1] Le Wang[1*] Sanping Zhou[1] Gang Hua[2]
[1]National Key Laboratory of Human-Machine Hybrid Augmented Intelligence,
National Engineering Research Center for Visual Information and Applications,
Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University
[2]Wormpex AI Research

Abstract

This supplementary material provides detailed implementations, more experiments and more visualizations to further evaluate the effectiveness of our proposed method.

1. Detailed Implementation

Trajectory Augmentation. In the process of global prediction, we obtain general motion modes represented by the trajectories with length T_{pred}. In the implementation, we extract multiple sub-trajectories with length T_{pred} for a single trajectory with length T by a sliding window with length T_{pred}, where $T = T_{obs} + T_{pred}$. In this case, the trajectories show more various motion behaviors and thus the obtained general motion modes could better cover the common motion behaviors of a pedestrian.

Trajectory Smooth. In global prediction, we use distance measurement (K-means) to obtain the general motion modes. To deal with the sensibility of K-means for the abnormal trajectory (*e.g.*, turning sharp), we smooth the trajectory before the K-means operation. In this paper, we use the moving average by sliding windows with length l to obtain smooth trajectories. $l = 3$ in our implementation.

Random Rotation. We randomly rotate the trajectory before the distance measurement in global prediction to deal with the trajectory imbalance among different scenes.

2. More Experiments

Comparison in Inference Speed We conduct extensive experiments to compare the methods without the post-processing in inference speed. As shown in Table I, our method outperforms the PECNet and is of the same order of magnitude in speed as SGAN and SIT. Moreover, our method owns the best accuracy performance as shown in Table 1. It strongly demonstrates that our method achieves a balance between speed and accuracy.

Smooth Label and One-hot Label In our method, we use a smooth label to supervise the classification, while the smooth label is sensitive to the trajectory state. Because labels with lower scores could disturb the model training. Thus, we make a comparison between the one-hot label (OHL) and the smooth label (SL). As shown in Table II, the experimental results on ETH-UCY and SDD demonstrate that the smooth label is better than the one-hot label. The reason could be that the smooth label allows the model to select diverse motion modes better than the one-hot label. Of course, the smooth label is a suboptimal alternative to the one-hot label due to the critical sensitives.

Number of motion modes on ETH The selection of the number of motion modes on ETH is shown in Table III.

More analysis about the Table 3 and Table 4. In mathematics, the worst result of p in brier-ADE/FDE is $1/k$, where k is the number of predicted trajectories. In this case, the predicted trajectory probabilities follow a uniform distribution. However, the goal of predicting trajectory probabilities is to reduce the indeterminacy in trajectory selection. It is invalid that directly add $1/k$ on the ADE/FDE to measure the ability to select the more likely predicted trajectories because the uniform distribution has maximum entropy and thus can not give any useful information on trajectory selection. Therefore, we have made efforts to build model variants of CAGN and SocialVAE+FPC to measure their ability in trajectory selection. Unfortunately, the model variants show inferior performance measured by brier-ADE/FDE. and the experimental results of model variants in brier-ADE/FDE are worse than the worst case. A speculative reason is that the classification task stresses the model leading to inferior predicted trajectories.

*Corresponding author.

N	SGAN	PECNet	SIT	Ours
10	0.0598s	0.1171s	0.0581s	0.0561s

Table I. Comparisons the methods without post-processing in inference time recorded in seconds on SDD dataset. The Lower the better.

	ETH	HOTEL	UNIV	ZARA1	ZARA2	SDD
OHL	0.44/0.63	0.13/0.19	**0.23**/0.43	**0.18/0.34**	**0.13/0.25**	7.80/12.83
SL	**0.40/0.61**	**0.11/0.18**	**0.23/0.42**	**0.18/0.34**	**0.13/0.25**	**7.76/12.69**

Table II. Comparison between the one-hot label and smooth label on ETH-UCY and SDD. The Lower the better.

	20	30	50	60	90	150
ETH	0.45/0.67	0.43/0.69	**0.40/0.61**	0.41/0.61	0.44/0.63	0.50/0.73

Table III. The ablation study of the number of motion modes on ETH in ADE/FDE. The lower the better.

3. More Visualizations

We provide more visualizations of predicted diverse trajectories randomly sampled from the dataset as shown in Figure I, Figure II and Figure III.

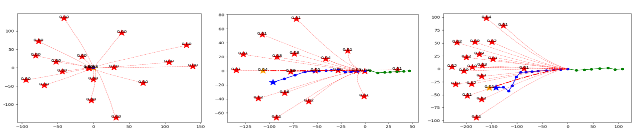

Figure I. More visualizations.

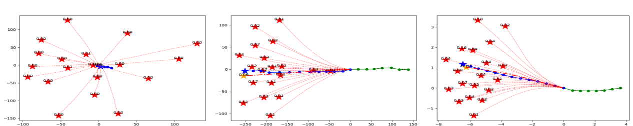

Figure II. More visualizations.

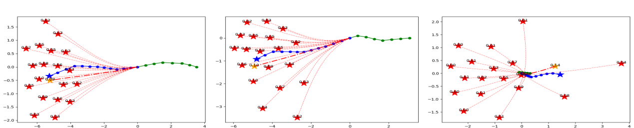

Figure III. More visualizations.