TrajFine: Predicted Trajectory Refinement for Pedestrian Trajectory Forecasting
— Supplementary Material —

1. Inference Details

As Section 3.3 of the main paper mentions, we introduce the noise terms ϵ_θ and ϵ_ϕ to enhance the sampling process for Y_{k-1}. We present the complete inference pipeline in Algorithm 1 to detail TrajFine. Under the DDIM sampling technique, we notate the DDIM timestamps as T, typically smaller than the maximum diffusion timestamps K, to expedite computations.

Algorithm 1 Sampling Pipeline

Input: X
Output: ϵ_θ, ϵ_ϕ, Y_0
$Y_T \sim \mathcal{N}(0, \mathbf{I})$
$c_{pc} \leftarrow \mathcal{F}_{pc}(X)$ extract past context
for $\tau = T, ..., 1$ **do**
  $\epsilon_\theta \leftarrow$ Intra-TNP(Y_τ, τ, c_{pc})
  Sample $\hat{Y}_{\tau-1}$
  $c_{fc} \leftarrow \mathcal{F}_{fc}(\hat{Y}_{\tau-1})$ extract future context
  $\epsilon_\phi \leftarrow$ Inter-TNP($\hat{Y}_{\tau-1}$, τ, c_{fc})
  $\hat{\epsilon} = \epsilon_\theta + \epsilon_\phi$
  DDIM sampling by $\hat{\epsilon}$
return Y_0

2. Supplementary Experiments

2.1. Performance Using Few Data

In this section, we explore the impact of using different amounts of training data on performance with $\mathcal{L}_{sep}$ training. We conducted training on datasets with proportions of 5%, 10%, 50%, 70%, and 100%. Table 1 shows that when the training data is scarce, the performance is notably below standard. This phenomenon is attributed to TrajFine's reliance on diffusion-based and transformer backbone methods, which exhibit a data-hungry characteristic. However, as the data increases to 50%, the performance approaches training on the entire dataset, with the ratio at 70%, even rivaling or slightly surpassing the performance achieved with the complete training dataset. This improvement is likely owing to the increased data complexity introduced by our Scene Mixup method, enhancing the model's robustness and generalization. Additionally, most of the complete training data consists of trajectories characterized by uniform walking speeds. This dominance of normal pathways may lead the model to overfit to such standard trajectories. Consequently, during inference, the model may struggle to accurately predict deviations in trajectories, such as sharp turns, halts, or U-turns, due to its excessive adaptation to the prevalent and routine walking patterns.

Training Data Ratio	ADE / FDE
5%	12.97 / 24.24
10%	10.00 / 19.16
50%	7.94 / 15.03
70%	**7.22 / 13.50**
100%	**7.22** / <u>13.79</u>

Table 1. Effect of training data ratios on the SDD Dataset. We present differently available training data ratios for TrajFine while employing identical training strategies.

2.2. Visualizing Various DDIM Timestamps

Figure 1 shows that as the number of DDIM timestamps increases, the predicted trajectories become smoother and align more closely with the ground truth. However, considering the trade-off between computation time and prediction performance, we choose a balance $T = 10$ between efficiency and predictive accuracy, which offers both reasonably high efficiency and a satisfactory level of prediction quality.

2.3. Components Analysis on ETH-UCY Dataset

In this section, we explore the effectiveness of each component on the ETH-UCY dataset with $\mathcal{L}_{sep}$ training. Table 2 shows that the proposed components, including TrajFine, Scene Mixup, and Length-aware CL, contribute to performance across these sub-datasets. We observe that each element brings a positive contribution but is marginal due

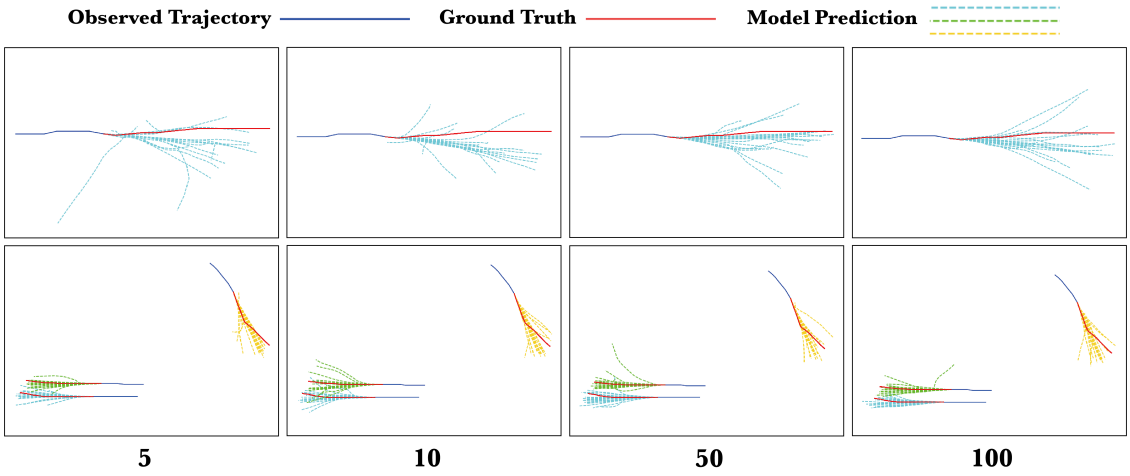

Figure 1. Visualization for hidden DDIM timestamps. We present two instances that are generated from different timestamps. The top row shows a single-agent case, while the bottom row illustrates the multi-agent situation. The numbers stand for timestamps T.

Components		ADE / FDE				
TrajFine	**Scene Mixup & Length CL**	**ETH**	**HOTEL**	**UNIV**	**ZARA1**	**ZARA2**
		0.35 / 0.63	0.13 / 0.22	0.22 / 0.48	0.18 / 0.39	0.13 / 0.29
	✓	0.35 / 0.62	0.13 / 0.24	0.21 / 0.44	0.17 / 0.37	0.12 / 0.27
✓		0.34 / 0.59	0.12 / 0.21	0.22 / 0.47	0.18 / 0.38	0.13 / 0.28
✓	✓	**0.34 / 0.58**	**0.11 / 0.19**	**0.21 / 0.44**	**0.17 / 0.35**	**0.12 / 0.27**

Table 2. Component Analysis on subdatasets of ETH-UCY. ✓denotes that the component is included. Length CL means length-aware CL.

to the saturated evaluation of ETH-UCY. As a result, the full model, *i.e.*, adopting the TrajFine module with Scene Mixup augmentation under a length-aware CL scenario, obtains the lowest error on all subsets, indicating the proposed method improves the accuracy for the trajectory prediction task.