This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TCYB.2025.3613498, IEEE Transactions on Cybernetics

Social Informer: Pedestrian Trajectory Prediction by Informer With Adaptive Trajectory Probability Region Optimization

Zihan Jiang, *Student Member, IEEE*, Rui Yang, *Senior Member, IEEE*, Yiqun Ma, Chengxuan Qin, Xiaohan Chen, *Student Member, IEEE*, Zidong Wang, *Fellow, IEEE*

***Abstract*—Pedestrian trajectory prediction is an important research area with significant applications in autonomous driving and intelligent surveillance. However, existing studies on pedestrian trajectory prediction often suffer from a noticeable discrepancy between predicted and actual trajectories, due to incomplete extraction of pedestrian trajectory features and the randomness of the pedestrian walking process. The key objective of this paper is to address this issue by proposing a method that can reasonably simulate the randomness of pedestrian walking and comprehensively extract pedestrian trajectory features. To achieve this, a novel social informer model built upon the informer model is proposed in this paper. The social informer utilizes a transformer encoder-based interaction module to comprehensively extract pedestrian trajectory features, which are input into the informer model for further processing. Additionally, an adaptive variance mechanism is proposed to determine the optimal variance and accurately simulate the random nature of pedestrian walking. Finally, the proposed model is evaluated in a comparative experiment on ETH and UCY datasets, with results demonstrating that the proposed model outperforms other models, exhibiting improved accuracy and performance.**

***Index Terms*—Trajectory prediction, adaptive variance mechanism, informer model, variety loss**

I. Introduction

The paper "125 Questions: Exploration and Discovery" published in Science in 2021, poses a challenging question about the feasibility of relying solely on self-driving cars in the future [41]. Nowadays, the field of self-driving technology [79] has gained significant attention [56], [61], [81], [82] and research focus [29], [51], [54], [77]. Predicting pedestrian trajectories is crucial in ensuring safe and reliable vehicle movement by enabling effective planning to prevent accidents [15], [17], [78]. Therefore, pedestrian trajectory prediction has become an essential task in the realm of self-driving technology [30], [36], [58], [65]. Existing methods for pedestrian trajectory forecasting can be broadly classified into three categories: (1) physical model-based methods, (2) classical machine learning-based methods, and (3) deep learning-based methods.

In physical model-based methods, pedestrian trajectory is predicted by modeling the intuitive characteristics of pedestrians (such as velocity, acceleration, and displacement) and generalizing past states to forecast future trajectories based on different physical models [35], [46], [58]. Classical machine learning-based methods for pedestrian trajectory prediction shift the focus from model-driven approaches to data-driven ones. Numerous algorithms have been widely utilized in this category [13], [59], [63], such as support vector machine (SVM), the Gaussian process (GP), dynamic Bayesian network (DBN), hidden Markov model (HMM), k-nearest neighbors (KNN), and decision tree [9], [18], [30], [36]. Deep learning-based methods for pedestrian trajectory prediction treat the problem as a sequence generation task, learning valuable information from past pedestrian trajectories to predict future ones. Nowadays, recurrent neural network (RNN) [1], generative adversarial networks (GAN) [14], [34], [38], and graph convolution neural networks (GCN) [22], [31], [47] are widely used in trajectory prediction studies.

Physical model-based methods predict pedestrian trajectories by modeling intuitive properties such as velocity, acceleration, and displacement, using various physical frameworks [35], [46], [58]. In contrast, classical machine learning approaches shift toward data-driven strategies, employing algorithms like support vector machine (SVM), the Gaussian process (GP), dynamic Bayesian network (DBN), hidden Markov model (HMM), k-nearest neighbors (KNN), and decision tree [9], [18], [30], [36]. More recently, deep learning methods formulate trajectory prediction as a sequence generation task, leveraging historical data through models such as recurrent neural network (RNN) [1], generative adversarial networks (GAN) [14], [34], [38], and graph convolution neural networks (GCN) [22], [31], [47].

Despite impressive progress, pedestrian trajectory prediction remains challenging due to incomplete tracking information caused by unknown destinations, random disturbances, or uncertain intentions. Human motion's inherently multi-modal and stochastic nature necessitates modeling diverse factors to

This research is supported by: Jiangsu Provincial Scientific Research Center of Applied Mathematics (BK20233002), Jiangsu Provincial Qinglan Project, Suzhou Science and Technology Programme (SYG202106), and Research Development Fund of XJTLU (RDF-20-01-18).

Z. Jiang is with the Shanghai Research Institute for Intelligent Autonomous Systems, Tongji University, Shanghai 200092, China, and also with School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, 215123, China (e-mail: Zihan.Jiang22@alumni.xjtlu.edu.cn);

Y. Ma, C. Qin and X. Chen are with School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, 215123, China, and also with School of Electrical Engineering, Electronics and Computer Science, University of Liverpool, Liverpool, L69 3BX, United Kingdom (e-mail: Yiqun.Ma22@alumni.xjtlu.edu.cn, C.Qin8@liverpool.ac.uk, Xiaohan.Chen20@alumni.xjtlu.edu.cn).

R. Yang is with School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, 215123, China (e-mail: R.Yang@xjtlu.edu.cn).

Z. Wang is with the Department of Computer Science, Brunel University London, Uxbridge, Middlesex, UB8 3PH, United Kingdom (e-mail: Zidong.Wang@brunel.ac.uk).

Corresponding author: R. Yang

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TCYB.2025.3613498, IEEE Transactions on Cybernetics

ensure accurate prediction. With advances in deep learning [2], [6], [8], [16], [37], [42], multi-modal trajectory prediction has gained increasing attention [14], [22], [38]. According to a recent survey [48], existing deep learning-based methods for multi-modal prediction can be categorized into four types: (1) GAN-based [22], [34], [38], [66]; (2) VAE-based [27], [28]; (3) reinforcement learning-based [7], [12], [43], [52]; and (4) transformer-based [20], [26], [50], [67], [72], [75].

Previous research in multi-modal pedestrian trajectory prediction has made significant contributions and successfully addressed numerous complex challenges [14], [22], [38]. According to the literature research, various efforts have been made to model the randomness of pedestrian movement in the absence of pedestrian-related information, such as the intention and perspective of pedestrians, and the relationship between pedestrians [10], [68]. While these approaches have resolved many issues [60], [70], [73], certain limitations persist, highlighting two main challenges in multi-modal pedestrian trajectory prediction: (1) how to model the social interactions of pedestrians effectively; (2) how to rationally model the randomness of pedestrian trajectories.

CNN has been widely adopted for modeling social interactions in pedestrian trajectory prediction [14], [22], [34], [38]. While CNNs excel at capturing local spatial features [4], [5], [40], [45], [53], [57], [76], [79], [80], they struggle with global spatial dependencies. Prior studies [14], [34] indicate that incorporating richer spatial information improves prediction accuracy, motivating the development of models with enhanced spatial modeling capabilities. To address trajectory uncertainty, many works [14], [22], [34], [38] use probabilistic sampling, typically assuming a standard Gaussian distribution. Although such an assumption aligns with common motion statistics, it limits prediction accuracy. The sampling variance should adapt to varying data and model conditions in real-world scenarios to better capture trajectory diversity.

Based on the above motivations, this paper proposes an improved model called the social informer. The social informer incorporates an adaptive variance mechanism, by representing the randomness of pedestrian trajectories, to determine the most appropriate variance of datasets and models through extensive training and comparison. Compared with the conventional informer model, this social informer employs a transformer encoder in the feature extraction module, enabling the extraction of both local and global spatial features of pedestrian relations. To summarize, this paper makes the following contributions:

1) Social Informer is a novel trajectory prediction framework that combines the long-sequence modeling strength of informer with a transformer-based interaction encoder. Unlike prior models relying on static graphs or fixed neighborhoods, it learns dynamic social patterns via self-attention, enabling adaptive modeling of diverse pedestrian behaviors;
2) We propose an adaptive variance mechanism that dynamically adjusts the trajectory sampling spread based on behavioral consistency, instead of using a fixed Gaussian prior. Our method optimizes the sampling variance based on downstream behavioral consistency, enabling the model to regulate the expressiveness of its prediction space dynamically;
3) This paper adopts a sampling-based multimodal prediction strategy guided by the variety loss, which ensures that the closest one to the ground truth contributes to the loss among multiple generated trajectories. In contrast to standard training, this encourages the model to explore the multimodal distribution space and reduces mode averaging.

The rest of this paper is presented as follows. Section II describes the principle of transformer and informer models. Section III gives the specific method of the proposed social informer. Section IV introduces the dataset and experiments used in this paper, and the experimental results are analyzed in detail. Finally, conclusions and discussions on relevant future works are presented in Section V.

II. Preliminary Knowledge

A. Informer

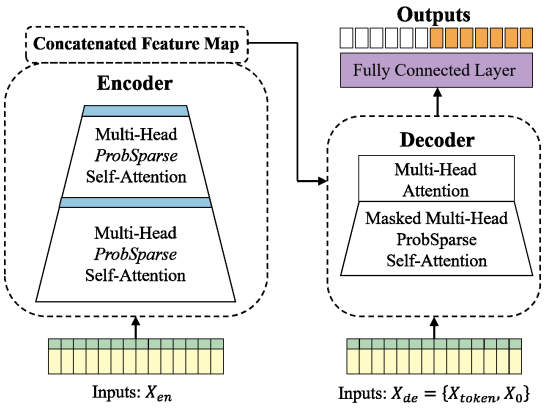


Fig. 1. Schematic diagram of informer model

The standard transformer model [49] faces challenges in long sequence time-series forecasting due to its high memory usage, quadratic complexity, and architectural inefficiencies. To address these limitations, the Informer model [74] introduces three key innovations: (1) ProbSparse self-attention to reduce computation cost; (2) a distillation mechanism to compress sequence representations; and (3) a generative decoder to produce predictions in one step.

Unlike transformer variants such as the multi-granularity scenarios understanding network [64], which rely on separate attention modules for scene and interaction modeling, Informer captures spatiotemporal dependencies through a unified architecture. This design avoids handcrafted attention fusion and enhances generalization while maintaining a lightweight structure. As shown in the schematic diagram in Figure 1, the informer model follows the encoder-decoder architecture, replacing canonical self-attention with ProbSparse self-attention:

$$\mathcal{A}(\mathbf{Q},\mathbf{K},\mathbf{V}) = \mathrm{Softmax}\left(\frac{\overline{\mathbf{Q}}\mathbf{K}^{\top}}{\sqrt{d}}\right)\mathbf{V} \tag{1}$$

where $\mathbf{Q}$ stands for the query vector, $\mathbf{K}$ for the key vector, $\mathbf{V}$ for the value vector, $\overline{\mathbf{Q}}$ represents the sparse matrix derived

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TCYB.2025.3613498, IEEE Transactions on Cybernetics

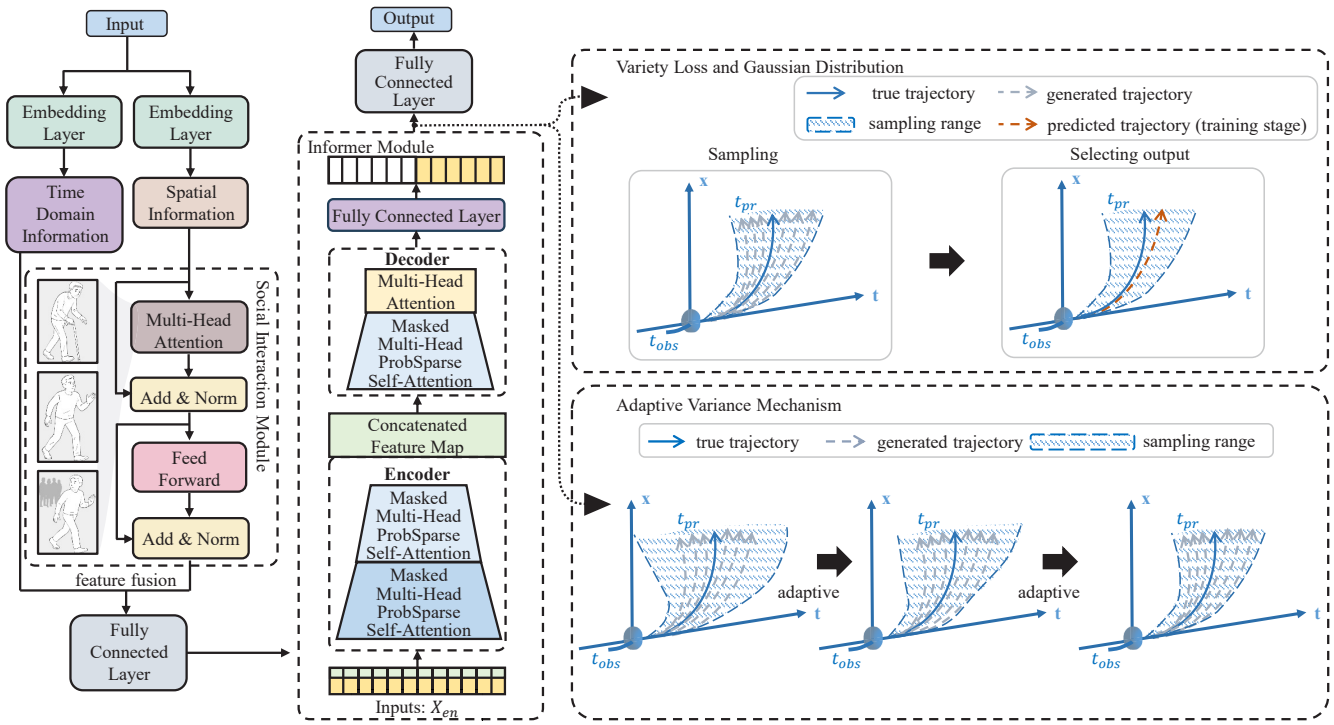


Fig. 2. The social informer model network diagram

from the sparsity measure, and d is the dimension of the $\mathbf{K}$-vector. Additionally, $\overline{\mathbf{Q}}$ consists of the first n query terms selected based on the sparsity measure $M(\mathbf{q}, \mathbf{K})$:

$$M\left(\mathbf{q}_i, \mathbf{K}\right) = \ln \sum_{j=1}^{L_K} e^{\frac{\mathbf{q}_i \mathbf{k}_j^\top}{\sqrt{d}}} - \frac{1}{L_K} \sum_{j=1}^{L_K} \frac{\mathbf{q}_i \mathbf{k}_j^\top}{\sqrt{d}} \tag{2}$$

where $\mathbf{q}_i$ represents the i-th row of the Q vector, and $\mathbf{k}_j$ represents the j-th column of the K vector.

B. Problem description

This paper formulates pedestrian trajectory prediction as a sequence-to-sequence generative task. Given a historical trajectory $X = x_1, x_2, \ldots, x_T$, where x_T encodes position and velocity at time T, the goal is to predict future trajectories $Y = y_1, y_2, \ldots, y_T$. We model the output as a probability distribution to bridge the inherent gap between predicted and actual trajectories. Prior studies have shown that appropriate probabilistic modeling enhances realism and prediction accuracy [14], [22]. The trajectory prediction task can thus be formulated as a probabilistic sequence generation problem, as shown in (3):

$$X_{actual}(t) = \hat{X}(t) + \epsilon(t) - p(t) \tag{3}$$

where $X_{actual}(t)$ denotes the actual position of the pedestrian at time t, $\hat{X}(t)$ denotes the model-predicted location of the pedestrian at time t, $\epsilon(t)$ denotes disturbance factors, and $p(t)$ denotes the probability distribution. Therefore, it can be inferred that accurate pedestrian trajectory prediction can be achieved when the difference between the probability distribution $p(t)$ and the disturbance factor $\epsilon(t)$ is closest to the difference between $X_{\text{actual}}(t)$ and $\hat{X}(t)$.

III. Social Informer

This section comprehensively explains the social informer model, illustrated in Figure 2. Building upon the informer model, the social informer encompasses three essential components, explained in the following subsections: (1) social interaction module; (2) our loss; and (3) adaptive variance mechanism.

A. Social interaction module

The social interaction module aims to capture not only spatiotemporal dependencies among pedestrians but also latent social cues such as interpersonal distance, movement synchronization, and implicit group dynamics—factors that often extend beyond mere positional proximity. Studies have shown that a pedestrian's perceptual range varies across individuals [3], [23], [24], influenced by a combination of physiological, psychological, and environmental factors. Physiologically, age and health conditions affect perception, with children, the elderly, and those experiencing fatigue exhibiting reduced awareness. Psychologically, attention, emotional state, and personality traits also shape perception—for example, anxiety narrows while extroversion broadens perceptual range. Environmental factors like lighting, weather, and spatial familiarity further modulate these effects. These complexities underscore the importance of incorporating heterogeneous perceptual modeling. This process is shown in (4):

$$P_r = f\left(\alpha \cdot F_{phy}, \beta \cdot F_{psy}, \gamma \cdot F_{env}\right) \tag{4}$$

where P_r represents the perceived range of pedestrians, F_{phy}, F_{psy}, and F_{env} represent the influence of physiological, psy-

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TCYB.2025.3613498, IEEE Transactions on Cybernetics

chological, and environmental factors, respectively, and α, β, γ are the weighting coefficients. The process is shown explicitly in Figure 3.

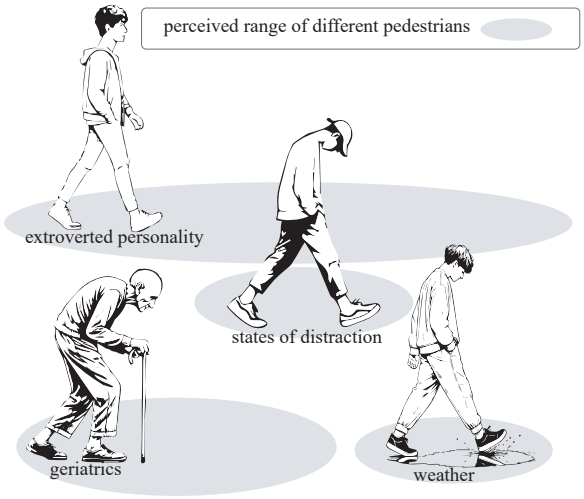


Fig. 3. Differences in the perceived range of different pedestrians

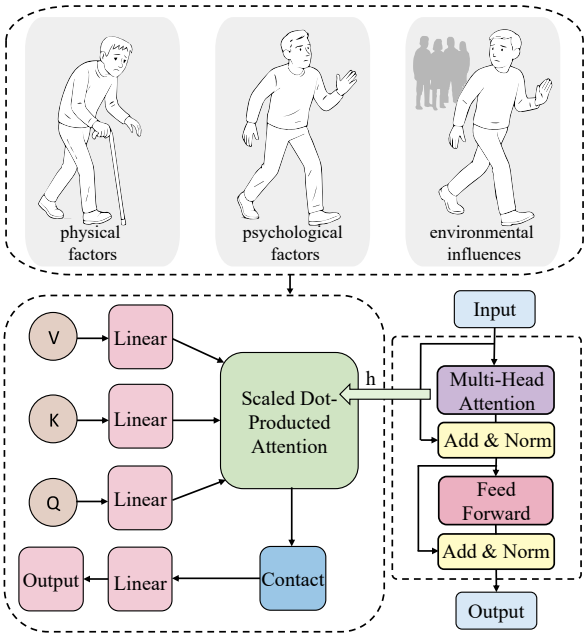


Fig. 4. Schematic diagram of transformer encoder

Traditional pedestrian interaction models often rely on fixed prior assumptions: the social force model assumes uniform psychological factors (F_{phy}) [11], CNN-based models presume shared physical characteristics (F_{psy}), and GCN-based methods assume identical environmental influences (F_{env}). In contrast, our Transformer-based interaction module avoids such hard-coded priors by learning social dynamics in a data-driven manner. Through self-attention, each pedestrian can attend to all others within the observation window, with interaction weights learned adaptively from data. This flexibility enables the model to capture complex and emergent social behaviors without explicit rules, as Figure 4 illustrates.

The transformer encoder serves as the interaction module in the proposed Social informer and is responsible for extracting spatiotemporal features from pedestrian trajectories. Figure 4 comprises three core components: multi-head attention, add & norm, and feed-forward layers. In the multi-head attention module, Q, K, and V denote the Query, Key, and Value matrices, respectively; "linear" refers to the learned projections applied to them; and h indicates the number of parallel scaled dot-product attention heads, as shown in (5):

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right) \cdot V \tag{5}$$

where d_k is the number of columns of matrices Q and K. This method allows each pedestrian to adaptively focus on others in the scene, regardless of physical proximity or predefined interaction graphs. Compared with CNN-based models that rely on local convolutional windows and GCN-based methods that require fixed neighborhood connections, the transformer encoder enables a fully data-driven modeling of social influence, automatically learning which agents are socially relevant in each situation.

Although the model does not explicitly encode behavioral factors, the learned attention weights implicitly capture latent social cues underlying pedestrian dynamics. These include: (1) physiological factors (e.g., walking speed or gait), as attention often focuses on agents with abrupt or conflicting motion; (2) psychological states (e.g., hesitation or uncertainty), reflected in temporal irregularities; and (3) environmental conditions (e.g., distance and crowd density), which influence spatial attention patterns. This formulation enables the model to represent diverse social behaviors without relying on hand-crafted priors or static interaction rules.

B. Uncertainty-Aware Variety Loss for Multi-modal Prediction

Various factors in pedestrian walking may lead to multiple plausible future trajectories. Traditional prediction methods often apply the L2 loss to minimize the distance between predicted and actual trajectories, as shown in (6):

$$L_2 = \sum_{i=t_{obs}}^{t_{obs}+t_{pre}} \left(Y_i - \hat{Y}_i\right)^2 \tag{6}$$

where Y_i and $\hat{Y}_i$ denote the ground truth and prediction, respectively.

However, the L2 loss tends to produce average trajectories (see Figure 5(a)), failing to reflect the pedestrian motion's inherent randomness. To address this, we adopt a variety loss based on (6), as shown in (7):

$$L_{\text{variety}} = \min_k \left|Y_i - \hat{Y}_i^k\right|_2 \tag{7}$$

which encourages generating diverse trajectory samples and selecting the closest one for loss computation. Where k represents the number of generated samples, the variety loss function encourages the model to explore multiple potential trajectories rather than relying on a single averaged prediction, as shown in Figure 5(b). However, it is crucial to note

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TCYB.2025.3613498, IEEE Transactions on Cybernetics

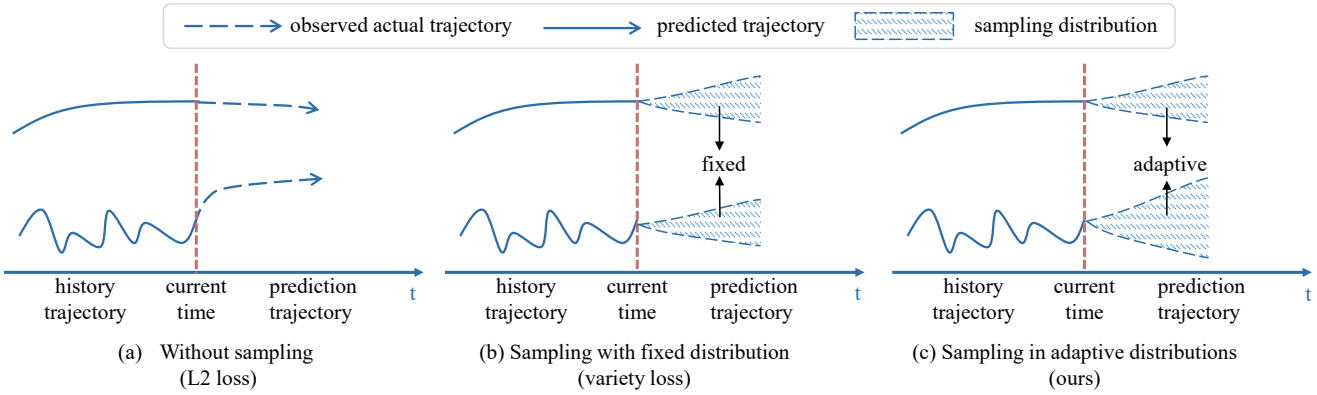


Fig. 5. Schematic comparison of different loss

that variety loss does not directly generate these trajectories. Instead, it operates with a sampling mechanism to generate a diverse set of candidate trajectories.

The proposed variety loss follows the widely adopted best-of-k formulation. k future trajectory samples are generated, and the one closest to the ground truth is selected to compute the training loss. While similar in structure, our method differs because the k hypotheses are not sampled from a fixed Gaussian distribution. Instead, the sampling distribution is adaptively shaped to reflect task-specific uncertainty, enabling the model to explore the trajectory hypothesis space under varying probabilistic spreads, as shown in Figure 5(c). This method allows the model to more effectively capture diverse futures while maintaining precision across heterogeneous scene contexts. The methods above are shown explicitly in Figure 5.

C. Adaptive variance mechanism

In multi-modal trajectory prediction, it is essential to construct a suitable sampling distribution to generate diverse future trajectories beyond using the variety loss. While prior works often adopt a fixed standard Gaussian distribution, this paper argues that prediction accuracy can reflect proximity to the ground truth: higher accuracy implies greater similarity, and vice versa. Relying on a static Gaussian variance may thus be suboptimal—particularly in dynamic datasets with evolving patterns. To address this, we propose an adaptive formulation (8) that explicitly links variance to predictive performance:

$$p\left(x_{t+1} \mid x_t\right)=\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(x_{t+1}-f\left(x_t\right)\right)^2}{2\sigma^2}\right) \tag{8}$$

where x_{t+1} represents the predicted value at time $t+1$, $f\left(x_t\right)$ represents the output of the prediction model, and σ^2 represents the predicted variance. It is evident from (8) that:

1) When the variance is small ($\sigma^2 \to 0$), the probability density function exhibits a spike, indicating that the generated trajectory closely aligns with the output of the prediction model;
2) When the variance is significant ($\sigma^2 \to \infty$), the probability density function becomes broad, covering a wide range of possible values, and suggesting that the generated trajectories are more dispersed than those with small variance.

The primary objective of the adaptive variance mechanism is to determine the optimal variance for superior performance in prediction tasks. The adaptive variance mechanism optimizes model parameters and prediction results by training under different initial variance values and subsequently integrating the results to identify the optimal variance value for the current task. The adaptive variance mechanism process is shown explicitly in Figure 6. The following paragraphs provide a detailed description of the algorithmic process of the adaptive variance mechanism.

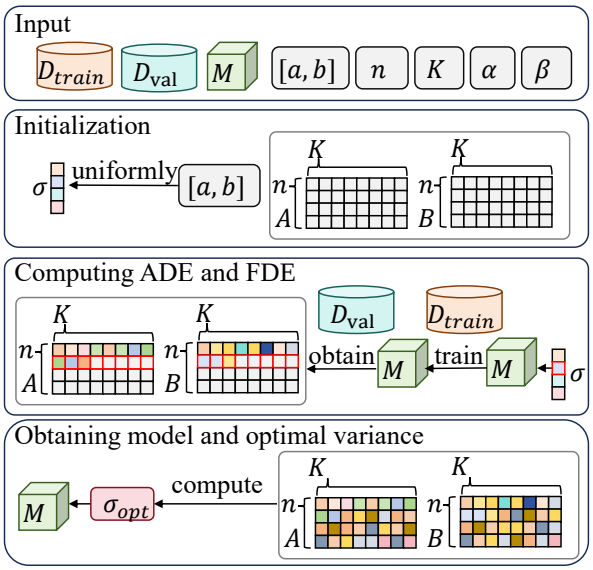


Fig. 6. The adaptive variance mechanism flowchart

1) Selecting initial variance values: choose n distinct initial variance values from the interval $[\mathrm{a}, \mathrm{b}]$, represented by $\sigma=\{\sigma_1, \sigma_2, \cdots, \sigma_n\}$, where $a \leq \sigma_i \leq b$ for $i=1,2, \cdots, n$;
2) Training models with different variance values: train the model for K epochs with different initial variance values. For each epoch, initialize the model parameters using current variance value σ_i, and perform iterative

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TCYB.2025.3613498, IEEE Transactions on Cybernetics

updates on the entire training set. During training, record the performance metrics ADE (average displacement error) and FDE (final displacement error) for the validation set at each epoch using (9) and (10) respectively:

$$\mathrm{ADE} = \frac{\sum_{i=1}^{U} \sum_{t-t_{\mathrm{obs}}+1}^{t_{\mathrm{pred}}} \left\| Y_i^t - \hat{Y}_i^t \right\|_2}{U * (t_{\mathrm{pred}} - t_{\mathrm{obs}} - 1)} \tag{9}$$

$$\mathrm{FDE} = \frac{\sum U_{i=1} \| Y_i^t \text{ pred } - \hat{Y}_i^t \text{ pred } \|_2}{U} \tag{10}$$

where Y_i and $\hat{Y}_i$ represent actual and predicted positions of pedestrian i at time t, and U denotes the total number of pedestrians.

3) Computing optimal variance value: based on the results obtained from training with n different variance values, compute the optimal variance value σ_{best}. First, compute the average ADE and FDE for each variance value:

$$\overline{\mathrm{ADE}}(\sigma_i) = \frac{1}{K} \sum_{k=1}^{K} \mathrm{ADE}(\sigma_i)_k \tag{11}$$

$$\overline{\mathrm{FDE}}(\sigma_i) = \frac{1}{K} \sum_{k=1}^{K} \mathrm{FDE}(\sigma_i)_k \tag{12}$$

where $\mathrm{ADE}(\sigma_i)_k$ and $\mathrm{FDE}(\sigma_i)_k$ represent the validation set ADE and FDE obtained during the k-th epoch using variance value σ_i. Subsequently, the optimal variance value can be determined by identifying the value that minimizes the weighted average of the ADE and FDE metrics:

$$\sigma_{\mathrm{best}} = \arg\min_{\sigma_i} \left(\alpha \overline{\mathrm{ADE}}(\sigma_i) + \beta \overline{\mathrm{FDE}}(\sigma_i) \right) \tag{13}$$

where α and β are weight parameters satisfying $\alpha, \beta \geq 0$ and $\alpha + \beta = 1$;

4) Training with optimal variance value: train the model using the optimal variance value σ_{best} until convergence. Finally, the trained model can be applied to the test set to evaluate its performance in the prediction task.

The adaptive variance mechanism dynamically adjusts sampling variance during training to improve trajectory-level consistency and forecasting performance across varying scene contexts. Task-specific configurations, such as the initial variance search space and training schedule, can modulate behavior. A detailed procedural description is provided in Algorithm 1.

Algorithm 1: Adaptive variance mechanism

Input: Training and validation sets D_{train} and D_{val}, initial variance range $[a, b]$, number of variance values n, training epochs K, and weight parameters α and β
Output: Optimal variance value σ_{opt} and model M

```
Initialize vector σ = [σ_1, σ_2, ..., σ_n] where each σ_i is uniformly selected from the range [a, b]
Initialize ADE matrix A and FDE matrix F
for i = 1 to n do
    for k = 1 to K do
        Initialize model parameters with variance value σ_i
        Train the model on the training set D_train
        Compute ADE and FDE on the validation set D_val
        Store ADE and FDE in matrices A and F, respectively: A[i, k] = computed ADE,
        F[i, k] = computed FDE by (11)-(12)
    end for
end for
Compute optimal σ_opt by minimizing the weighted loss function using computed ADE and FDE values by (13)
Train model M using optimal variance value σ_opt
```

IV. Experiments and analysis

This section evaluates the performance of the proposed method through experiments conducted on three datasets, ETH [33], UCY [25], and self-made datasets. A 5-fold cross-validation methodology was employed to validate the model's efficacy. Additionally, this paper compares the performance of the proposed method against several other deep learning techniques.

A. Experimental setup and dataset introduction

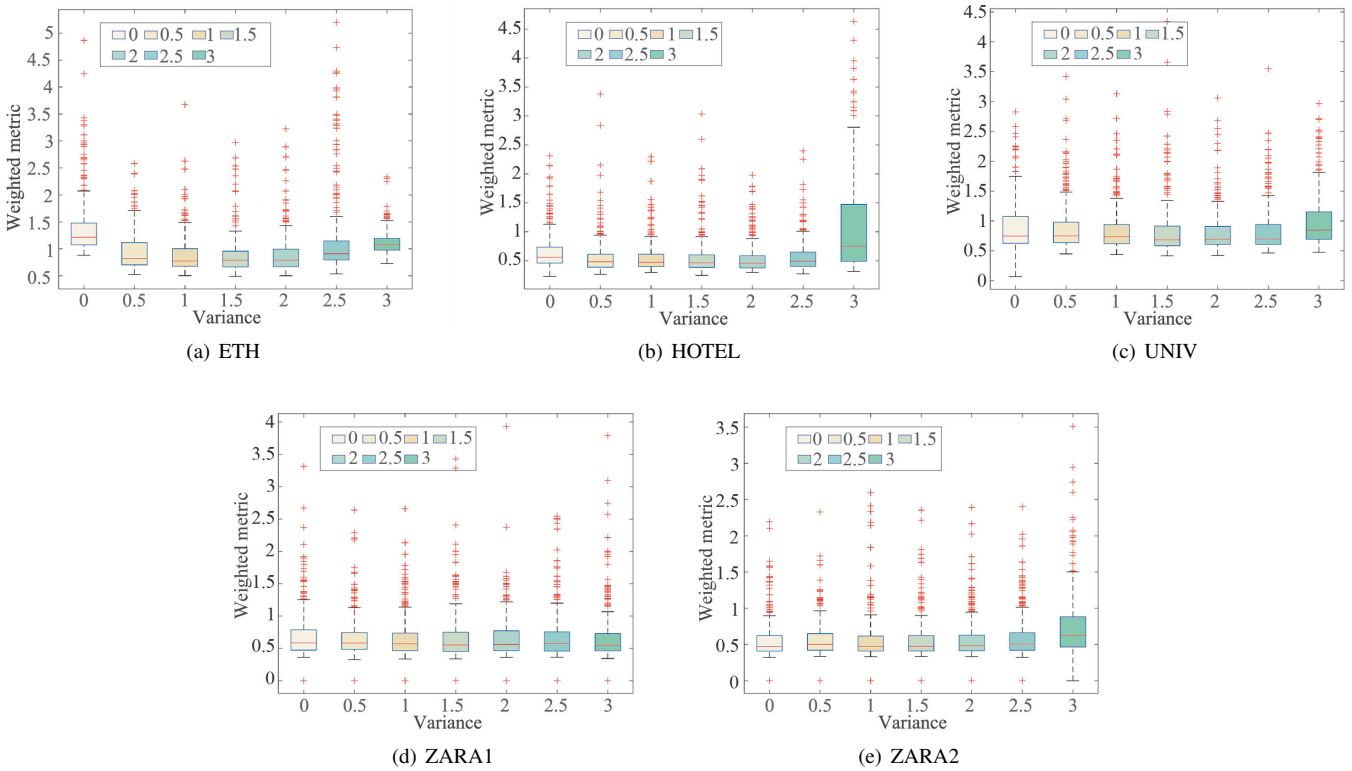
The experiments presented in this paper were conducted on a hardware environment consisting of an i7-11800H processor, an NVIDIA GeForce RTX 3080 graphics card, and Windows 11 operating system. The software environment employed for the experiments was Python 3.8. Table I presents the specific hyperparameter settings employed in the experiments.

TABLE I
Hyperparameter settings

Hyperparameter	Value
Learning rate	0.0015
Epochs	300
Batch size	4
Observation time	3.2s (8 frames)
Prediction time	4.8s (12 frames)
Sequence Length	8s (20 frames)
Sampling number	20
Optimizer	Adam

This study employs both public and self-collected datasets.

1) Public datasets: ETH and UCY, comprising five subsets (ETH, HOTEL, UNIV, ZARA01, ZARA02), are widely used pedestrian trajectory benchmarks. The ETH dataset includes 5,400 frames of 450 subjects in a busy square, with trajectories lasting up to 15 seconds and featuring conversational or stationary behaviors. The UCY dataset provides top-view pedestrian recordings in unconstrained public scenes, capturing diverse multi-person interactions with minimal environmental obstructions.
2) Self-collected datasets: Covering five real-world scenarios, Street Intersection, Community, Market, Road, and Piazza. This dataset was recorded over 40 minutes and



(a) ETH (b) HOTEL (c) UNIV

(d) ZARA1 (e) ZARA2

Fig. 7. Experimental results with different variances under different datasets

processed using YOLOv11 and DeepSORT. It includes individual and interactive pedestrian trajectories across varied environments, offering complementary realism to benchmark datasets.

B. Validity test of adaptive variance mechanism

In this section, a preliminary experimental verification is conducted using five public datasets to examine the correlation between prediction accuracy and sampling variance in trajectory prediction tasks. The findings of the experiment are presented in Figure 7, illustrating a significant gap in the model's prediction accuracy across different variances. The results from almost all datasets indicate a clear relationship between model prediction accuracy and variance.

TABLE II
MATCHED SAMPLES T-TEST RESULTS

Index	Mean diff.	Std. deviation diff.	p-value
ADE	-0.0136	0.009477	0.00141
FDE	-0.039900	0.034093	0.004914
Weighted metric	-0.030037	0.023377	0.002828

As the adaptive variance mechanism relies on sampling, further tests are required to assess its effectiveness and reduce chance influences. Both ADE and FDE metrics are combined into a weighted metric, with weights of 0.375 for ADE and 0.625 for FDE, computed by the entropy weight method as shown in (14):

$$H = -\sum_{i=1}^{m} f_i \cdot \ln f_i \tag{14}$$

where f_i represents the probability of event occurrence, and H represents the information entropy associated with that event.

This study employs a matched samples t-test to evaluate the effectiveness of the adaptive variance mechanism. We hypothesize that the enhanced model (model A) outperforms the baseline (model B). Both models are trained for the same number of total epochs to ensure fairness, as shown in (15):

$$E_{\mathrm{B}^{\mathrm{train}}} = E_{\mathrm{A}^{\mathrm{var}}} + E_{\mathrm{A}^{\mathrm{train}}} \tag{15}$$

where $E_{\mathrm{B}^{\mathrm{train}}}$, $E_{\mathrm{A}^{\mathrm{var}}}$, and $E_{\mathrm{A}^{\mathrm{train}}}$ denote the training epochs of model B, the adaptive variance phase, and the training epochs of model A, respectively. Both models were trained ten times, and a matched samples t-test was conducted on the collected data. Table II reports each metric's mean differences, standard deviations, and p-values.

As shown in Table II, model A consistently achieves lower ADE, FDE, and weighted errors than model B, with smaller standard deviations and p-values below 0.05. These results confirm the statistical significance and effectiveness of the adaptive variance mechanism. Additional experiments were conducted to assess the trade-off between time and performance of the adaptive variance mechanism. With a fixed number of variance candidates n, training epochs K were

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TCYB.2025.3613498, IEEE Transactions on Cybernetics

TABLE III
RELATIONSHIP BETWEEN ADAPTIVE VARIANCE MECHANISM TIME CONSUMPTION AND MODEL PERFORMANCE

The number of K	ETH	HOTEL	UNIV	ZARA01	ZARA02	Average	Weighted metric
0 epochs	0.37/0.64	0.21/0.47	0.33/0.65	0.24/0.54	0.21/0.48	0.27/0.56	0.45
5 epochs	0.37/0.65	0.20/0.40	0.33/0.66	0.26/0.55	**0.20**/0.48	0.27/0.56	0.45
10 epochs	0.35/0.62	0.19/0.37	0.31/0.63	**0.23/0.52**	**0.20**/0.46	0.26/0.53	0.43
15 epochs	0.34/0.63	0.18/0.33	0.30/0.63	0.25/0.53	0.21/**0.45**	0.26/**0.51**	0.42
20 epochs	**0.33/0.61**	**0.17/0.31**	**0.29/0.62**	0.25/0.53	0.21/0.46	**0.25/0.51**	**0.41**

TABLE IV
THE COMPARATIVE EXPERIMENTAL RESULTS

Method (ADE/FDE)	ETH	HOTEL	UNIV	ZARA1	ZARA2	Average	Weighted metric
SGAN [14]	0.87/1.62	0.67/1.37	0.60/1.26	0.34/0.69	0.42/0.84	0.58/1.12	0.92
SoPhie [38]	0.7/1.43	0.76/1.67	0.54/1.24	0.30/0.63	0.38/0.78	0.54/1.15	0.92
Social BiGAT [22]	0.69/1.29	0.49/1.01	0.55/1.32	0.30/0.62	0.36/0.75	0.48/1.00	0.81
Social STGCNN [31]	0.64/1.11	0.49/0.85	0.44/0.79	0.34/0.53	0.30/0.48	0.44/0.75	0.63
RSGB [39]	0.80/1.53	0.33/0.64	0.59/1.25	0.40/0.86	0.30/0.65	0.48/0.99	0.80
Informer [74]	0.61/1.21	0.25/0.55	0.38/0.79	0.29/0.65	0.25/0.56	0.36/0.75	0.60
SGCN [44]	0.63/1.03	0.32/0.55	0.37/0.70	0.29/0.53	0.25/0.45	0.37/0.65	0.55
BR-GAN [34]	0.73/1.37	0.55/1.13	0.53/1.07	0.35/0.71	0.35/0.72	0.50/1.00	0.81
SEEM [55]	0.62/1.20	0.61/1.21	0.50/1.04	0.31/0.61	0.36/0.68	0.48/0.95	0.77
IA-LSTM [65]	0.43/0.77	0.50/0.80	0.48/0.73	0.44/0.46	0.36/0.55	0.44/0.66	0.58
TPPO [66]	0.75/1.27	0.36/0.70	0.39/0.74	0.22/0.37	0.23/0.45	0.39/0.71	0.59
Social NSTransformers [21]	0.40/0.71	0.29/0.47	0.39/0.73	0.34/0.62	0.31/0.57	0.35/0.62	0.52
TP-EGTv1 [69]	0.55/0.82	0.18/**0.30**	0.33/**0.60**	**0.20/0.32**	**0.21/0.35**	0.29/**0.48**	**0.41**
Social informer	**0.34/0.61**	**0.17**/0.33	**0.30**/0.63	0.24/0.52	**0.21**/0.44	**0.25**/0.51	**0.41**

varied, excluding the adaptive variance phase from the total training count, as shown in (16):

$$E_{\text{total}} = E_{\text{var}} + E_{\text{train}} \tag{16}$$

where E_{total} represents the total epochs, and E_{var} and E_{train} represent the epochs of the adaptive variance mechanism and the final training, respectively. The final experimental results are shown in Table III.

Table III investigates how varying the number of training epochs (K) per variance value in the adaptive variance mechanism affects model performance. Results show a general improvement across datasets as K increases: performance is lowest at $K = 0$, improves notably at $K = 5$ and $K = 10$, but plateaus by $K = 20$, suggesting diminishing returns. The adaptive variance mechanism consistently enhances model performance under different training schedules.

C. Comparative Experimental Results and Discussion

Table IV presents ADE/FDE results on five standard datasets. Results are presented in Table IV, with bold for the best and underlined for the suboptimal results. Social Informer achieves the best overall performance, with the lowest average ADE (0.25) and FDE (0.51), and ties for the best weighted metric (0.41).

Social informer ranks first in ETH, HOTEL, and UNIV, and remains competitive in ZARA1 and ZARA2. These results indicate the model's strong generalization ability across dense and sparse scenes. Compared to prior methods, our model consistently reduces displacement errors. The improvements demonstrate the effectiveness of our attention-based interaction encoding and diverse trajectory modeling.

D. The ablation experiment results and analysis

Table V illustrates the enhancement in prediction performance achieved by including the interaction module, variety loss, and adaptive variance mechanism. Notably, the combined application of multiple components further improves the model's performance. The proposed social informer model, incorporating all components, achieves the highest prediction performance. These findings emphasize the significance of interaction, diversity, and adaptive variance in improving model performance in crowd behavior prediction.

E. Experimental results and analysis of long-distance pedestrian trajectory prediction

Further experiments were conducted to validate the social informer's effectiveness in long-distance pedestrian trajectory prediction. With the observation time set at 8 frames, the prediction time was extended to evaluate performance over longer durations. To ensure consistency of the experimental results, the long-distance pedestrian trajectory prediction experiments were conducted on five sub-datasets: ETH, HOTEL, UNIV, ZARA01, and ZARA02. Table VI shows the specific experimental results.

The results in Table VI confirm that the social informer maintains competitive performance as the prediction distance increases. At 12 frames, the model achieves a strongly weighted metric of 0.41, demonstrating effective short-term trajectory prediction. Even at 24 frames, it performs robustly with a metric of 0.61. While performance degradation is expected due to long-term prediction uncertainty, it remains within an acceptable range for long-distance modeling. The increasing ADE and FDE values across datasets highlight the growing difficulty of predicting over longer time horizons.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TCYB.2025.3613498, IEEE Transactions on Cybernetics

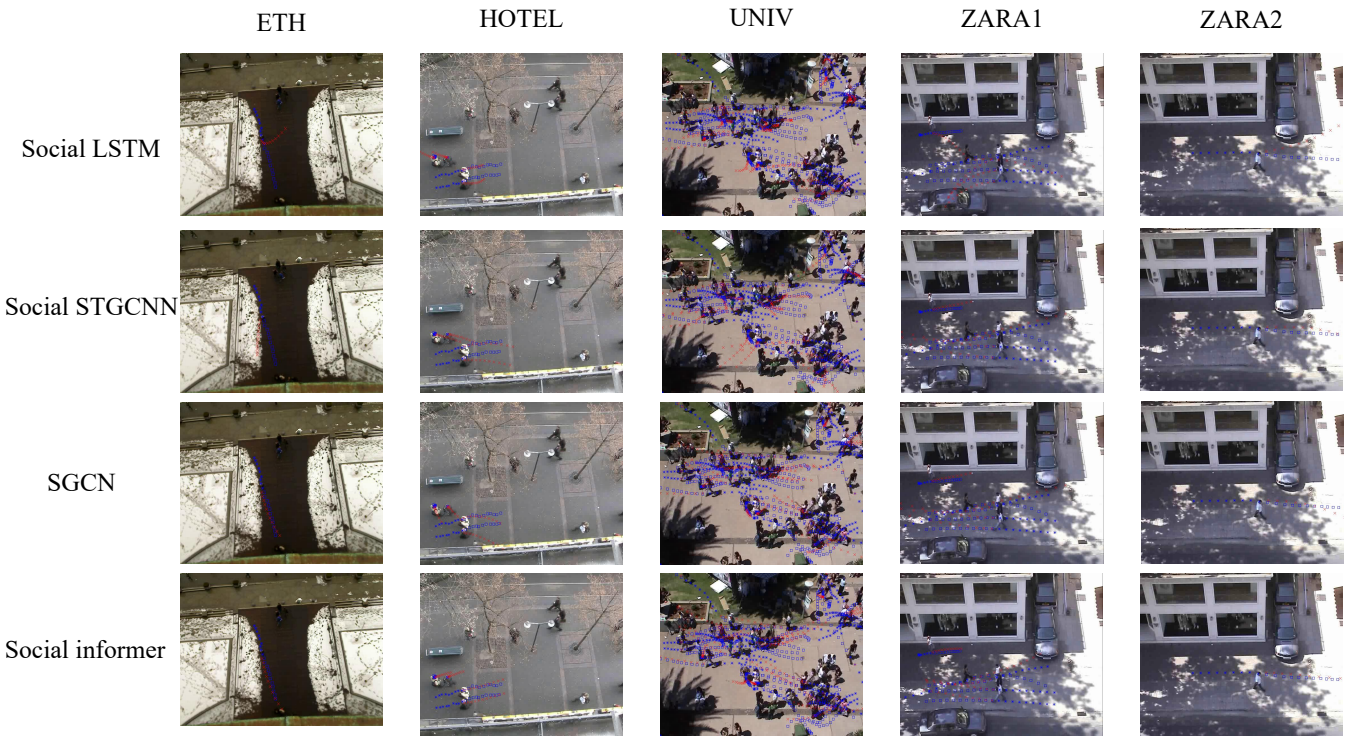


Fig. 8. The comparative experimental visualization results: blue asterisks represent observed values, blue boxes represent labeled values, and red boxes represent predicted values.

TABLE V
THE RESULTS OF ABLATION EXPERIMENTS

Method (ADE/FDE)	ETH	HOTEL	UNIV	ZARA01	ZARA02	Average	Weighted metric
Informer [74]	0.61/1.21	0.25/0.55	0.38/0.79	0.29/0.65	0.25/0.56	0.36/0.75	0.60
Informer+Interaction module	0.39/0.78	0.24/0.53	0.37/0.76	0.27/0.62	0.23/0.52	0.30/0.64	0.51
Informer+Variety loss	0.42/0.76	0.22/0.50	0.36/0.76	0.27/0.55	0.23/0.49	0.30/0.61	0.49
Informer+Adaptive variance mechanism	0.38/0.68	0.22/0.49	0.34/0.70	0.26/0.55	0.23/0.48	0.29/0.58	0.47
Informer+Interaction module+Variety loss	0.37/ 0.64	0.21/0.47	0.33/0.65	**0.24**/0.54	**0.21**/0.48	0.27/0.56	0.45
Social informer	**0.34/0.61**	**0.17/0.33**	**0.30/0.63**	**0.24/0.52**	**0.21/0.44**	**0.25/0.51**	**0.41**

TABLE VI
EXPERIMENTAL RESULTS OF LONG-DISTANCE PEDESTRIAN TRAJECTORY PREDICTION

Prediction distance (ADE/FDE)	ETH	HOTEL	UNIV	ZARA01	ZARA02	Average	Weighted metric
12 frames	**0.34/0.61**	**0.17/0.33**	**0.30/0.63**	**0.24/0.52**	**0.21/0.44**	**0.25/0.51**	**0.41**
15 frames	0.38/0.70	0.20/0.39	0.32/0.66	0.25/0.53	0.24/0.51	0.28/0.56	0.46
18 frames	0.39/0.73	0.23/0.46	0.35/0.72	0.26/0.56	0.27/0.60	0.30/0.61	0.49
21 frames	0.41/0.76	0.27/0.52	0.38/0.87	0.27/0.61	0.28/0.68	0.32/0.69	0.55
24 frames	0.44/0.80	0.33/0.59	0.43/0.94	0.31/0.65	0.32/0.77	0.37/0.75	0.61

TABLE VII
EXPERIMENTAL RESULTS OF REAL-WORLD ENVIRONMENTS

Model weight sources (ADE/FDE)	Street intersection	Community	Market	Roads	Piazza	Average	Weighted metric
ETH	0.053/0.142	0.043/0.099	0.041/0.090	0.037/0.076	0.060/0.151	0.047/0.112	0.088
HOTEL	0.062/0.181	0.045/0.135	0.047/0.142	0.046/0.149	0.077/0.244	0.055/0.170	0.127
UNIV	0.051/0.159	0.035/0.097	0.041/0.121	0.035/0.094	0.081/0.210	0.049/0.136	0.103
ZARA01	0.180/0.595	0.158/0.479	0.154/0.474	0.154/0.466	0.164/0.489	0.162/0.501	0.374
ZARA02	0.236/0.654	0.190/0.500	0.182/0.488	0.169/0.435	0.153/0.450	0.186/0.505	0.386
Complete training	**0.012/0.019**	**0.010/0.026**	**0.021/0.059**	**0.015/0.036**	**0.031/0.107**	**0.018/0.049**	**0.037**

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TCYB.2025.3613498, IEEE Transactions on Cybernetics

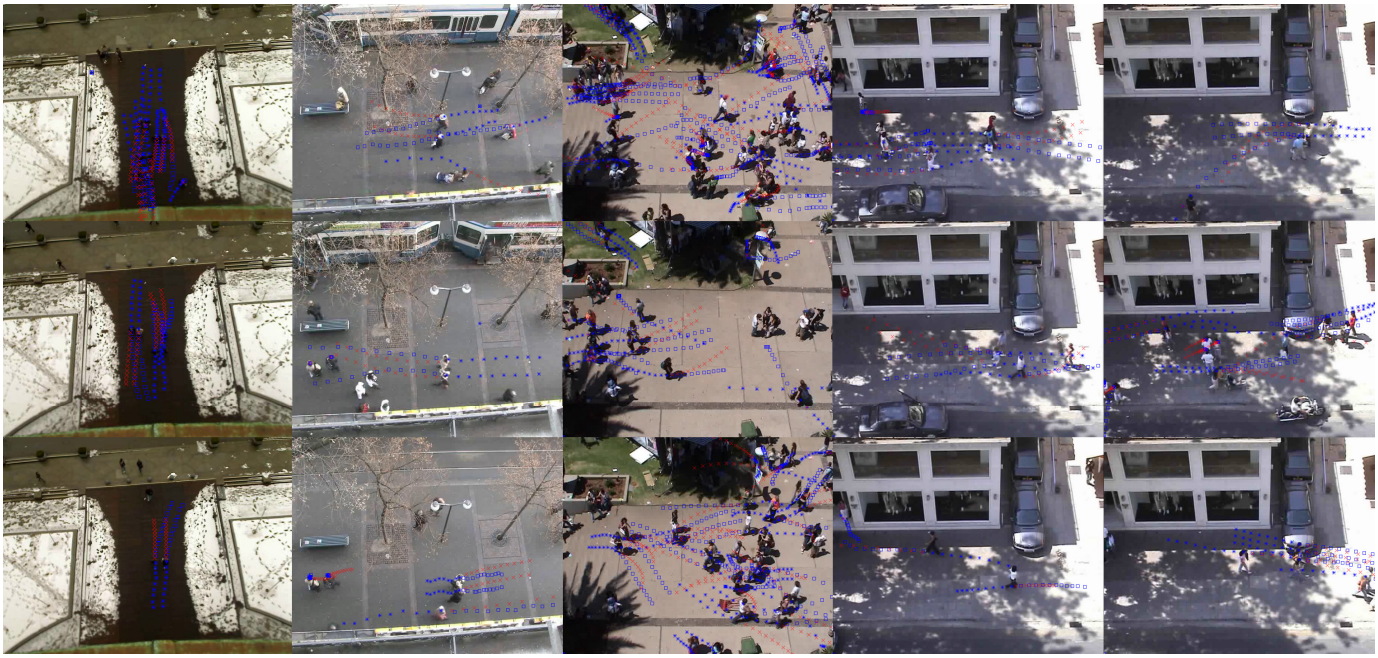


Fig. 9. The social informer model experimental effect: blue asterisks represent observed values, blue boxes represent labeled values, and red boxes represent predicted values.

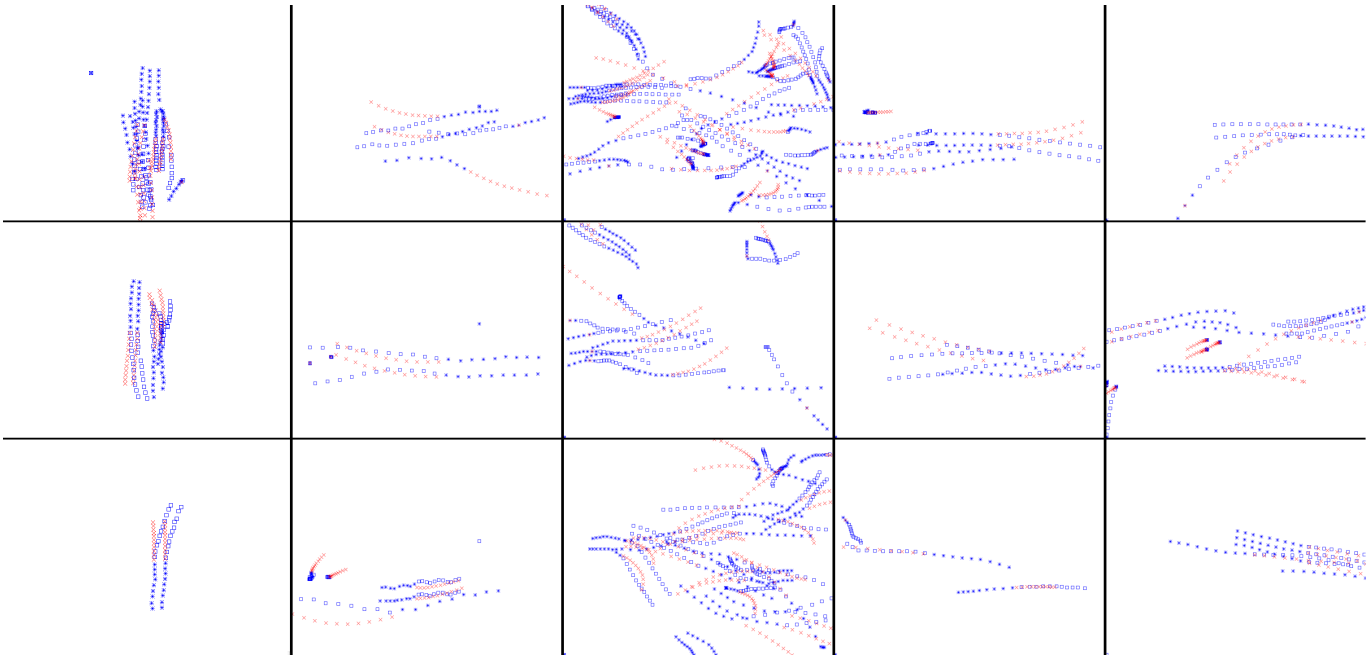


Fig. 10. The experimental effect of social informer modeling without backgrounds: blue asterisks represent observed values, the blue boxes are labeled values, and the red boxes represent predicted values.

Nevertheless, the social informer maintains its ability to capture social interactions, proving its adaptability in long-range predictions.

F. Comparison of experimental results with different distribution assumptions

Table VIII illustrates the comparative performance of the proposed adaptive variance mechanism against three common distribution assumptions—uniform, Poisson ($\lambda = 1$), and standard Gaussian. Across all five datasets, the adaptive variance mechanism achieves the best results in ADE and FDE metrics, consistent improvements over the standard Gaussian baseline (e.g., 0.25/0.51 vs. 0.28/0.56 on average). While the standard Gaussian distribution already performs better than uniform and Poisson assumptions, its fixed nature limits adaptability to diverse scenes and pedestrian behaviors. In contrast, the adaptive variance mechanism dynamically tunes the sampling distribution, yielding more realistic and diverse predictions.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TCYB.2025.3613498, IEEE Transactions on Cybernetics

Fig. 11. The results of real environment experiment visualization: the blue asterisks represent observed values, the blue boxes are labeled values, and the red boxes represent predicted values.

Importantly, this gain in prediction accuracy does not incur additional computational cost—the inference time remains nearly identical (0.32s), affirming the method's efficiency and scalability. These results demonstrate the proposed mechanism's effectiveness and innovation in accuracy and runtime.

G. Experimental qualitative analysis

In Figure 9, blue asterisks represent observed values, blue boxes are labeled values, and red boxes show predicted values. The analysis of Figure 9 and Table IV shows that the social informer model predicts pedestrian trajectories well when the trajectories are steady, indicating its ability to capture underlying patterns. However, accuracy drops significantly when trajectories change abruptly or when multiple pedestrians interact. These challenges arise from the difficulty of predicting sudden behavioral changes and modeling complex pedestrian interactions. Despite these limitations, the social informer outperformed other models regarding similarity to actual values and precision in predicting future trajectories, making it suitable for most pedestrian movement prediction scenarios. Additionally, Figure 10 shows the experimental results without a background map, where all trajectories correspond to those in Figure 9.

H. Experimental results and analysis of real-world environments

This paper collects real-world datasets to test the social informer model's generalization ability. The experimental approach involves training the model on public datasets and then using the trained weights to predict pedestrian trajectories in real scenarios that the model has not seen during training. The results of real-world experiments are shown in Table VII.

As shown in Table VII, the results highlight the social informer model's excellent generalization ability in pedestrian trajectory prediction. After training on public datasets, the model performed well on self-collected datasets, demonstrating strong generalization. The complete training model achieved the lowest error, validating its robustness and potential for few-shot and zero-shot learning.

Additionally, this paper conducted further visualization experiments to demonstrate the model's prediction capabilities in real-world environments. As shown in the Figure 11, the predicted trajectories align closely with the ground truth, even in complex settings, further highlighting the model's effectiveness in capturing dynamic social interactions and pedestrian behaviors.

I. Experimental results and analysis of model performance indicator evaluation

Table IX compares trajectory prediction models regarding FLOPS, parameters, and inference time. While early models like Social LSTM [1] are less efficient in handling complex dynamics, newer ones like STGAT [19] and STGCNN [31] reduce computational load but compromise on speed. Models like MemoNet [62], STAR [71], and GE-Trans [32] improve accuracy at the cost of higher complexity. In contrast, Social Informer achieves a good balance—offering low FLOPS (984.12 Mb), minimal parameters (0.56 Mb), and fast infer-

TABLE VIII
EXPERIMENTAL RESULTS COMPARING DIFFERENT DISTRIBUTION ASSUMPTIONS

Method (ADE/FDE)	ETH	HOTEL	UNIV	ZARA01	ZARA02	Average	Inference time
Social informer with the Uniform distribution	0.41/0.74	0.25/0.52	0.39/0.70	0.31/0.60	0.29/0.53	0.33/0.62	0.31s
Social informer with the Poisson distribution (λ=1)	0.39/0.70	0.25/0.49	0.37/0.68	0.28/0.58	0.26/0.49	0.31/0.59	0.33s
Social informer with the standard Gaussian distribution	0.38/0.66	0.22/0.47	0.33/0.64	0.25/0.54	0.23/0.47	0.28/0.56	0.32s
Social informer with the adaptive variance mechanism	**0.34/0.61**	**0.17/0.33**	**0.30/0.63**	**0.24/0.52**	**0.21/0.44**	**0.25/0.51**	0.32s

TABLE IX
MODELS PERFORMANCE COMPARISON RESULTS

Method	FLOPS	Model parameters	Inference time
Social LSTM [1]	26.50 Mb	0.26 Mb	0.04 s
STGAT [19]	6.47 Mb	0.0047 Mb	0.01 s
STGCNN [31]	0.19 Mb	0.01 Mb	2.05 s
MemoNet [62]	2868.03 Mb	5.18 Mb	3.51 s
STAR [71]	11933.95 Mb	0.80 Mb	0.45 s
GE-Trans [32]	26861.96 Mb	44.15 Mb	0.05 s
Social informer	984.12 Mb	0.56 Mb	0.32 s

ence (0.32s)—making it suitable for real-time deployment in complex scenes.

V. CONCLUSION

This paper proposes social informer, a novel multi-modal pedestrian trajectory prediction model that addresses two core challenges: learning social interactions and modeling behavioral uncertainty. The model dynamically adapts to diverse data distributions by leveraging an adaptive variance mechanism, enabling accurate and flexible prediction without relying on predefined interaction priors. Future extensions could integrate computer vision to enhance interaction modeling, and the framework holds promise for broader applications such as robotics, autonomous driving, and surveillance, underscoring its versatility.

REFERENCES

[1] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, F. Li and S. Savarese, "Social LSTM: Human Trajectory Prediction in Crowded Spaces," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, 2016, pp. 961-971.

[2] A. Belhadi, Y. Djenouri, D. Djenouri, T. Michalak and J. C. -W. Lin, "Deep Learning Versus Traditional Solutions for Group Trajectory Outliers," *IEEE Transactions on Cybernetics*, vol. 52, no. 6, pp. 4508-4519, 2022.

[3] A. Chetverikov, and J. Jehee, "Motion direction is represented as a bimodal probability distribution in the human visual cortex," *Nature Communications*, vol. 14, no. 1, pp. 7634, 2023.

[4] X. Chen, R. Yang, Y. Xue, M. Huang, R. Ferrero and Z. Wang, "Deep Transfer Learning for Bearing Fault Diagnosis: A Systematic Review Since 2016," *IEEE Transactions on Instrumentation and Measurement*, vol. 72, pp. 1-21, 2023.

[5] Y. Chen, R. Yang, M. Huang, Z. Wang, and X. Liu, "Single-Source to Single-Target Cross-Subject Motor Imagery Classification Based on Multisubdomain Adaptation Network," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 1992-2002, 2022.

[6] H. Cheng, Z. Wang, Z. Wei, L. Ma and X. Liu, "On Adaptive Learning Framework for Deep Weighted Sparse Autoencoder: A Multiobjective Evolutionary Algorithm," *IEEE Transactions on Cybernetics*, vol. 52, no. 5, pp. 3221-3231, 2022.

[7] S. Dai, J. Liu and N. -M. Cheung, "Uncertainty-Aware Pedestrian Crossing Prediction via Reinforcement Learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 34, no. 10, pp. 9540-9549, 2024.

[8] F. Deng, Y. Ming and B. Lyu, "CCE-Net: causal convolution embedding network for streaming automatic speech recognition," *International Journal of Network Dynamics and Intelligence*, vol. 3, no. 3, art. no. 100019, 2024.

[9] N. Deo, A. Rangesh and M. M. Trivedi, "How Would Surround Vehicles Move? A Unified Framework for Maneuver Classification and Motion Prediction," *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 2, pp. 129-140, 2018.

[10] A. Dong, A. Starr and Y. Zhao, "Neural network-based parametric system identification: a review," *International Journal of Systems Science*, vol. 54, no. 13, pp. 2676–2688, 2023.

[11] Helbing D, Molnar P. "Social force model for pedestrian dynamics", *Physical review E*, vol. 51, no. 5, pp. 4282, 1995.

[12] T. Fernando, S. Denman, S. Sridharan and C. Fookes, "Neighbourhood Context Embeddings in Deep Inverse Reinforcement Learning for Predicting Pedestrian Motion Over Long Time Horizons," *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, Seoul, Korea (South), 2019, pp. 1179-1187.

[13] W. Fang, B. Shen, A. Pan, L. Zou and B. Song, "A cooperative stochastic configuration network based on differential evolutionary sparrow search algorithm for prediction," *Systems Science & Control Engineering*, vol. 12, no. 1, art. no. 2314481, 2024.

[14] A. Gupta, J. Johnson, F. Li, S. Savarese and A. Alahi, "Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, USA, 2018, pp. 2255-2264.

[15] Y. Guan et al., "Integrated Decision and Control: Toward Interpretable and Computationally Efficient Driving Intelligence," *IEEE Transactions on Cybernetics*, vol. 53, no. 2, pp. 859-873, 2023.

[16] D. Huang, W. -A. Zhang, F. Guo, W. Liu and X. Shi, "Wavelet Packet Decomposition-Based Multiscale CNN for Fault Diagnosis of Wind Turbine Gearbox," *IEEE Transactions on Cybernetics*, vol. 53, no. 1, pp. 443-453, 2023.

[17] M. Huang, Z. Jiang and K. Ozbay, "Learning-Based Adaptive Optimal Control for Connected Vehicles in Mixed Traffic: Robustness to Driver Reaction Time," *IEEE Transactions on Cybernetics*, vol. 52, no. 6, pp. 5267-5277, 2022.

[18] Y. Huang, J. Du, Z. Yang, Z. Zhou, L. Zhang, and H. Chen, "A Survey on Trajectory-Prediction Methods for Autonomous Driving," *IEEE Transactions on Intelligent Vehicles*, vol. 7, no. 3, pp. 652-674, 2022.

[19] Y. Huang, H. Bi, Z. Li, T. Mao and Z. Wang, "STGAT: Modeling Spatial-Temporal Interactions for Human Trajectory Prediction," *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, Seoul, Korea (South), 2019, pp. 6271-6280.

[20] Z. Jiang, C. Qin, R. Yang, B. Shi, F. E. Alsaadi and Z. Wang, "Social Entropy Informer: A Multi-Scale Model-Data Dual-Driven Approach for Pedestrian Trajectory Prediction," *IEEE Transactions on Intelligent Transportation Systems*, early access.

[21] Z. Jiang et al., "Social NSTransformers: Low-Quality Pedestrian Trajectory Prediction," *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 11, pp. 5575-5588, 2024.

[22] V. Kosaraju, A. Sadeghian, R. Martín-Martín, I. Reid, H. Rezatofighi, and S. Savarese, "Social-BiGAT: Multimodal Trajectory Forecasting using Bicycle-GAN and Graph Attention Networks," *2019 Advances in Neural Information Processing Systems (NIPS)*, Vancouver, Canada, 2019, vol.32.

[23] F. Kessler, J. Frankenstein, and CA. Rothkopf. "Human navigation strategies and their errors result from dynamic interactions of spatial uncertainties," *Nature Communications*, vol. 15, no. 1, pp. 5677, 2024.

[24] H. Lapid, S. Shushan, A. Plotkin, H. Voet, Y. Roth, T. Hummel, E. Schneidman, and N. Sobel, "Neural activity at the human olfactory epithelium reflects olfactory perception," *Nature neuroscience*, vol. 14, no. 11, pp. 1455-1461, 2011.

[25] A. Lerner, Y. Chrysanthou, and D. Lischinski, "Crowds by Example." *Computer Graphics Forum*, vol. 26, no. 3, pp. 655-664, 2007.

[26] L. Li, M. Pagnucco and Y. Song, "Graph-based Spatial Transformer with Memory Replay for Multi-future Pedestrian Trajectory Prediction," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, 2022, pp. 2221-2231.

[27] N. Lee, W. Choi, P. Vernaza, C. B. Choy, P. H. S. Torr and M. Chandraker, "DESIRE: Distant Future Prediction in Dynamic Scenes with Interacting Agents," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, 2017, pp. 2165-2174.

[28] R. Liang, Y. Li, J. Zhou and X. Li, "STGlow: A Flow-Based Generative Framework With Dual-Graphormer for Pedestrian Trajectory Prediction," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 11, pp. 16504-16517, 2024.

[29] L. Li, B. Zhou, J. Lian, X. Wang and Y. Zhou, "Multi-PPTP: Multiple Probabilistic Pedestrian Trajectory Prediction in the Complex Junction Scene," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 8, pp. 13758-13768, 2022.

[30] Y. Li, X.-Y. Lu, J. Wang, and K. Li, "Pedestrian Trajectory Prediction Combining Probabilistic Reasoning and Sequence Learning," *IEEE Transactions on Intelligent Vehicles*, vol. 5, no. 3, pp. 461-474, 2020.

[31] A. Mohamed, K. Qian, M. Elhoseiny and C. Claudel, "Social-STGCNN: A Social Spatio-Temporal Graph Convolutional Neural Network for Hu-

man Trajectory Prediction," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2020, pp. 14412-14420.

[32] J. Ma, C. Yang, S. Mao, J. Zhang, S. C. Periaswamy and J. Patton, "Human Trajectory Completion with Transformers," *ICC 2022 - IEEE International Conference on Communications*, Seoul, Korea, Republic of, 2022, pp. 3346-3351.

[33] S. Pellegrini, A. Ess, and L. Van Gool, "Improving data association by joint modeling of pedestrian trajectories and groupings," *Computer Vision–ECCV 2010: 11th European Conference on Computer Vision (ECCV)*, Heraklion, Crete, Greece, 2010, pp.452-465.

[34] S. M. Pang, J. X. Cao, M. Y. Jian, J. Lai, and Z. Y. Yan, "BR-GAN: A Pedestrian Trajectory Prediction Model Combined With Behavior Recognition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 12, pp. 24609-24620, 2022.

[35] R. Quintero Mínguez, I. Parra Alonso, D. Fernández-Llorca, and M. Á. Sotelo, "Pedestrian Path, Pose, and Intention Prediction Through Gaussian Process Dynamical Models and Pedestrian Activity Recognition," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 5, pp. 1803-1814, 2019.

[36] S. Qiao, D. Shen, X. Wang, N. Han, and W. Zhu, "A Self-Adaptive Parameter Selection Trajectory Prediction Approach via Hidden Markov Models," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 1, pp. 284-296, 2015.

[37] R. Ran, L. -J. Deng, T. -X. Jiang, J. -F. Hu, J. Chanussot and G. Vivone, "GuidedNet: A General CNN Fusion Framework via High-Resolution Guidance for Hyperspectral Image Super-Resolution," *IEEE Transactions on Cybernetics*, vol. 53, no. 7, pp. 4148-4161, 2023.

[38] A. Sadeghian, V. Kosaraju, A. Sadeghian, N. Hirose, H. Rezatofighi and S. Savarese, "SoPhie: An Attentive GAN for Predicting Paths Compliant to Social and Physical Constraints," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, 2019, pp. 1349-1358.

[39] J. Sun, Q. Jiang and C. Lu, "Recursive Social Behavior Graph for Trajectory Prediction," *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2020, pp. 657-666.

[40] J. Sun, Z. Wang, H. Yu, S. Zhang, J. Dong and P. Gao, "Two-Stage Deep Regression Enhanced Depth Estimation From a Single RGB Image," *IEEE Transactions on Emerging Topics in Computing*, vol. 10, no. 2, pp. 719-727, 2022

[41] SANDERS, S. "125 questions: Exploration and Discovery," *Science, AAAS Custom Publishing Office: Washington, DC, USA*, 2021.

[42] Y. Sun, B. Xue, M. Zhang, G. G. Yen and J. Lv, "Automatically Designing CNN Architectures Using the Genetic Algorithm for Image Classification," *IEEE Transactions on Cybernetics*, vol. 50, no. 9, pp. 3840-3854, 2020.

[43] K. Saleh, M. Hossny and S. Nahavandi, "Contextual Recurrent Predictive Model for Long-Term Intent Prediction of Vulnerable Road Users," *IEEE Transactions on Intelligent Transportation Systems*, vol. 21, no. 8, pp. 3398-3408, 2020.

[44] L. Shi et al., "SGCN:Sparse Graph Convolution Network for Pedestrian Trajectory Prediction," 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 2021, pp. 8990-8999.

[45] X. Song, K. Chen, X. Li, J. Sun, B. Hou, Y. Cui, B. Zhang, G. Xiong, and Z. Wang, "Pedestrian Trajectory Prediction Based on Deep Convolutional LSTM Network," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 6, pp. 3285-3302, 2021.

[46] Z. Su, C. Wang, H. Cui, N. Djuric, C. Vallespi-Gonzalez and D. Bradley, "Temporally-Continuous Probabilistic Prediction using Polynomial Trajectory Parameterization," *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Prague, Czech Republic, 2021, pp. 3837-3843.

[47] H. Tang, P. Wei, J. Li, and N. Zheng, "EvoSTGAT: Evolving Spatiotemporal Graph Attention Networks for Pedestrian Trajectory Prediction," *Neurocomputing*, vol. 491, pp. 333-342, 2022.

[48] N. Uhlemann, F. Fent and M. Lienkamp, "Evaluating Pedestrian Trajectory Prediction Methods With Respect to Autonomous Driving," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 10, pp. 13937-13946, 2024.

[49] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is All You Need," *Advances in Neural Information Processing Systems (NIPS)*, Long Beach, CA, USA, vol. 30, 2017.

[50] F. Wen, M. Li and R. Wang, "Social Transformer: A Pedestrian Trajectory Prediction Method based on Social Feature Processing Using Transformer," *2022 International Joint Conference on Neural Networks (IJCNN)*, Padua, Italy, 2022, pp. 1-7.

[51] H. Wang, R. Fan, Y. Sun and M. Liu, "Dynamic Fusion Module Evolves Drivable Area and Road Anomaly Detection: A Benchmark and Algorithms," *IEEE Transactions on Cybernetics*, vol. 52, no. 10, pp. 10750-10760, 2022.

[52] Y. Wang, Y. Niu, W. Zhu, W. Chen, Q. Li and T. Wang, "Predicting Pedestrian Crossing Behavior at Unsignalized Mid-Block Crosswalks Using Maximum Entropy Deep Inverse Reinforcement Learning," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 5, pp. 3685-3698, 2024.

[53] Y. Wang, C. Wen and X. Wu, "Fault detection and isolation of floating wind turbine pitch system based on Kalman filter and multi-attention 1DCNN," *Systems Science & Control Engineering*, vol. 12, no. 1, art. no. 2362169, 2024.

[54] Y. Wang, C. Shen, J. Huang and H. Chen, "Model-free adaptive control for unmanned surface vessels: a literature review," *Systems Science & Control Engineering*, vol. 12, no. 1, art. no. 2316170, 2024.

[55] D. Wang, H. Liu, N. Wang, Y. Wang, H. Wang and S. McLoone, "SEEM: A Sequence Entropy Energy-Based Model for Pedestrian Trajectory All-Then-One Prediction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 1070-1086, 2023.

[56] H. Wang, Z. Liu, Z. Han, Y. Wu and D. Liu, "Rapid Adaptation for Active Pantograph Control in High-Speed Railway via Deep Meta Reinforcement Learning," *IEEE Transactions on Cybernetics*, vol. 54, no. 5, pp. 2811-2823, 2024.

[57] Y. Wu, X. Huang, Z. Tian, X. Yan and H. Yu, "Emotion contagion model for dynamical crowd path planning," *International Journal of Network Dynamics and Intelligence*, vol. 3, no. 3, art. no. 100014, 2024.

[58] Y. Wang, Z. Liu, Z. Zuo, Z. Li, L. Wang and X. Luo, "Trajectory Planning and Safety Assessment of Autonomous Vehicles Based on Motion Prediction and Model Predictive Control," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 9, pp. 8546-8556, 2019.

[59] W. Wang and M. Wang, "Adaptive neural event-triggered output-feedback optimal tracking control for discrete-time pure-feedback nonlinear systems," *International Journal of Network Dynamics and Intelligence*, vol. 3, no. 2, art. no. 100010, 2024.

[60] C. Wong, B. Xia, Z. Zou, Y. Wang and X. You, "SocialCircle: Learning the Angle-based Social Interaction Representation for Pedestrian Trajectory Prediction," *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2024, pp. 19005-19015.

[61] Z. Xiao et al., "Understanding Private Car Aggregation Effect via Spatio-Temporal Analysis of Trajectory Data," *IEEE Transactions on Cybernetics*, vol. 53, no. 4, pp. 2346-2357, 2023.

[62] C. Xu, W. Mao, W. Zhang and S. Chen, "Remember Intentions: Retrospective-Memory-based Trajectory Prediction," *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, LA, USA, 2022, pp. 6478-6487.

[63] J. Xue and B. Shen, "A survey on sparrow search algorithms and their applications," *International Journal of Systems Science*, vol. 55, no. 4, pp. 814–832, 2024.

[64] B Yang, J Yang, R Ni, C Yang, and X Liu, "Multi-granularity scenarios understanding network for trajectory prediction," *Complex & Intelligent Systems*, vol. 9, no. 1, pp. 851-864, 2023.

[65] J. Yang, Y. Chen, S. Du, B. Chen and J. C. Principe, "IA-LSTM: Interaction-Aware LSTM for Pedestrian Trajectory Prediction," *IEEE Transactions on Cybernetics*, vol. 54, no. 7, pp. 3904-3917, 2024.

[66] B. Yang, C. He, P. Wang, C. Chan, X. Liu and Y. Chen, "TPPO: A Novel Trajectory Predictor With Pseudo Oracle," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 54, no. 5, pp. 2846-2859, 2024.

[67] B. Yang, K. Yan, C. Hu, H. Hu, Z. Yu and R. Ni, "Dynamic Subclass-Balancing Contrastive Learning for Long-Tail Pedestrian Trajectory Prediction With Progressive Refinement," *IEEE Transactions on Automation Science and Engineering*, vol. 22, pp. 8645-8658, 2025.

[68] J. Yin, F. Yan, Y. Liu, G. He and Y. Zhuang, "An overview of simultaneous localisation and mapping: towards multi-sensor fusion," *International Journal of Systems Science*, vol. 55, no. 3, pp. 550–568, 2024.

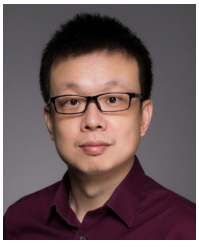
[69] B. Yang, F. Fan, R. Ni, H. Wang, A. Jafaripournimchahi and H. Hu, "A Multi-Task Learning Network With a Collision-Aware Graph Transformer for Traffic-Agents Trajectory Prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 7, pp. 6677-6690, 2024.

[70] B. Yang, J. Zhu, Z. Yu, F. Fan, X. Liu and R. Ni, "Fast Adaptation Trajectory Prediction Method Based on Online Multisource Transfer Learning," *IEEE Transactions on Automation Science and Engineering*, vol. 22, pp. 1289-1304, 2025.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.
Citation information: DOI: 10.1109/TCYB.2025.3613498, IEEE Transactions on Cybernetics

[71] C. Yu, X. Ma, J. Ren, H. Zhao, and S. Yi, "Spatio-temporal graph transformer networks for pedestrian trajectory prediction," *Computer Vision–ECCV 2020: 16th European Conference*, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII 16.
[72] C. Yang and Z. Pei, "Long-Short Term Spatio-Temporal Aggregation for Trajectory Prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 4, pp. 4114-4126, 2023.
[73] B Yang, Y Lu, R Wan, H Hu, C Yang, R Ni, "Meta-IRLSOT++: A meta-inverse reinforcement learning method for fast adaptation of trajectory prediction networks," *Expert Systems with Applications*, vol. 240, pp. 122499, 2024.
[74] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, and W. Zhang, "Informer: Beyond Efficient Transformer for Long Sequence Time-series Forecasting," *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, 2021, vol. 35, no. 12, pp. 11106-11115.
[75] K. Zhang, X. Feng, L. Wu, and Z. He, "Trajectory Prediction for Autonomous Driving Using Spatial-Temporal Graph Attention Transformer," *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, no. 11, pp. 22343-22353, 2022.
[76] N. Zeng, P. Wu, Z. Wang, H. Li, W. Liu, and X. Liu, "A Small-Sized Object Detection Oriented Multi-Scale Feature Fusion Approach with Application to Defect Detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1-14, 2022.
[77] X. Zhang, H. Chen, W. Yang, W. Jin and W. Zhu, "Pedestrian Path Prediction for Autonomous Driving at Un-Signalized Crosswalk Using W/CDM and MSFM," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 5, pp. 3025-3037, 2021.
[78] Y. Zhang et al., "Adaptive Safe Reinforcement Learning With Full-State Constraints and Constrained Adaptation for Autonomous Vehicles," *IEEE Transactions on Cybernetics*, vol. 54, no. 3, pp. 1907-1920, 2024.
[79] K. Zhu, Z. Wang, D. Ding, J. Hu and H. Dong, "Cloud-Based Collision Avoidance Adaptive Cruise Control for Autonomous Vehicles Under External Disturbances With Token Bucket Shapers," *IEEE Transactions on Industrial Informatics*, early access.
[80] N. Zeng, Z. Wang, B. Zineddin, Y. Li, M. Du, L. Xiao, X. Liu and T. Young "Image-Based Quantitative Analysis of Gold Immunochromatographic Strip via Cellular Neural Network Approach," *IEEE Transactions on Medical Imaging*, vol. 33, no. 5, pp. 1129-1136, 2014.
[81] L. Zhang, B. Wang, Y. Zhao, Y. Yuan, T. Zhou and Z. Li, "Collaborative Multimodal Fusion Network for Multiagent Perception," *IEEE Transactions on Cybernetics*, early access.
[82] S. Zhang and W. Che, "Observer-Based Self-Triggered Resilient Control for Multiagent Systems: A k-Connected Graph Construction Approach," *IEEE Transactions on Cybernetics*, vol. 54, no. 2, pp. 706-716, 2024.

Zihan Jiang (Student Member, IEEE) received the B.Eng. degree from the Northeast Forestry University, Harbin, China, in 2022 and the M.Sc. degree from the University of Liverpool, UK, in 2024. He is pursuing a Ph.D. degree at Tongji University, Shanghai, China. He has served as a reviewer for several journals including IEEE Transactions on Intelligent Transportation Systems, IEEE Transactions on Mobile Computing, and IEEE Internet of Things Journal. His research interests include trajectory prediction, robot and computer vision.



Rui Yang (Senior Member, IEEE) received the B.Eng. degree in Computer Engineering and the Ph.D. degree in Electrical and Computer Engineering from National University of Singapore in 2008 and 2013 respectively.

He is currently an Associate Professor in the School of Advanced Technology, Xi'an Jiaotong-Liverpool University, Suzhou, China, and an Honorary Lecturer in the Department of Computer Science, University of Liverpool, Liverpool, United Kingdom. His research interests include machine learning based data analysis and applications. He is the author or co-author of several technical papers and also a very active reviewer for many international journals and conferences. Dr. Yang is currently serving as Associate Editors for IEEE Transactions on Instrumentation and Measurement, Neurocomputing, and Cognitive Computation.

Yiqun Ma received the B.Eng. degree in computer science and technology from Nanjing University of Information Science and Technology in 2021. She is currently pursuing the M.Sc. degree at the University of Liverpool, UK. Her research interests include computer vision and semi-supervised learning.

Chengxuan Qin received the MRes degree in Pattern Recognition and Intelligent Systems from the University of Liverpool in 2023. He is pursuing a Ph.D. degree at the University of Liverpool, China. His research interests include brain-computer interfaces, temporal signal analysis, and machine learning.

Xiaohan Chen received the M.Eng. degree from Beijing University of Chemical Technology, Beijing, China, in 2020 and the Ph.D. degree from University of Liverpool, Liverpool, United Kingdom, in 2024. His research interests include transfer learning, few-shot learning, self-supervised learning and applications in rotating machinery fault diagnosis.

Zidong Wang (Fellow, IEEE) received the B.Sc. degree in mathematics from Suzhou University, Suzhou, China, in 1986, and the M.Sc. degree in applied mathematics and the Ph.D. degree in electrical engineering from the Nanjing University of Science and Technology, Nanjing, China, in 1990 and 1994, respectively.

From 1990 to 2002, he held teaching and research appointments in universities in China, Germany, and the U.K. He is currently a Professor of dynamical systems and computing with the Department of Computer Science, Brunel University London, Uxbridge, U.K. He has published a number of papers in international journals. His research interests include dynamical systems, signal processing, bioinformatics, control theory, and applications.

Prof. Wang is a member of the Academia Europaea and the European Academy of Sciences and Arts, an Academician of the International Academy for Systems and Cybernetic Sciences, a fellow of the Royal Statistical Society, and a member of the Program Committee of many international conferences. He holds the Alexander von Humboldt Research Fellowship of Germany, the JSPS Research Fellowship of Japan, and the William Mong Visiting Research Fellowship of Hong Kong. He serves (or has served) as the Editor-in-Chief for International Journal of Systems Science, Neurocomputing, and Systems Science and Control Engineering; and an Associate Editor for 12 international journals, including IEEE Transactions on Automatic Control, IEEE Transactions on Control Systems Technology, IEEE Transactions on Neural Networks and Learning Systems, IEEE Transactions on Signal Processing, and IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Reviews.