SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

July 2021, Vol. 64 172207:1–172207:13
https://doi.org/10.1007/s11432-020-3071-8

Trajectory prediction of cyclist based on dynamic Bayesian network and long short-term memory model at unsignalized intersections

Hongbo GAO[1,2], Hang SU[3], Yingfeng CAI[4], Renfei WU[5*], Zhengyuan HAO[1], Yongneng XU[6], Wei WU[6], Jianqing WANG[7], Zhijun LI[1] & Zhen KAN[1]

[1]*Department of Automation, University of Science and Technology of China, Hefei* 230026, *China;*
[2]*Institute of Advanced Technology, University of Science and Technology of China, Hefei* 230088, *China;*
[3]*The Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Milano* 20133, *Italy;*
[4]*Automotive Engineering Research Institute, Jiangsu University, Zhenjiang* 212013, *China;*
[5]*College of Traffic and Transportation, Southeast University, Nanjing* 210018, *China;*
[6]*School of Automation, Nanjing University of Science and Technology, Nanjing* 210094, *China;*
[7]*School of Vehicle and Mobility, Tsinghua University, Beijing* 100084, *China*

Received 18 April 2020/Revised 22 June 2020/Accepted 1 September 2020/Published online 18 May 2021

Abstract Cyclist trajectory prediction is of great significance for both active collision avoidance and path planning of intelligent vehicles. This paper presents a trajectory prediction method for the motion intention of cyclists in real traffic scenarios. This method is based on dynamic Bayesian network (DBN) and long short-term memory (LSTM). The motion intention of cyclists is hard to predict owing to potential large uncertainties. The DBN is used to infer the distribution of cyclists' intentions at intersections to improve the prediction time. The LSTM with encoder-decoder is used to predict the cyclists' trajectories to improve the accuracy of prediction. Therefore, the DBN and LSTM are adopted to guarantee prediction accuracy and improve the prediction time. The experiment results are presented to show the effectiveness of the predict strategies.

Keywords trajectory prediction, dynamic Bayesian network (DBN), long short-term memory (LSTM), unsignalized intersections, motion intention

Citation Gao H B, Su H, Cai Y F, et al. Trajectory prediction of cyclist based on dynamic Bayesian network and long short-term memory model at unsignalized intersections. Sci China Inf Sci, 2021, 64(7): 172207, https://doi.org/10.1007/s11432-020-3071-8

1 Introduction

Cyclists are the vulnerable group in road traffic accidents. The "Annual Statistical Report of 2018 for Road Traffic Accidents in China" shows that cyclists accounted for 21.98% of the total number of traffic fatalities with 46074 cyclists, and 20.07% of the injured with 12800 cyclists in 2017 [1]. Especially, medium-sized cities and campuses generally have higher levels of bicycle commuting, so the traffic accident rate is higher. According to situation awareness theory [2], it is not possible for a vehicle to arrive at its destination safely by merely judging the current state of the environment. Instead, it has to cope with the predictions of cyclists in the surrounding traffic environment [3–5].

Cyclist trajectory prediction aims to predict the position of the cyclist based on the model of cyclist motion. In the beginning, some modeled cyclist motion treats the cyclist as a rigid body, including a first power motion model and a quadratic motion model according to the motion complexity. The first power motion model includes the constant position model, constant velocity model, constant position model, and so on. The first power motion model assumes linear motion of the cyclist in a plane with two perpendicular velocities, where the velocity directions are invariant. Quadratic motion models are nonlinear models including the constant turn rate and velocity model, and the constant turn rate and

* Corresponding author (email: wurenfeidora@163.com)

acceleration model. Quadratic motion models are often used to model aircraft and automobiles. However, such models do not perform well in this work owing to potential large errors.

Based on the typical motion model, Kalman filter (KF) was used to track and predict the motion of the cyclist [6]. In order to improve the prediction accuracy of a single model, multi-motion models, such as the mixture of linear dynamical system and switching linear dynamical system, are often employed. Goldhammer et al. [7] and Gao et al. [8] combined multi-layer perception (MLP) and polynomial fitting to predict the future trajectory of the rider. Bieshaar et al. [9] used MLP to detect the starting intention of cyclists at intersections based on smart devices and infrastructure. In addition, long short-term memory (LSTM) was increasingly used as a time recurrent neural network for trajectory prediction [10,11]. Saleh et al. [12] compared one-way and two-way LSTM in the trajectory prediction of cyclists. The results showed that the prediction accuracy of both networks is higher than that of traditional MLP.

However, the above methods are not applicable to scenarios when cyclists changing directions because they treat the cyclist as a rigid body and less consider the intention in cyclists' mind. Pool et al. [13] modeled the rider's motion as a regular or a hybrid linear dynamic system [14,15]. The hybrid method used road intersection as features to distinguish the cyclist's motion intention such as going straight or making turns. The accuracy of the predicted right turn was increased by 20% when the forecast duration was 1.0 s compared with the traditional constant velocity model. Refs. [9,16,17] extracted the rider's pose characteristics from history cyclist images and employed SVM and CNN to classify the rider's intentions. The system was used to predict the cyclist's intention of waiting or moving at intersections with traffic lights. The test results show that both methods perform well to decide whether the rider is ready to move, but the method of CNN requires a shorter processing time. Refs. [14,18–20] used a dynamic Bayesian network (DBN) to determine the motion intention. According to the characteristics of the rider's arm movement and the relative position between the rider, vehicles and road structure of the intersection, the probabilities of cyclist's motions such as going straight or making turns can be well predicted. Compared with linear dynamic systems and hybrid linear dynamic systems, DBN can take into account more cyclist-related features and effectively improve recognition accuracy.

This paper mainly studies an integrated method combining the intention recognition and trajectory prediction of the cyclist in an unsignalized intersection scene. The design algorithm realizes the intent-based cyclist trajectory prediction, and uses the test vehicle platform to collect real road data. Furthermore, the proposed method verification and comparison with other methods are conducted.

The main contributions of this paper are summarized as follows.

(1) The DBN is used to infer the distribution of cyclists' intentions at intersections. Therefore, the problem of short prediction time can be solved.

(2) The LSTM with encoder-decoder is used to predict the cyclists' trajectories. Therefore, the prediction inaccuracy problem can be solved.

(3) The DBN and LSTM are adopted to guarantee prediction accuracy and improve the prediction time.

The rest of this paper is structured as follows. Section 2 provides a method overview of the framework. Section 3 describes the process of intention inference. Section 4 introduces trajectory prediction. Section 5 provides experiments and analysis about intention inference and trajectory prediction. Section 6 concludes the article.

2 Method overview of framework

In this paper, the system consists of four parts, and the system overview is shown in Figure 1.

The first block is an input module. Information of the cyclist, ego vehicle, and road environments are collected from sensors equipped on the ego vehicle. A data-collecting scenario is set at an unsignalized intersection on campus, where cyclists pass frequently and cyclist-vehicle accidents are likely to occur.

The second block is the intention inference. A cyclist's motion features, ego vehicle movement and environmental features are incorporated in a graphical model of DBN, and the intention of the cyclist can be estimated through the probabilistic reasoning. Details of this part are shown in Section 4.

The third block is trajectory prediction. LSTM with encoder-decoder is selected because of its strong capabilities of acquiring information and representing intrinsic features when dealing with time sequences. The trajectory prediction model incorporates historical movement and road limit. The intention result from the second block is also incorporated. The model details are shown in Section 5.

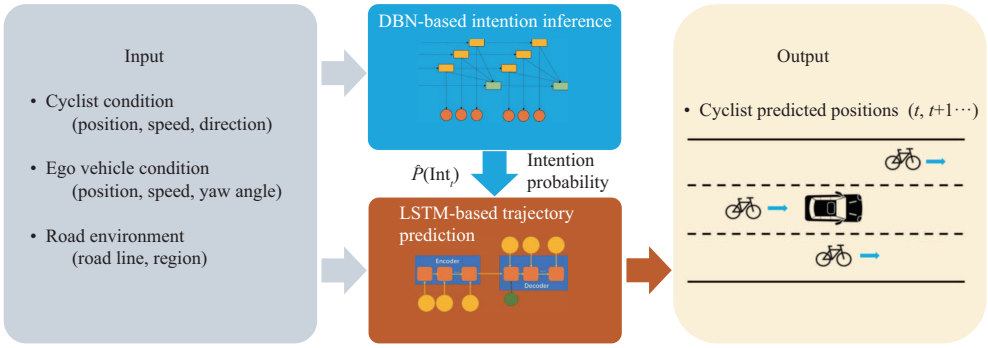

Figure 1 (Color online) Block diagram of the proposed method.

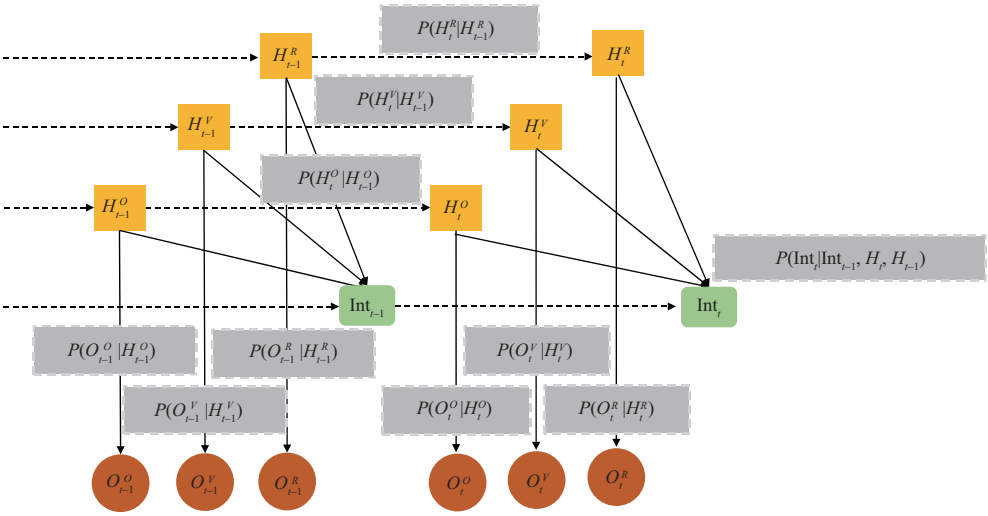

Figure 2 (Color online) The DBN-based model for cyclist intention inference from time step $t-1$ to t. The observed variables (circle) correspond to the hidden variables. Intentions (shaded rectangle) are inferred via given hidden variables.

In the fourth piece, the predicted positions of cyclists are implemented for verification and application.

3 Intention inference

A DBN is employed to combine qualitative intentions with additional context information. The intent-related position, motion, and feature details are regarded as hidden variables and observables, which are variables collected from the vehicle sensors.

As shown in Figure 2, the hidden variables are selected as cyclist intention node Int, which has three possible intentions: going straight, turning right and turning left. Node O is the set of observables from the view of vehicle sensors, and node H is the set of hidden variables in the view of cyclists [20, 21].

$$O = \{O^O, O^V, O^R\}, \tag{1}$$

$$H = \{H^O, H^V, H^R\}, \tag{2}$$

where O^O is the cyclist velocity orientation, O^V is the time of collision between the cyclist and the ego vehicle, and O^R is the relative position of cyclist in the road. In the set of hidden variables H, H^O denotes the orientation of cyclist's temporary target, which is related to O^O; H^V denotes whether the cyclist feels any danger when riding, which is related to O^V; H^R denotes whether the cyclist is ready to go straight, turn right or turn left; H^O, H^V and H^R are discrete variables; and O^O, O^V, O^R are continuous variables. H^O, H^V and H^R are independent variables with temporal and probabilistic transitions. Therefore, the temporal transition for hidden variables H is defined as

$$P(H_t|H_{t-1}) = P(H_t^O|H_{t-1}^O) \times P(H_t^V|H_{t-1}^V) \times P(H_t^R|H_{t-1}^R), \tag{3}$$

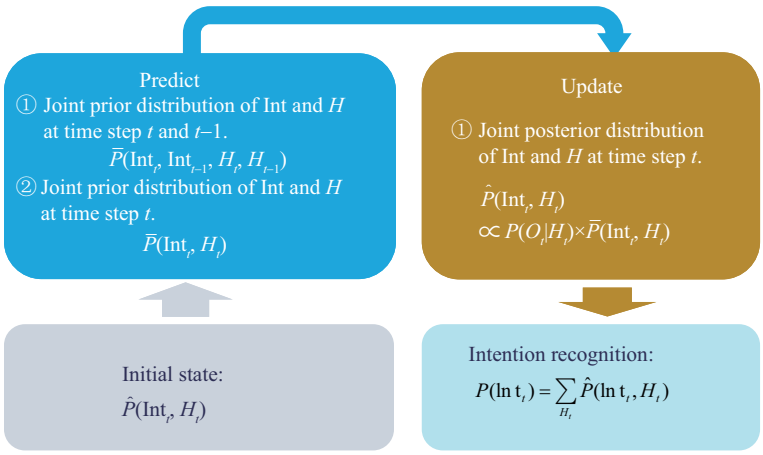

Figure 3 (Color online) Inferring process steps of DBN.

where $P(H_t^O|H_{t-1}^O)$, $P(H_t^V|H_{t-1}^V)$ and $P(H_t^R|H_{t-1}^R)$ are the temporal transitions for hidden variables H from time step $t-1$ to t.

The state of cyclist motion orientation is not directly related to the observed cyclist's speed and the cyclist's coordinates on the road map. The probability relationships $O^O - H^O$, $O^V - H^V$ and $O^R - H^R$ are independent of each other. Therefore, the relationship between the observed variables O and the hidden variables H^V could be written as follows:

$$P(O_t|H_t) = P(O^O|H^O) \times P(O^V|H^V) \times P(O^R|H^R), \tag{4}$$

where $P(O^O|H^O)$, $P(O^V|H^V)$ and $P(O^R|H^R)$ indicate the relationships between $O^O - H^O$, $O^V - H^V$ and $O^R - H^R$ respectively in the same time step.

$P(O^O|H^O)$ accords with a Gaussian distribution with a mean value μ_0 and a standard deviation σ_0; $P(O^V|H^V)$ conforms to a Gamma distribution with a parameter λ_V; and $P(O^R|H^R)$ accords with a polynomial distribution. Their empirical distributions and parameters are estimated using maximum likelihood estimation on the training data, which are manually annotated with the context states.

The inference DBN process employs assumed density filtering as the approximate inference method [20] and is divided into the "predict" and "update" steps as shown in Figure 3. Prior probability $\overline{P}(\cdot)$ and posterior probability $\hat{P}(\cdot)$ are then introduced in the inference process, which uses the posterior probability of the previous step $\hat{P}(\mathrm{Int}_{t-1}, H_{t-1})$ as the basis for the "predict" step. The prior probabilities $\overline{P}(\mathrm{Int}_t, \mathrm{Int}_{t-1}, H_t, H_{t-1})$ and $\overline{P}(\mathrm{Int}_t, H_t)$ are calculated, and then the posterior probability is updated to $\hat{P}(\mathrm{Int}_t, H_t)$ with the observables.

The "predict" step is

$$\overline{P}(\mathrm{Int}_t, \mathrm{Int}_{t-1}, H_t, H_{t-1}) = P(\mathrm{Int}_t|\mathrm{Int}_{t-1}, H_t) \times P(H_t|H_{t-1}) \times \hat{P}(\mathrm{Int}_{t-1}, H_{t-1}), \tag{5}$$

$$\bar{P}(\mathrm{Int}_t, H_t) = \sum_{\mathrm{Int}_{t-1}} \sum_{H_{t-1}} \bar{P}(\mathrm{Int}_t, \mathrm{Int}_{t-1}, H_t, H_{t-1}). \tag{6}$$

The "update" step is

$$\hat{P}(\mathrm{Int}_t, H_t) \propto P(O_t|H_t) \times \overline{P}(\mathrm{Int}_t, H_t), \tag{7}$$

$$\hat{P}(\mathrm{Int}_t) = \sum_{H_t} \hat{P}(\mathrm{Int}_t, H_t). \tag{8}$$

At every time step, the network is updated with all available observables, so the posterior probability of crossing function $\hat{P}(\mathrm{Int}_t)$ is inferred. The whole intention influence process is shown in Algorithm 1.

4 Trajectory prediction

LSTM is employed for trajectory prediction based on the intention result. LSTM is a special recurrent neural network (RNN) and capable of learning long-term dependencies. It shows strong information

Algorithm 1 Intention inference model

Initialize: $P(H_t^O|H_{t-1}^O), P(H_t^V|H_{t-1}^V), P(H_t^R|H_{t-1}^R), P(O^O|H^O), P(O^V|H^V), P(O^R|H^R), P(\mathrm{Int}_t|\mathrm{Int}_{t-1}, H_t)$;
for episode $= 1,\ M$ **do**
  Initialize $\hat{P}(\mathrm{Int}_{t-1}, H_{t-1})$;
  Obtain the temporal transitions for hidden variables with time step: $P(H_t|H_{t-1})$;
  Obtain the relationship between the observed variables and the hidden variables: $P(O_t|H_t)$;
  Obtain joint prior distribution of Int and H at time step t and $t-1$: $\overline{P}(\mathrm{Int}_t, \mathrm{Int}_{t-1}, H_t, H_{t-1})$;
  Obtain joint prior distribution of Int and H at time step t : $\overline{P}(\mathrm{Int}_t, H)$;
  Obtain and output the posterior probability of crossing function $\hat{P}(\mathrm{Int}_t)$;
end for

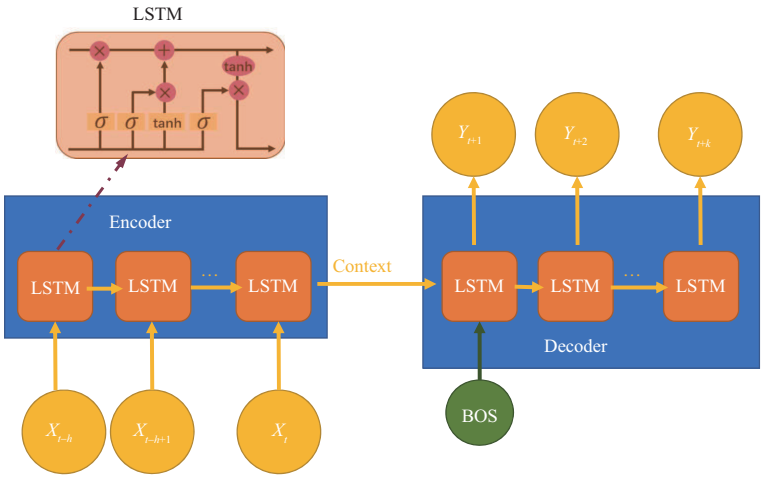

Figure 4 (Color online) Framework of trajectory prediction.

acquisition capabilities and can represent intrinsic features when dealing with time sequences. Also, LSTM can reliably predict trajectories in complicated scenes because of its ability of combining all feature into one semantic vector. Since trajectory prediction belongs to a time sequence problem, LSTM is selected to predict trajectories in future through historical motion data and environment information.

Encoder-decoder is chosen as our trajectory prediction model, since it can handle cases of unrelated outputs at different time steps and deal with seq-to-seq problems. The encoder-decoder structure is shown in Figure 4.

4.1 Encoder

The information of target cyclist's motion and road structure are put into the encoder, and then the input sequence is transformed into intermediate semantic representations — context by non-linear transformation. The context is the summary of hidden layer information.

(1) The system input is

$$X = \{X_{t-h}, X_{t-h+1}, \ldots, X_t\}, \tag{9}$$

$$X_t = \{C_t, R_t, \mathrm{Int}_t\}, \tag{10}$$

where C_t is motion information of target cyclist and $C_t = [x_t, y_t, v_t, a_t]$; R_t is information of road structure, Int_t is the result of intention inference. (x_t, y_t) is the position of the target cyclist, v_t is velocity and a_t presents acceleration.

(2) Output. The output is the context which is a semantic vector and is a summary of hidden layer information.

4.2 Decoder

The output of the encoder is the input of this step. Using the intention inference result, the predicted position in the next time step is obtained. A new position is achieved by inputting the result of the previous step iteratively. And we can get the predicted sequence of the trajectory. Here the mean squared error (MSE) is used as the loss function and mean absolute error (MAE) is defined as the evaluation index.

(1) Input. The input is the context which is a semantic vector and is a summary of hidden layer information. BOS (begin-of-study) is set as a starting condition and represents the starting position of predicted sequence. Intention with $[w_1, w_2, w_3]$ is the output from DBN-based intention recognition model, where, w_1, w_2 and w_3 represent the probabilities of going straight, turning left and turning right respectively.

(2) Loss function MSE is set as follows:

$$\mathrm{MSE} = \frac{\sum_{i=1}^{n}(Y_i - Y_{ip})^2}{n}, \tag{11}$$

where n is the number of input steps. Y_i and Y_{ip} represent the real and the predicted positions respectively.

(3) Evaluation index MAE is set as follows:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n}(|Y_i - Y_{ip}|)}{n}. \tag{12}$$

(4) Output:

$$Y = \{Y_{t+1}, Y_{t+2}, \ldots, Y_{t+k}\}. \tag{13}$$

$$Y_{t+1} = [x_{t+1}, y_{t+1}], \tag{14}$$

where the output Y is the predicted position and $[x_t, y_t]$ represents the coordinates of that. k is the number of time steps about prediction. In order to avoid information bottleneck and overfitting phenomenon, the number of hidden cells of LSTM cell body is 128, and the deep cyclic neural network structure is constructed by stacking three cyclic layers. The dropout ratio between different layers is 0.2. Using Adam optimizer, the learning rate is $\alpha = 0.0005$, and the attenuation rate is 0.9. The whole trajectory prediction process is shown in Algorithm 2.

Algorithm 2 Trajectory prediction model

Initialize: Encoder network $Q_e(X_t, w)$, and decoder network $Q_d(C_t, \mathrm{BOS}, v)$ (w and v are the network parameters);
for episode $=$ 1, M **do**
  Initialize encoder input X_t;
  Initialize starting condition BOS;
  for $t = h, H$ **do**
    Obtain the output of encoder C_t;
    Obtain the output of decoder Y_{t+1};
    Minimize the loss function to update networks: $\mathrm{MSE} = \frac{\sum_{i=1}^{n}(Y_i - Y_{ip})^2}{n}$;
  end for
end for

5 Experiments and analysis

5.1 Data collecting

Since the sampling frequency is relatively low in existing public benchmark dataset on cyclist detection [22], a new dataset is collected in this work for intent-based trajectory prediction. The dataset consists of 94 sequences, 7358 frames collected by a moving vehicle equipped with LiDAR (Velodyne VLP-16, 10 fps) and a mono camera (IDS UI-5250CP-C-HQ, 15 fps). The sensor data are fused with 10 fps. Each sequence involves one cyclist approaching the intersection with an intention to cross, for a duration of several seconds (min/max/mean: 3.7 s/12.2 s/7.8 s).

The experiment is conducted at an unsignalized intersection. The road has two lanes with mixed traffic, where vehicles and cyclists can share the same lane. The road has a speed limit of 40 km/h.

The positional ground truth (GT) of cyclists was obtained by manual labeling from the plan view of LiDAR point cloud data. To acquire information about the road geometry ahead, an intelligent vehicle can localize itself in digital map. In the confined road area for data collection, a 2D static map of the road geometry is built offline by LiDAR SLAM [23] on collected LiDAR data. The coordinate system on the map is aligned with the intersection layout. The scenarios for data collection are shown in Figure 5.

Based on the directions of relative movement between the cyclist and the ego vehicle, three classes are distinguished: longitudinal, crossing and oncoming [24]. The dataset is also divided into sub-scenarios

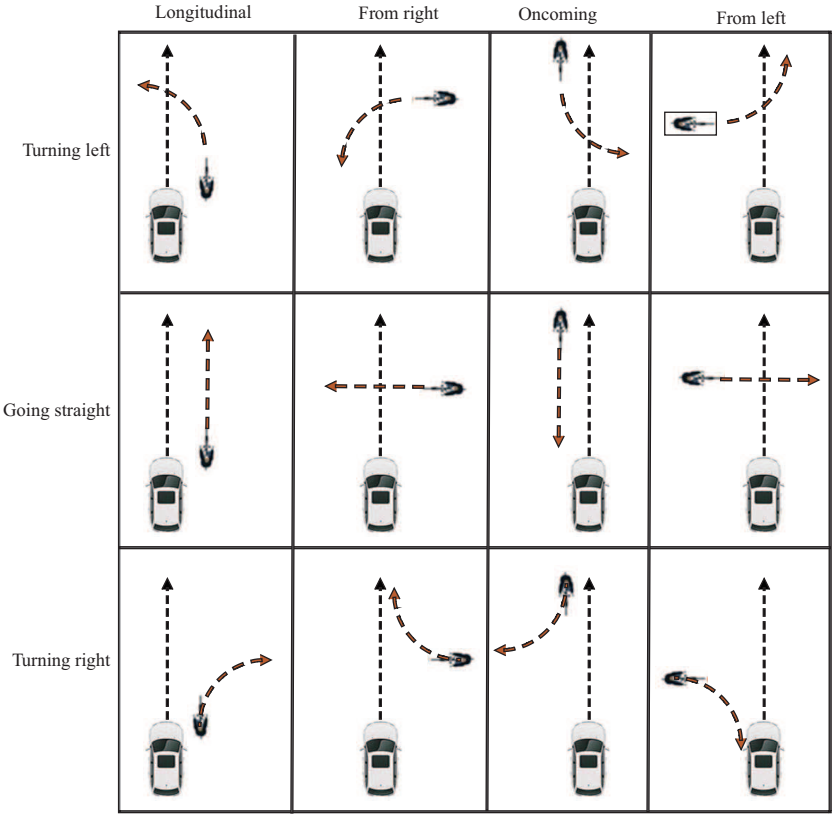

Figure 5 (Color online) Scenarios for data collection.

according to the high-level behaviors of cyclists at cross-shaped intersection as listed in Table 1. Three sub-scenarios are "going straight", "turning left" and "turning right", which correspond to the cyclist intentions when approaching the intersection.

Recorded frames are further labeled with time-to-event (TTE) values following the literature [25]. For all cyclists, the first frame where they passed the entrance line of the intersection is labeled with TTE = 0. All frames before TTE = 0 have negative TTE values, and all frames after the event have positive TTE values. The intersection region is determined by four entrance lines. If the cyclist passes the entrance line into the intersection region, the frames are labeled as "at intersection".

The proposed approach is further verified by the cyclist dataset. In the experiments, cross validation is used to separate training and test sequences, covering three sub-scenarios (i.e. "going straight", "turning left" and "turning right").

5.2 Analysis of intention inference results

We compare the recognition accuracy, recognition precision and recognition rate in different scenarios, using the method of LSTM. Moreover, we evaluate the classification performance over TTE on test sequences. Each time step corresponds to 100 ms. In each sub-scenario, four instances of cyclists coming from different directions related to ego vehicle are selected to analyze and evaluate the specific performance.

LSTM is a typical neural network method with the advantages of solving timing problems that was chosen for comparing with DBN in intention inference. In LSTM network setting, the deep cyclic neural network structure is constructed by stacking two cyclic layers with 800 training sessions. The learning rate is 0.001, and the attenuation rate is 0.8. In order to balance the training speed and training effect, the batch size of the small batch gradient descent method used in this paper was set to 5. During the training process, five sets of training data were randomly selected for learning model parameters. Variable learning rates were used to learn model parameters.

The confusion matrix was used to compare the intent recognition effects of DBN and LSTM. The confusion matrix is defined in Table 2.

Table 1 Number of frames in the cyclist dataset

	Longitudinal	From left	From right	Oncoming	Total
Going straight	750	750	1126	657	3283
Turning left	735	571	898	490	2694
Turning right	425	372	372	212	1381
Total	1910	1693	2396	1359	7358

Table 2 Confusion matrix definition

Inference	Reality	
	Yes	No
Yes	True positive (TP)	False positive (FP)
No	True negative (TN)	False negative (FN)

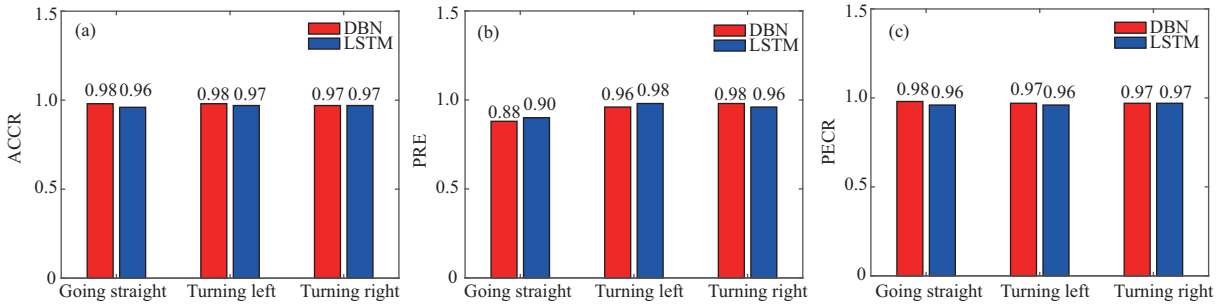

Figure 6 (Color online) Comparison of intention inference methods. (a) ACCR; (b) PRE; (c) RECR.

In each category, the accuracy rate (ACCR), precision (PRE) and recognition rate (RECR) were defined. Since intention inference was a multi-category task, the results of various intention classifications were analyzed in three 2×2 confusion matrices separately. For example, for the straight category, it is divided into going straight and non-going straight, similarly turning left and non-turning left, and turning right and non-turning right for the other two categories.

$$\mathrm{ACCR} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{TN} + \mathrm{FP} + \mathrm{FN}}, \tag{15}$$

$$\mathrm{PRE} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}}, \tag{16}$$

$$\mathrm{RECR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}}. \tag{17}$$

Figure 6 shows ACCR, PRE and RECR of different models for different intention inferences. It can be seen that the DBN has higher ACCR in inferring the intention of going straight and turning left. The inference precision reflects the credibility of the model as positive timing. In the comparison of inference precision, it can be seen that PRE of DBN for going straight is significantly lower than that of LSTM, but the PRE of DBN for turning right is higher than LSTM. The RECR indicates the ability of the model to identify positive samples. RECR of DBN is better than LSTM for all cases of going straight, turning left, or turning right.

By assuming that each intention is distinct in all scenarios, the classification performance on test tracks is evaluated, and the results demonstrate that the developed approach could classify the intention of each track correctly. The probabilities of cyclist's intention changing with TTE in the case of going straight and turning left are shown in Figure 7. The time when the cyclist goes into the intersection is defined as TTE = 0. This shows that the intent inference method could accurately classify the cyclist intentions within 17 frames after cyclists entering the intersection region. For turning cases, at the time when the corresponding intent probability raises up to 0.8 that indicates relative clear intention, the TTE of the turning-left cases is larger than that of the turning-right cases by comparing Figure 7(b) with Figure 7(c). This is related to the fact that the left turning radius is greater than the right turning radius under the rule of driving on the right. Considering the probability of three cyclist intentions varying with time in three sub-scenarios, the proposed method has great performance in turning right cases with the intention

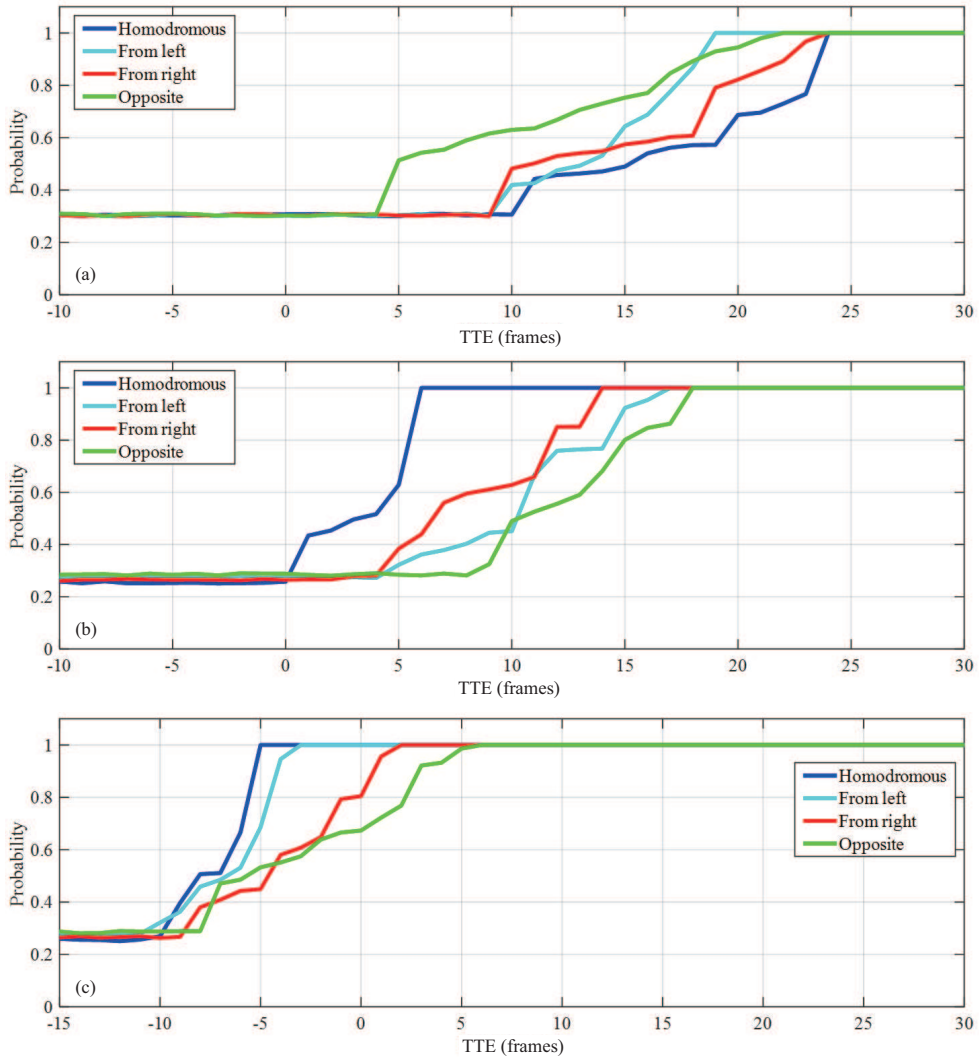

Figure 7 (Color online) The probability of cyclist intentions changes over TTE in three cases. (a) Going straight; (b) turning left; (c) turning right.

probability achieving 0.8 before 2 frames of entering the intersection for four types of cyclists as shown in Figure 7(c).

Above all, the developed approach shows a good performance in predicting cyclist intentions at intersections. The approach could infer the cyclist intentions within 0.9 s after the cyclist enters the intersection. The intent inference method based on DBN, benefiting from the intent-related cyclist motion, ego vehicle motion and environment features, not only is conducive to generating a reasonable trajectory, but also saves time for the decision layer of intelligent vehicles to take maneuvers to protect cyclists.

5.3 Analysis of trajectory prediction results

Conventional prediction methods, e.g., the constant velocity KF, do not consider the intentions of cyclists, which limits its performance in long-term prediction. To evaluate the proposed DBN+LSTM, we compare the prediction performance to the baseline method for all sub-scenarios. We use DBN for intention inference and KF for trajectory prediction [25] as the baseline methods. Section 4 shows that our approach is able to successfully predict cyclist intentions at intersections. This provides a probabilistic qualitative basis for the long-term prediction of reasonable trajectories rather than a deterministic prediction.

Euclidean metric ε was adopted to evaluate the performance of prediction results quantitatively in time domain. As shown in (18), we use the Euclidean metric of the predicted positions with respect to the ground truth (GT) which is generated by the tracked position. At prediction time step t, the predicted

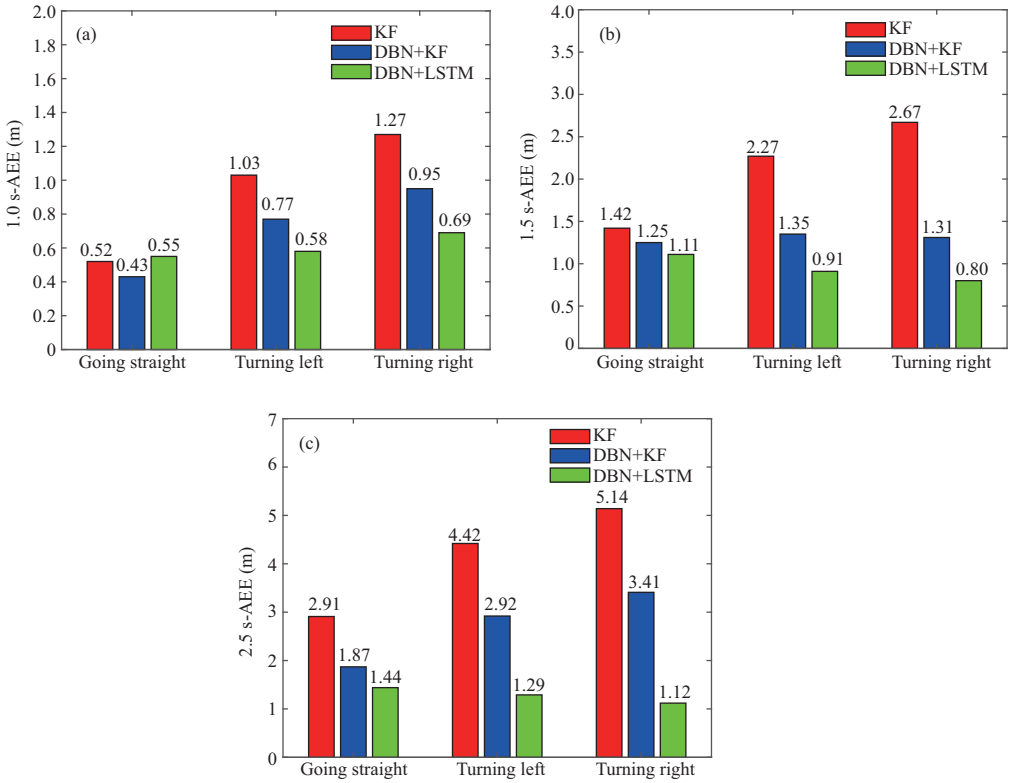

Figure 8 (Color online) Error comparison of trajectory prediction methods. (a) 1.0 s-AEE; (b) 1.5 s-AEE; (c) 2.5 s-AEE.

position of the cyclist is $(x_{\mathrm{pre}}(t), y_{\mathrm{pre}}(t))$ and the GT position is $(x_{\mathrm{GT}}(t), y_{\mathrm{GT}}(t))$.

$$\varepsilon(t) = \sqrt{(x_{\mathrm{pre}}(t) - x_{\mathrm{GT}}(t))^2 + (y_{\mathrm{pre}}(t) - y_{\mathrm{GT}}(t))^2}. \tag{18}$$

One evaluation metric based on Euclidean distance is employed, which considers a specific prediction time horizon named average Euclidean error (AEE). When all of the predictions in the test data with a specific prediction time horizon are given, AEE shows the average error at each time step [26]. The prediction horizons are chosen as 1.0 s, 1.5 s and 2.5 s respectively to evaluate the performance at different time steps.

The results of trajectory prediction methods for the prediction horizons of 1.0 s, 1.5 s and 2.5 s are illustrated in Figure 8. With the increase of prediction time, the AEE of KF and DBN+KF rises obviously, while the proposed approach keeps a relatively stable prediction performance with a slight increase of AEE. The AEE of the approach is below 1.44 m in three scenarios with the prediction time horizon of 2.5 s. For the going straight scenario, we received an AEE of 0.55 m with DBN+LSTM and 0.43 m with DBN+KF for a prediction time horizon of 1.0 s. For a prediction time horizon of 2.5 s we received 1.44 m and 1.87 m, while the KF approach showed an AEE of 0.52 m for 1 s and 2.91 m for 2.5 s. For the turning right scenario, a prediction time horizon of 1.0 s led to an AEE with DBN+LSTM of 0.69 m and 1.12 m for a horizon of 2.5 s. Using the KF we obtained 1.27 m and 5.14 m, respectively. In the turning left scenario, it shows that DBN+LSTM also performs a lower AEE than DBN+KF and KF.

The KF method does not consider any context information while the DBN+KF method does not take road structures into consideration. In contrast, our approach considers context information when anticipating the cyclist intentions. Since the road structure limits the curvature of trajectory, our approach outperforms the KF and DBN+KF not only in turning cases, but also for long prediction horizons especially when the prediction horizon is longer than 1.5 s.

Figure 9 shows the Euclidean error varies with TTE in sub-scenarios for a prediction horizon of 1.5 s. Compared to KF, the developed approach performs better in predicting trajectories for going straight and turning. The KF cannot determine if the cyclist will make a turn or keep moving straight, resulting in failures to predict the lateral movement of the cyclist. As a result, in turning left cases, the proposed approach shows similar accuracy to KF before turning, and improves accuracy after the turn is initiated.

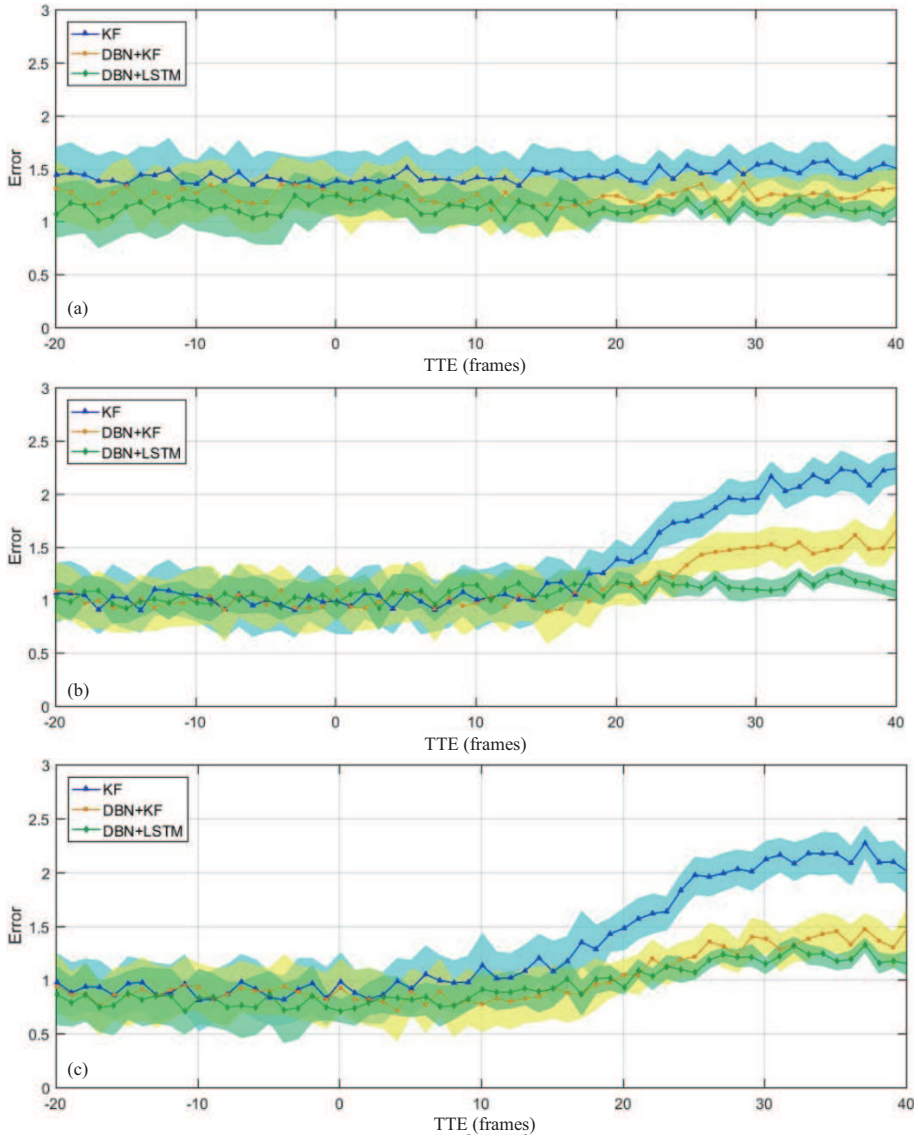

Figure 9 (Color online) Prediction error over TTE in three cases for prediction time horizon 1.0 s. (a) Going straight; (b) turning left; (c) turning right.

Compared to DBN+KF, the developed approach demonstrates improved accuracy in predicting positions based on the road structure. DBN+KF performs well for going straight. However, it does not perform well for turning cases. As shown in turning right cases, the proposed approach and DBN+KF method have similar accuracy before turning. During turning, the error of the DBN+KF goes up obviously, while the proposed approach keeps a relatively stable prediction performance with minor errors. Also as shown in Figure 8, the error in turning right cases of DBN+KF is larger than that in turning left cases. It is because the cyclist has a smaller turning angle in turning right cases.

6 Conclusion and future work

In this paper we proposed an integrated framework for cyclist trajectory prediction based on intent inference at unsignalized intersections. The intentions of cyclists at intersections are inferred via a DBN, incorporating the cyclist-related motion, ego vehicle movement and environment features. Then, taking the advantage of LSTM with encoder-decoder, an online trajectory prediction approach is presented considering not only motion dynamics, but also the cyclist intentions and environment constraints. The performance of the proposed approach is verified on the collected cyclist dataset. The developed approach

could in advance infer the cyclist intentions within 0.9 s after entering the intersection.

Compared to the baseline KF and DBN+KF, the developed approach has shown significant improvements in average error for different prediction horizons. The proposed approach connects the qualitative intent inference with the quantitative trajectory prediction domain to make a more accurate prediction within a longer prediction horizon than conventional methods KF and DBN+KF, which is significant for intelligent vehicles on both vulnerable road user protection system and the module of path-planning.

The cyclist trajectory prediction problem has been solved successfully; however, there are still some issues such as the cyclist and ego vehicle's interaction. The robustness as well as the interpretability of cyclist trajectory prediction methods need to be researched in the future.

Acknowledgements This work was supported in part by National Natural Science Foundation of China (Grant Nos. U1804161, U2013601, U20A20225), Key Research and Development Plan of Anhui Province (Grant No. 202004a05020058), Fundamental Research Funds for the Central Universities, Science and Technology Innovation Planning Project of Ministry of Education of China, NVIDIA NVAIL program, and Key Laboratory of Advanced Perception and Intelligent Control of High-end Equipment of Ministry of Education (Anhui Polytechnic University, Wuhu, China, 241000) (Grant No. GDSC202007). And experiments are conducted on NVIDIA DGX-2.

References

1 The Ministry of Public Security Traffic Management Bureau. Road Traffic Accidents in China. Statistical Report. 2018
2 Endsley M R. Toward a theory of situation awareness in dynamic systems. Hum Factors, 1995, 37: 32–64
3 Duan J, Li R, Hou L, et al. Driver braking behavior analysis to improve autonomous emergency braking systems in typical Chinese vehicle-bicycle conflicts. Accident Anal Prevention, 2017, 108: 74–82
4 Li D, Gao H. A hardware platform framework for an intelligent vehicle based on a driving brain. Engineering, 2018, 4: 464–470
5 Ma X, Luo D. Modeling cyclist acceleration process for bicycle traffic simulation using naturalistic data. Transpation Res Part F-Traffic Psychol Behav, 2016, 40: 130–144
6 Wu X, Li Z, Kan Z, et al. Reference trajectory reshaping optimization and control of robotic exoskeletons for human-robot co-manipulation. IEEE Trans Cybern, 2020, 50: 3740–3751
7 Goldhammer M, Kohler S, Doll K, et al. Camera based pedestrian path prediction by means of polynomial least-squares approximation and multilayer perceptron neural networks. In: Proceedings of 2015 SAI Intelligent Systems Conference (IntelliSys), London, 2015. 390–399
8 Gao H, Luo L, Pi M, et al. EEG-based volitional control of prosthetic legs for walking in different terrains. IEEE Trans Automat Sci Eng, 2021, 18: 530–540
9 Bieshaar M, Zernetsch S, Depping M, et al. Cooperative starting intention detection of cyclists based on smart devices and infrastructure. In: Proceedings of 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, 2017. 1–8
10 Alahi A, Goel K, Ramanathan V, et al. Social LSTM: human trajectory prediction in crowded spaces. In: Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, 2016. 961–971
11 Xue H, Huynh D Q, Reynolds M. SS-LSTM: a hierarchical LSTM model for pedestrian trajectory prediction. In: Proceedings of 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Tahoe, 2018. 1186–1194
12 Saleh K, Hossny M, Nahavandi S. Cyclist trajectory prediction using bidirectional recurrent neural networks. In: Proceedings of the 31st Australasian Joint Conference on Advances in Artificial Intelligence, Wellington, 2018. 11–14
13 Pool E A I, Kooij J F P, Gavrila D M. Using road topology to improve cyclist path prediction. In: Proceedings of 2017 IEEE Intelligent Vehicles Symposium (IV), Los Angeles, 2017. 289–296
14 Kooij J F P, Schneider N, Gavrila D M. Analysis of pedestrian dynamics from a vehicle perspective. In: Proceedings of 2014 IEEE Intelligent Vehicles Symposium, 2014. 1445–1450
15 Gao H, Zhu J, Su H, et al. Automatic parking control of unmanned vehicle based on switching control algorithm and backstepping. IEEE/ASME Trans Mech, 2020. doi: 10.1109/TMECH.2020.3037215
16 Xie G, Gao H, Qian L, et al. Vehicle trajectory prediction by integrating physics- and maneuver-based approaches using interactive multiple models. IEEE Trans Ind Electron, 2018, 65: 5999–6008
17 Gao H, Cheng B, Wang J, et al. Object classification using CNN-based fusion of vision and LIDAR in autonomous vehicle environment. IEEE Trans Ind Inf, 2018, 14: 4224–4231
18 Zhang X, Gao H, Li G, et al. Multi-view clustering based on graph-regularized nonnegative matrix factorization for object recognition. Inf Sci, 2018, 432: 463–478
19 Li Z, Huang B, Ye Z, et al. Physical human-robot interaction of a robotic exoskeleton by admittance control. IEEE Trans Ind Electron, 2018, 65: 9614–9624
20 Kooij J F P, Flohr F, Pool E A I, et al. Context-based path prediction for targets with switching dynamics. Int J Comput Vis, 2019, 127: 239–262
21 Gao H, Zhu J, Zhang T, et al. Situational assessment for intelligent vehicles based on stochastic model and Gaussian distributions in typical traffic scenarios. IEEE Trans Syst Man Cy-S, 2020. doi: 10.1109/TSMC.2020.3019512

22 Wakim C F, Capperon S, Oksman J. A Markovian model of pedestrian behavior. In: Proceedings of 2004 IEEE International Conference on Systems, Man and Cybernetics, The Hague, 2004. 4028–4033

23 Levinson J, Thrun S. Robust vehicle localization in urban environments using probabilistic maps. In: Proceedings of 2010 IEEE International Conference on Robotics and Automation, Anchorage, 2010. 4372–4378

24 Koehler S, Goldhammer M, Bauer S, et al. Stationary detection of the pedestrian's intention at intersections. IEEE Intell Transp Syst Mag, 2013, 5: 87–99

25 Huang L, Wu J. Study on the cyclist behavior at signalized intersections. In: Proceedings of the 2003 IEEE International Conference on Intelligent Transportation Systems, Shanghai, 2003. 317–322

26 Ling H, Wu J. A study on cyclist behavior at signalized intersections. IEEE Trans Intell Transp Syst, 2004, 5: 293–299