Accident Analysis and Prevention 148 (2020) 105799

Contents lists available at ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Prediction of pedestrian-vehicle conflicts at signalized intersections based on long short-term memory neural network☆

Shile Zhang *, Mohamed Abdel-Aty, Qing Cai, Pei Li, Jorge Ugan

Department of Civil, Environmental & Construction Engineering, University of Central Florida, Orlando, FL 32816, USA

ARTICLE INFO

Keywords:
Pedestrian-vehicle conflicts
LSTM
Neural network
Connected vehicles

ABSTRACT

Pedestrian protection is an important component of road safety. Intersections are dangerous locations for pedestrians with mixed traffic. This paper aims to predict potential traffic conflicts between pedestrians and vehicles at signalized intersections. Using detection and tracking techniques in computer vision, pedestrians' and vehicles' features are extracted from video data. An LSTM (Long Short-term Memory) neural network is proposed to predict the pedestrian-vehicle conflicts 2 s ahead. The established model reaches an accuracy of 88.5 % at one signalized intersection. It is further tested at a new intersection, reaching the accuracy of 84.9 %, while the new data merely takes up 30 % of the training data set. This indicates that the proposed model is promising to be implemented at different locations. Moreover, the proposed model can also be applied to develop collision warning systems under the Connected Vehicles' environment.

1. Introduction

Pedestrians are regarded as vulnerable road users (VRUs). Each year, thousands of pedestrian and cyclist deaths are caused by traffic collisions, which take up 16 % of the total road fatalities and injuries in the U.S. (FHWA, 2018). In 2018, there were 6283 pedestrian deaths, ranked the highest in the last three decades. Among them, 26 % happened from 6 pm and 9 pm (NHTSA, 2019a). The intersections are one of the dangerous road locations with relatively high pedestrian volume, and mixed traffic volume of pedestrians and vehicles. Annually, approximately 18 % of pedestrian fatalities happened at the intersections in the U.S. (NHTSA, 2019b). To further improve intersection safety, studies have been carried out to investigate collision-related variables (Gårder, 1989; Lee and Abdel-Aty, 2005). With more safety countermeasures to be implemented, it is also important to model and predict potential collisions in advance to warn drivers, with the development of Connected Vehicle (CV) technologies.

Traditional traffic safety studies mainly used crash data. However, crash data were not usually complete or accurate, and sometimes failed to reveal the true contributing factors of collisions (Ismail et al., 2009). Surrogate Safety Measures (SSMs) were proposed (Tarko et al., 2009) and used to measure collision risks (Fu et al., 2018; Khosravi et al., 2018; Wu et al., 2019). Indicators of SSMs included Post-Encroachment Time (PET, Cooper (1984)), Time to Collision (TTC, Hayward (1972)), Gap Time (GT, Vogel (2002)), etc. PET was regarded as an appropriate indicator to capture conflicts between pedestrians and vehicles (Ismail et al., 2009). Given a predetermined threshold, small PET values could denote the proximity of collisions (Cooper, 1984; Ismail et al., 2009; Mizoguchi et al., 2017).

Previous studies investigated the factors contributing to pedestrian-vehicle conflicts. Environmental factors, such as the signal timing (Gårder, 1989) and geometric design of the intersections (Gårder, 1989; Salamati et al., 2011), could influence pedestrian-vehicle conflicts. Chen et al. (2017) conducted safety evaluation on an intersection and found more severe conflicts (small PET values) outside the crosswalks. Besides, the drivers' yielding behaviors (Fu et al., 2018), pedestrians' acceptable gaps, and pedestrians' yielding behaviors (Tageldin et al., 2017) were found to be significant factors for pedestrian safety at the intersections.

Pedestrians' violation behaviors were found to be significant for pedestrian-vehicle conflicts at the signalized intersections, as violating pedestrians were exposed to motorized traffic without the protection of traffic signals. For example, the pedestrians' spatial violations, i.e., crossing outside the crosswalks, were found to be positively correlated with the number of traffic conflicts (Zaki et al., 2013). In addition, pedestrians' characteristics affected crossing behaviors, which could increase the irregularities of their motions. For example, pedestrians

☆ This paper has been handled by associate editor Tony Sze.

* Corresponding author.
E-mail address: shirleyzhang@Knights.ucf.edu (S. Zhang).

https://doi.org/10.1016/j.aap.2020.105799
Received 15 March 2020; Received in revised form 18 June 2020; Accepted 21 September 2020
Available online 17 October 2020

walking in groups had lower walking velocities and higher commonality (Hediyeh et al., 2014). And females were found to walk slower than males (Montufar et al., 2007). More emphasis should be placed on integrating pedestrians' characteristics into the pedestrian crossing safety.

Compared with traditional data sources, including radar, loop detector, and Bluetooth sensors, video data could offer a microscopic view for pedestrian safety analysis (Wang et al., 2012; Ka et al., 2019; Wu et al., 2019; Zhang et al., 2020). Wang et al. (2012) developed a smartphone application to alarm pedestrians. It could recognize vehicles from both the front view and the back view. But as the application used embedded cameras of smartphones, it could not be used if the smartphone was in the pocket or facing the ground. Ka et al. (2019) predicted pedestrians' red- light violations based on the pedestrians' characteristics (age, gender, head orientation, etc), and send warnings to drivers. These studies offered new insights into improving pedestrian safety using video data. However, the pedestrians-involved conflicts were sophisticated, taking into consideration various kinds of pedestrian-vehicle interactions (Schneemann and Gohl, 2016; Yue et al., 2020). For example, pedestrians were found to take different evasive actions when approaching vehicles had different velocities, and either party could make abrupt movements during an interaction process (Ni et al., 2016).

To summarize, the research gap lies in modeling pedestrian safety based on the features of vehicles and pedestrians throughout the pedestrian-vehicle interactions. To fill the gap, the trajectories generated from videos can bring more possibilities to model and predict the occurrences of pedestrian-vehicle conflicts.

To better handle time series data in the transportation field, neural networks have been successfully applied (Manh and Alaghband, 2018; Cai et al., 2019; Du et al., 2019; Gong et al., 2019). And LSTM (Long Short-term Memory) neural network (Hochreiter and Schmidhuber, 1997), as an advanced Recurrent Neural Network (RNN), could connect the information between the last time window to the next time window. Thus, it is more effective at capturing the sequential information lying in the time series data (Altché and Fortelle, 2017; Yuan et al., 2019; Li et al., 2020). LSTM neural network proved to be effective in predicting pedestrians' motions such as trajectory predictions (Manh and Alaghband, 2018;), and predictions of pedestrians' interactions in crowded spaces (Alahi et al., 2016). It can be better used to predict the occurrences of pedestrian-vehicle conflicts using trajectory data generated from videos.

Based on the above discussion, this study is intended to predict the pedestrian-vehicle conflicts at signalized crosswalks using an LSTM neural network. With video data generated from real traffic scenes, pedestrian-vehicle interactions are collected and divided into three categories, safe interactions, slight conflicts, and severe conflicts. Pedestrians' characteristics are generated using computer vision techniques. The LSTM model is established and well trained. The experiment result at the original intersection shows that the LSTM model achieves the accuracy of 88.5 %. The external experiment result at another intersection shows that the model achieves the accuracy of 84.9 % (around 85 %), when merely 30 % external data are fed during the training process. This indicates the model is promising to generalize well at different locations. In addition, it can be further implemented in the collision warning systems under the Connected Vehicle (CV) environment.

2. Data collection

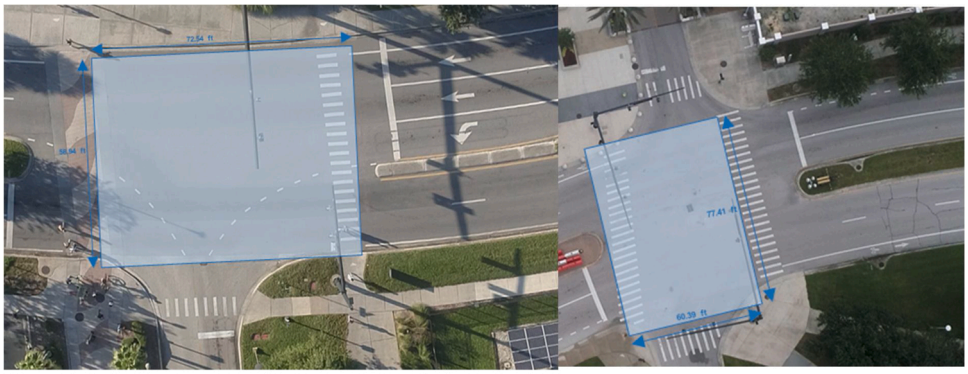
To analyze pedestrians' crossing behaviors at intersections, video data from two intersections (Intersection 1: (latitude: 28.5963094, longitude: −81.1993496); Intersection 2: (latitude: 28.606160, longitude: −81.197373)) were collected, as shown in Fig. 1. The traffic volumes at the two intersections were around 200veh/h and 150veh/h, respectively. The videos were collected on different days using GOPRO HERO 7 camera. Data from both intersections (marked in shadow areas) were collected during daytime (16:00−17:00) and evening (18:00−19:00, with street light) on sunny clear weekdays in October 2019. All video data are of good quality to generate the trajectories of road users.

2.1. Evaluation of the crosswalk safety at the studied site

Post-Encroachment Time (PET) was defined as the time difference between the moment when the first road user left the potential conflict zone and the moment when the second user reached it (Allen et al., 1978). This was an indicator typically used for denoting pedestrian safety in previous studies (Ismail et al., 2009; ; Wu et al., 2019), as it could measure the proximity of the road users to analyze crossing conflicts. As shown in Fig. 2(a), t_1 was the moment when the pedestrian left the conflict zone, and t_2 was the moment when the vehicle reached the same zone. The Fig. 2(b) shows the converse case when the vehicle left first and the pedestrian reached. And the time difference between t_2 and t_1 was defined as PET, as shown in Eq. 1. According to the literature (Radwan et al., 2016; Kathuria and Vedagiri, 2020), if the PET value during the interaction was smaller than 3 s, the situation was regarded as severe conflict. If the PET value was between 3 s and 6 s, the situation was regarded as a slight conflict. If the PET value was larger than 6 s, the situation was regarded as a safe interaction.

$$PET = t_2 - t_1 \tag{1}$$

t_2: *the moment when the vehicle (pedestrian) reached the area of potential collision*



(a) Intersection 1 (b) Intersection 2

Fig. 1. The studied locations.

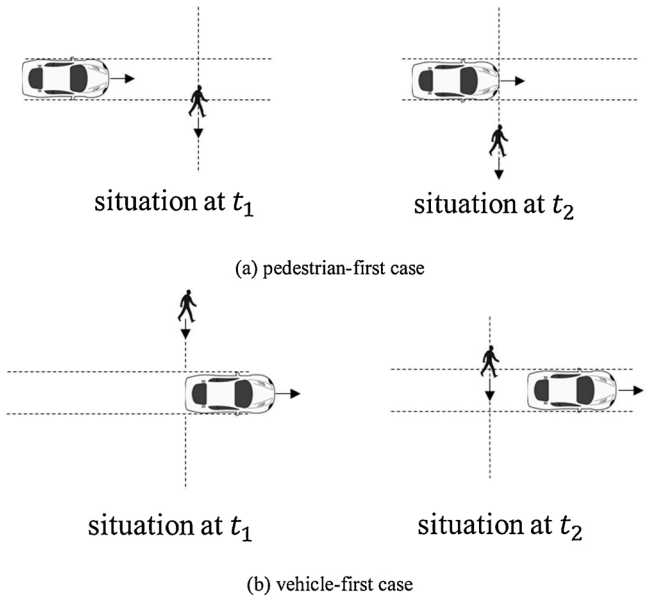


Fig. 2. Illustration of PET calculation.

t_1 : *the moment when the pedestrian(vehicle) left the area of potential collision*

Pedestrian-vehicle conflicts were manually collected to ensure analysis accuracy. At the first crosswalk, 334 pedestrians and 69 traffic conflicts were collected. Forty happened during daytime, and 29 happened during evening. At the second crosswalk, videos of 254 pedestrians and 62 traffic conflicts were collected. Among them, 48 happened during daytime and 14 happened during evening. The details can be found in Fig. 3.

2.2. Video processing

To generate the trajectories of pedestrians and vehicles, computer vision techniques including object detection, object tracking, and perspective transformation were used.

2.2.1. Object detection

YOLO (You only look once) is a real-time object detection model first proposed by Redmon et al. (2016). It could apply a single neural network to the full image, dividing different areas (anchor boxes) and classifying the objects in these areas at the same time. This characteristic made it more efficient to use, compared with two-stage models such as R-CNN (Girshick, 2015; Ren et al., 2015). YOLOv3 model (Redmon and Farhadi, 2018) improved the original model by using multi-scale images, data augmentation, and batch normalization during the training procedure. YOLOv3 proved to be effective on the COCO dataset (Lin et al., 2014), a standardized large-scale data set for evaluating the performance of object detection algorithms. YOLOv3 has been used to detect road users from traffic video data in previous studies (Jana et al., 2018; Lin and Sun, 2018).

2.2.2. Object tracking

To follow the movements of multiple road users appearing in the scene, the Deep SORT model (Wojke et al., 2017; Wojke and Bewley, 2018; Qidian et al., 2020) was used. The model assigned unique tracker IDs to each pedestrian and each vehicle recognized by the detection model, and followed their movements. As shown in Fig. 4, the blue bounding boxes were generated from the YOLOv3 model, and the white bounding boxes were from the Deep SORT model. The green numbers were the tracker IDs for pedestrians, and the white numbers were the tracker IDs for vehicles.

Deep SORT was evaluated on the MOT16 Challenge benchmark (Milan et al., 2016), a standardized benchmark for evaluating the performance of different Multiple Object Tracking algorithms. Deep SORT outperformed previous models from the perspectives of MOTA score (Multi-object tracking accuracy), and reducing FN (false negatives), etc (Wojke et al., 2017). The Deep SORT had a few applications in transportation field (Arvind et al., 2019; Hou et al., 2019).

2.2.3. Perspective transformation

The purpose of the perspective transformation was to create a mapping from the image plane to the world plane. A homograph matrix $\boldsymbol{h}$ was used to transform the coordinates extracted from videos to the world coordinates. Matrix $\boldsymbol{h}$ was composed of nine values from h_1 to h_9. As shown in Eq. 2, the points correspondences from videos and Google

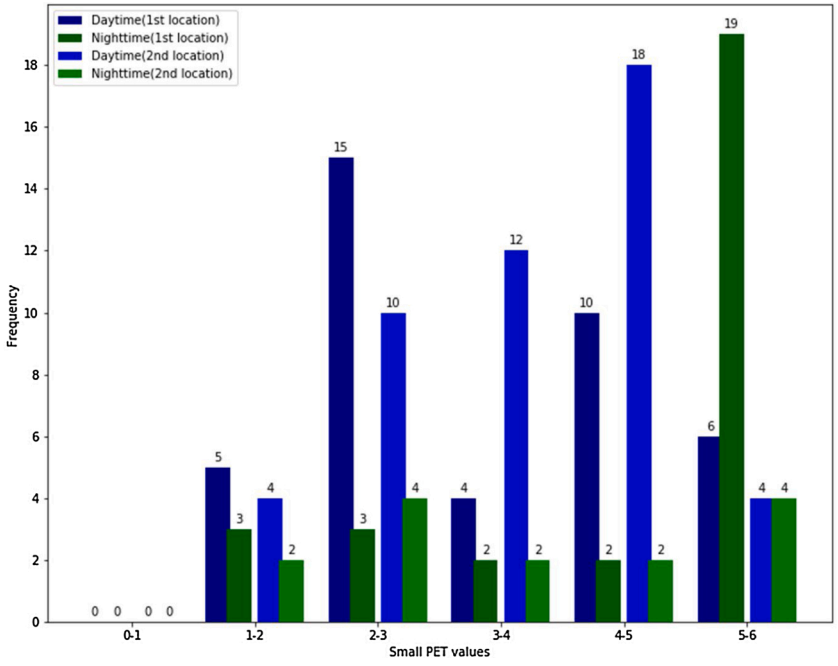


Fig. 3. Distribution of small PET values.

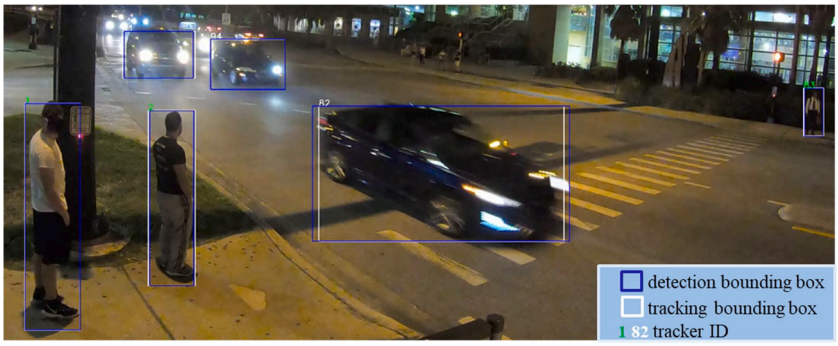


Fig. 4. Object detection and object tracking at the studied area.

Maps©, i.e., (u_i, v_i), and (X_i, Y_i), formed matrix A. The Singular Value Decomposition (SVD) could be used to solve Eq. 2 with of $h_9 = 1$ (Naphade et al., 2019; Španhel et al., 2019; Tang et al., 2019). After generating matrix $\boldsymbol{h}$, the image coordinates could be converted to the world coordinates through the inverse matrix of $\boldsymbol{h}$, thus generating the correct world coordinates of road users.

$$A * \boldsymbol{h} = \begin{bmatrix} 0 & 0 & 0 & -X_1 & -Y_1 & 1 & v_1X_1 & v_1Y_1 & v_1 \\ X_1 & Y_1 & 1 & 0 & 0 & 0 & -u_1X_1 & -u_1Y_1 & -u_1 \\ 0 & 0 & 0 & -X_2 & -Y_2 & 1 & v_2X_2 & v_2Y_2 & v_2 \\ X_2 & Y_2 & 1 & 0 & 0 & 0 & -u_2X_2 & -u_2Y_2 & -u_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \\ h_7 \\ h_8 \\ h_9 \end{bmatrix} = 0 \quad (2)$$

(u_i, v_i) : *coordinate of each point on image plane*

(X_i, Y_i) : *coordinate of each point on world plane*

$\boldsymbol{h} = [h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8, h_9]^T$

From this step, location information of pedestrians and vehicles was generated at the frequency of 60 Hz. The trajectories of road users (pedestrians and vehicles) were generated from videos, as shown in Fig. 5. Different road users are marked with different colors.

Another transformation procedure was conducted to create a generic coordinate system for different intersections. As shown in Eq. 3, the matrix $\boldsymbol{M}$ is used to convert world coordinates (X_i, Y_i) to the scaled coordinates $(X_{i\_scale}, Y_{i\_scale})$ according to the scale of the intersections. Matrix $\boldsymbol{M}$ could be calculated using the four-points perspective transformation method offered by OpenCV packages (Szeliski, 2010). And the four points were the four corners at this intersection, as shown in Fig. 6. Basically, the camera covers the areas within the boundary formed by the four points. Take the top-left point for example, the world coordinate of this point is $(X_1, Y_1) = (-81.1993496, 28.5963094)$, and the scaled coordinate is $(X_{1\_scale}, Y_{1\_scale}) = (0, 58.94)$, as the height of the rectangle area is 58.94 ft. Though this step, the world coordinates were converted to the scaled coordinates, and then normalized to feed into the model. This ensured that the model can be further implemented to more intersections with different geometric designs.

$$\begin{pmatrix} X_{i\_scale} \\ Y_{i\_scale} \\ 1 \end{pmatrix} = \boldsymbol{M} \begin{pmatrix} X_i \\ Y_i \\ 1 \end{pmatrix} \quad (3)$$

2.3. Variables description

From the above procedures, variables obtained from both intersections are listed in Table 1. The independent variables were composed of pedestrians' features, such as gender (male/female), pedestrian coordinates $\left(X_i^{ped}, Y_i^{ped}\right)$, walking directions (towards/away from camera), whether the pedestrians crossed during the red light (yes/no), as well as vehicle coordinates $\left(X_j^{veh}, X_j^{veh}\right)$. The variables except for pedestrians' genders and pedestrians crossed during red light were automatically generated from videos. The pedestrians' and vehicles' coordinates were preprocessed to feed into the model, as mentioned above. The dependent variables in this study were traffic conflicts between pedestrians and vehicles, which were divided into three categories, i.e., severe conflicts ($PET \leq 3s$), slight conflicts ($3s < PET \leq 6s$), and no conflicts/safe interactions ($PET > 6s$).

For predicting purpose, also to implement the system prototype under the CV environment, suppose the driver can get warning for pedestrian θ units of time ahead. Considering drivers' reaction time, θ was taken as 2 s (Wilson et al., 1997; Obeid et al., 2017). The trajectories of pedestrians and vehicles were extracted 2 s before reaching the conflict zones.

(a) pedestrians' trajectories (b) vehicles' trajectories

Fig. 5. Trajectories of pedestrians & vehicles at 1st location on Google Maps.

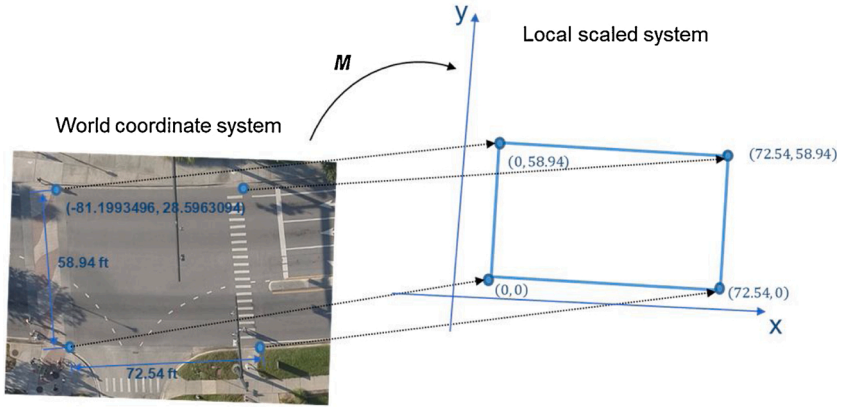


Fig. 6. Transformation from world coordinates to scaled coordinates*.

Table 1
Summary of variable descriptive statistics.

Variable	Description	Distribution
Gender	"male" or "female"	("male" = 389, "female" = 199)
Crossed during red light	"yes" or "no"	("yes" = 182, "no" = 406)
Walking directions	"towards" or "away"	("towards" = 305, "away" = 283)
Pedestrian locations X_i^{ped}	Coordinates	(0, 1)
Pedestrian location Y_i^{ped}	Coordinates	(0, 1)
Vehicle locations X_j^{veh}	Coordinates	(0, 1)
Vehicle locations Y_j^{veh}	Coordinates	(0, 1)
Traffic conflicts	**"no", "slight", "severe"**	**("no" = 457, "slight" = 85, "severe" = 46)**

*Note: "gender" and "crossed during red light" were manually labeled. **Bold** marked is the dependent variable.

3. Methodologies

As trajectory data are time series data, an LSTM neural network (Hochreiter and Schmidhuber, 1997) model can better capture the temporal relationships lying in the data. LSTM neural network is an advanced Recurrent Neural Network (RNN). As Recurrent Neural Networks are less effective to learn long-term dependency from time series data (Graves et al., 2013), the LSTM neural network is proposed to solve this problem.

An LSTM neural network model is usually composed of the input layer, multiple hidden layers, and the output layer. The advantage of the LSTM model is the design of its units inside the hidden layer. An LSTM unit, with the unique memory cell, can remember historical information lying in the sequence data, such as speech (temporal sequence) or image data (spatial sequence). The illustration of an LSTM unit is shown in Fig. 7. Given the time window t, the unit is composed of an input gate i_t, a forget gate f_t, an output gate o_t. These three gates control the information flow in the unit. The i_t, f_t, and o_t are calculated using weight matrices W, the input sequence x_t, and the last layer output h_{t-1}. c_t is the cell activation vector, which is formed by two elementwise products of the vectors. It is also referred to as the cell state, containing the information which the unit is going to store. The output h_{t-1}, and the cell state c_{t-1} from the last layer are used as the input for the current layer, which is similar to the normal recurrent neural networks. The input sequence x_t is computed by Eqs. 4–8 to generate the hidden layer output h_t, which is a vector of probabilities. So, the output sequence y_t for the neural network is calculated iteratively from hidden layer output h_t, as shown in Eq. 9.

(adapted from (Graves et al., 2013; Kang et al., 2017)).

$$i_t = \sigma(W_{xi}x_t + W_{hi}h_{t-1} + W_{ic}c_{t-1} + b_i) \quad (4)$$

$$f_t = \sigma(W_{fx}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (5)$$

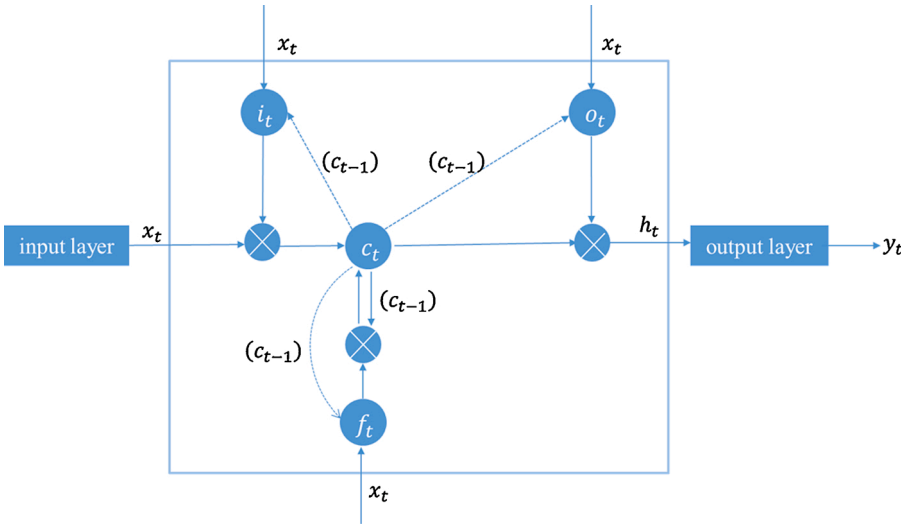


Fig. 7. Schematic of LSTM unit.

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (6)$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \emptyset(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (7)$$

$$h_t = o_t \otimes \emptyset(c_t) \quad (8)$$

$$y_t = W_{hy}h_t + b_y \quad (9)$$

σ: logistics sigmoid function
⊗: elementwise product of the vectors
∅: activation function tanh

The model architecture used in this study is shown in Fig. 8. The model contains two stacked LSTM layers and a dense layer. Features from four time slices are fed into the model as the input of the next layer. The output neurons denote the probabilities that one record is classified as each of the targeted categories. Softmax function is the activation function. Adam (Kingma and Ba, 2014) is the optimization function. The decay value for the optimization function is set to be 0.1 to avoid overfitting. The model is implemented in the Keras framework (Chollet, 2015).

4. Experiment and results

The proposed model is first trained and tested using data from one intersection (Fig. 1(a)). There are 35,647 records in the data set after slicing and stacking the features from different time slices, with a ratio of 28:2:1 between the targeted classes "safe interaction", "slight conflicts", and "severe conflicts". Eighty percent of the data are used as the training data set, and twenty percent of the data are used as the test data set. An over-sampling method SMOTE (Synthetic Minority Over-Sampling Technique, Chawla et al. (2002)) is used on the training data set to increase the number of records for the two minority classes, i.e., "slight conflicts", and "severe conflicts". SMOTE is a popular over-sampling method, which can generate new minority class records by interpolating between several minority class examples that lie together. It should be noted that SMOTE is only applied to the training data set, while the test data set still uses original records.

The proposed LSTM model is well trained before getting overfitted. As shown in Table 2, the batch size is selected as 1000, the learning rate is 0.005, and the unit number in the LSTM layer is 64. The epoch number for the training process is 30.

Table 2
Hyperparameters.

Hyperparameter	Tuning range	Selected value
Batch size	100, 500, 1000, 5000	1000
Learning rate	0.001, 0.005, 0.01	0.005
LSTM unit number	32, 64, 128	64

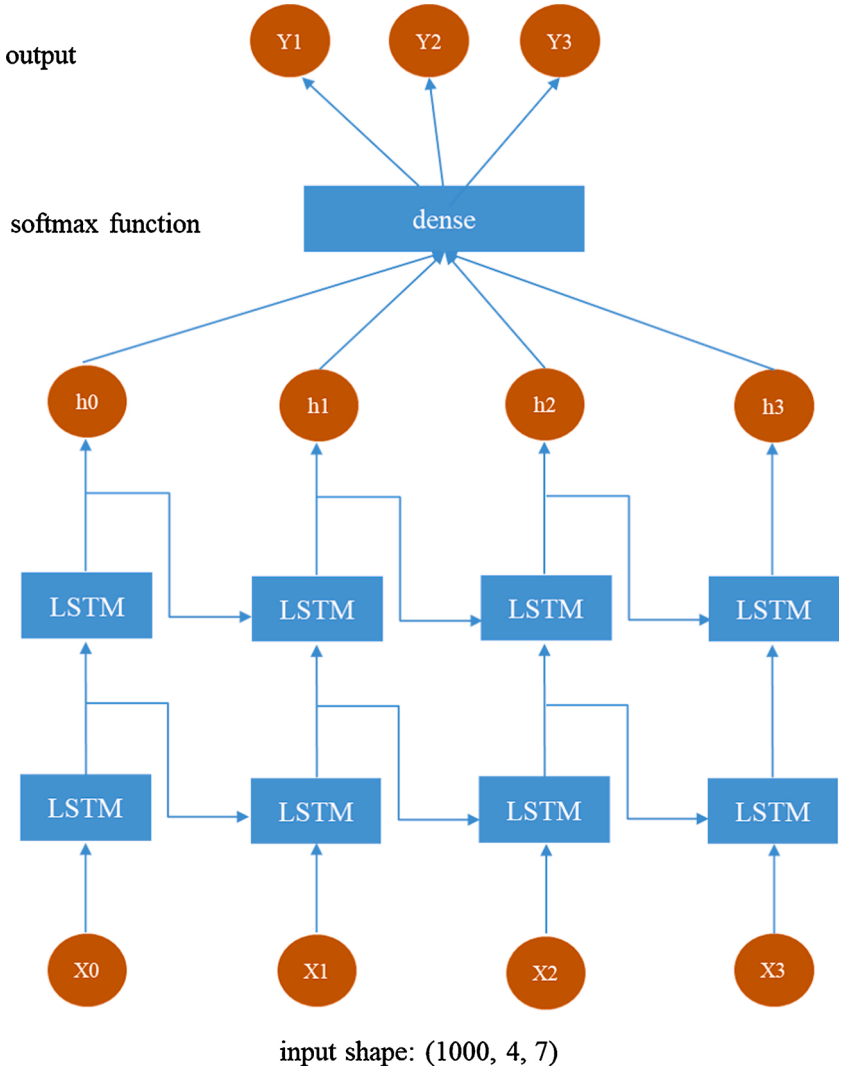


Fig. 8. Model architecture.

4.1. Evaluating metrics

The diagram for evaluation metrics calculation is shown in Table 3. TP (true positive) means the number of actual positive samples that are correctly classified. FP (false positive) means the number of actual negative samples that are wrongly classified. FN (false negative) is the number of actual negative samples that are wrongly classified. TN (true negative) is the number of actual negative samples that are correctly classified.

Using these four values, the equations of calculating the metrics are listed in Eqs. 10–13. Precision, also called positive predictive value (PPV), is the ratio of actual positive samples to the classified positive samples. Recall, also called sensitivity, is the proportion of the actual positive samples that are correctly classified. F1 score is an integrated metric taking into consideration both precision value and recall value. Accuracy is defined as the ratio of the correctly classified samples in the whole data set, taking into consideration both positive samples and negative samples.

$$Precision = \frac{TP}{TP + FP} \tag{10}$$

$$Recall = \frac{TP}{TP + FN} \tag{11}$$

$$F1\ score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{12}$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \tag{13}$$

4.2. Experiment results

The experiment is first carried out using the data from the first intersection, which is also regarded as internal testing. For comparison, a machine learning model, Support Vector Machine (SVM) (Cortes and Vapnik, 1995) and a simple Deep Neural Network (DNN) are also used. The SVM model is implemented using the function "SVC" with a linear kernel from the Scikit-learn package (Pedregosa et al., 2011). DNN model is also referred to as Artificial Neural Network (ANN), with multiple layers used to generate the output from the input data. The DNN model here is composed of three dense layers, which are the main differences with the LSTM model. The dense layer is not capable of learning sequential information. The optimization function and the activation function are the same as the LSTM model. The model is well tuned to reach the best result.

Among the three models, the LSTM model achieves the best results from the perspectives of all evaluation metrics. As shown in Table 4, the SVM model reaches the accuracy of 56.1 %, and the DNN model is 76.6 %. The LSTM model achieves the precision of 88.9 % over both classes. And the recall value is 88.2 %, the F1 score is 88.5 %. The model achieves an overall accuracy of 88.5 % on the test data set.

4.3. Experiment results (external testing)

The data set collected at the second intersection (shown in Fig. 1(b)) is regarded as external data set, which contains 20,335 records in total, with the ratio at around 22:4:1 between three targeted classes. SMOTE is used to balance different classes. The objective of the external testing is to prove that the LSTM model can be implemented at different locations.

Table 3
Confusion matrix.

	Actual value	
Classified value	Positive	Negative
Positive	TP (true positive)	FP (false positive)
Negative	FN (false negative)	TN (true negative)

Table 4
Internal testing result.

	Training set	Test set			
Model	Accuracy	Precision	Recall	F1 score	Accuracy
SVM	0.670	0.831	0.561	0.670	0.561
DNN	0.800	0.861	0.766	0.811	0.766
LSTM	0.889	0.889	0.882	0.885	0.885

The idea is to further train the model by gradually increasing more external data in the training set, while keeping the original training data from the 1st intersection. As shown in Table 5, the first column shows the ratios of external data in the training data set, and the second column shows the experiment results on the external test set. When there are no external data in the training data set, the previous model achieves the precision rate of 62.0 %, the recall rate of 61.6 %, and the overall accuracy of 61.6 %. With more external data used in the training process, the model's performance at the external location gets improved as well. When the external data take up 10 % of the training data set, the model achieves the precision of 79.4 % and the accuracy of 75.9 %. When the external data take up 30 % of the training data set, the model achieves the accuracy of 84.9 %. If the external data continue to increase, up to 50 % of the training data set, the model achieves performances that are similar to the original intersection, from the perspective of accuracy. Models are well trained before getting overfitted. And each model achieves an accuracy of around 88.5 % on the test data set of the first intersection (the same result as internal testing). In other words, the model's performance doesn't get worse at the original intersection, with the external data increasing in the training data set.

The experiment results indicate that when there are two intersections, the ratio of 30 % of external data will improve the model's accuracy to 84.9 % (around 85 %) on the test data set of the new location, which can be seen as an ideal accuracy rate. This indicates that the model can be further trained and implemented with more external data, to be implemented at different intersections.

5. Conclusion and discussion

In this paper, an LSTM neural network model is employed to predict the pedestrians' safety situations, denoting by different PET values. Based on detection and tracking techniques in computer vision, the characteristics of pedestrians and vehicles are fed into the model. The proposed model achieves the accuracy of 88.5 % at one signalized intersection. The external test indicates that including 30 % new data significantly improves the model performance at a different location to an ideal accuracy (84.9 %). The results imply that the characteristics during pedestrian-vehicle interaction processes will reflect the potentially dangerous situations of pedestrians. Moreover, the model can be further trained and implemented at different locations with a smaller size of new data set required.

Different from traditional studies, this paper predicts the pedestrians' conflicts in time series before the pedestrians and vehicles reach

Table 5
External testing result.

Ratio of external data in the training data set	Prediction result on the test set (external data)			
	Precision	Recall	F1 score	Accuracy
0 %	0.620	0.616	0.618	0.616
10 %	0.794	0.759	0.776	0.759
20 %	0.800	0.797	0.798	0.803
30 %	0.815	0.806	0.810	0.849
40 %	0.819	0.817	0.818	0.854
50 %	0.883	0.883	0.883	0.885

the conflict zones. And different geometric designs of intersections are taken into considerations by transforming the location coordinates. More research can be further conducted to implement the model in the field experiments with Connected Vehicles' technologies, to better warn drivers.

The limitation of the study is only the PET indicator is used. As various indicators of SSMs have different features, other indicators such as TTC, GT, or the integrations of multiple indicators, can also be used to identify the dangerous situations of pedestrians.

CRediT authorship contribution statement

Shile Zhang: Conceptualization, Methodology, Software, Writing - original draft. **Mohamed Abdel-Aty:** Conceptualization, Methodology, Validation, Writing - review & editing. **Qing Cai:** Formal analysis, Validation, Writing - review & editing. **Pei Li:** Methodology, Data curation, Software, Validation. **Jorge Ugan:** Data curation, Writing - original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., Savarese, S., 2016. Social lstm: human trajectory prediction in crowded spaces. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 961–971.

Allen, B.L., Shin, B.T., Cooper, P.J., 1978. Analysis of Traffic Conflicts and Collisions.

Altché, F., Fortelle, A.D.L., 2017. An lstm network for highway trajectory prediction. 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC) 353–359.

Arvind, C., Jyothi, R., Mahalakshmi, K., Vaishnavi, C., Apoorva, U., 2019. Vision based driver assistance for near range obstacle sensing under unstructured traffic environment. In: 2019 IEEE Symposium Series on Computational Intelligence (SSCI). IEEE, pp. 1163–1170.

Cai, Q., Abdel-Aty, M., Sun, Y., Lee, J., Yuan, J., 2019. Applying a deep learning approach for transportation safety planning by using high-resolution transportation and land use data. Transp. Res. Part A Policy Pract. 127, 71–85.

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. Smote: synthetic minority over-sampling technique. J. Artif. Int. Res. 16 (1), 321–357.

Chen, P., Zeng, W., Yu, G., Wang, Y., 2017. Surrogate Safety Analysis of Pedestrian-vehicle Conflict at Intersections Using Unmanned Aerial Vehicle Videos.

Chollet, F., 2015. Keras. GitHub.

Cooper, P.J., 1984. Experience With Traffic Conflicts in Canada with Emphasis on "post Encroachment Time" Techniques.

Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20 (3), 273–297.

Du, X., Vasudevan, R., Johnson-Roberson, M., 2019. Bio-lstm: a biomechanically inspired recurrent neural network for 3-d pedestrian pose and gait prediction. IEEE Robot. Autom. Lett. 4 (2), 1501–1508.

Fhwa, U.S.D.O.T., 2018. Pedestrian & Bicycle Safety.

Fu, T., Miranda-Moreno, L., Saunier, N., 2018. A novel framework to evaluate pedestrian safety at non-signalized locations. Accid. Anal. Prev. 111, 23–33.

Gårder, P., 1989. Pedestrian safety at traffic signals: a study carried out with the help of a traffic conflicts technique. Accid. Anal. Prev. 21 (5), 435–444.

Girshick, R., 2015. Fast r-cnn. In: Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV). IEEE Computer Society, pp. 1440–1448.

Gong, Y., Abdel-Aty, M., Cai, Q., Rahman, M.S., 2019. Decentralized network level adaptive signal control by multi-agent deep reinforcement learning. Trans. Res. Interdisciplinary Perspectives 1, 100020.

Graves, A., Mohamed, A., Hinton, G., 2013. Speech recognition with deep recurrent neural networks. In: 2013 IEEE International Conference on Acoustics. Speech and Signal Processing, pp. 6645–6649.

Hayward, J.C., 1972. Near Miss Determination through Use of a Scale of Danger.

Hediyeh, H., Sayed, T., Zaki, M.H., Ismail, K., 2014. Automated analysis of pedestrian crossing speed behavior at scramble-phase signalized intersections using computer vision techniques. Int. J. Sustain. Transp. 8 (5), 382–397.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780.

Hou, X., Wang, Y., Chau, L., 2019. Vehicle tracking using deep sort with low confidence track filtering. 2019 16th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) 1–6.

Ismail, K., Sayed, T., Saunier, N., Lim, C., 2009. Automated analysis of pedestrian–vehicle conflicts using video data. Transp. Res. Rec. 2140 (1), 44–54.

Jana, A.P., Biswas, A., Mohana, 2018. Yolo based detection and classification of objects in video records. In: 2018 3rd IEEE International Conference on Recent Trends in Electronics. Information & Communication Technology (RTEICT), pp. 2448–2452.

Ka, D., Lee, D., Kim, S., Yeo, H., 2019. Study on the framework of intersection pedestrian collision warning system considering pedestrian characteristics. Transportation Research Record: Journal of the Transportation Research Board.

Kang, D., Lv, Y., Chen, Y., 2017. Short-term traffic flow prediction with lstm recurrent neural network. 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC) 1–6.

Kathuria, A., Vedagiri, P., 2020. Evaluating pedestrian vehicle interaction dynamics at un-signalized intersections: a proactive approach for safety analysis. Accid. Anal. Prev. 134, 105316.

Khosravi, S., Beak, B., Head, K.L., Saleem, F., 2018. Assistive system to improve pedestrians' safety and mobility in a connected vehicle technology environment. Transp. Res. Rec. 2672 (19), 145–156.

Kingma, D.P., Ba, J., 2014. Adam: a Method for Stochastic Optimization. arXiv preprint arXiv:1412.6980.

Lee, C., Abdel-Aty, M., 2005. Comprehensive analysis of vehicle–pedestrian crashes at intersections in florida. Accid. Anal. Prev. 37 (4), 775–786.

Li, P., Abdel-Aty, M., Yuan, J., 2020. Real-time crash risk prediction on arterials based on lstm-cnn. Accid. Anal. Prev. 135, 105371.

Lin, J., Sun, M., 2018. A yolo-based traffic counting system. 2018 Conference on Technologies and Applications of Artificial Intelligence (TAAI) 82–85.

Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L., 2014. Microsoft coco: Common objects in context. In: European Conference on Computer Vision. Springer, pp. 740–755.

Manh, H., Alaghband, G., 2018. Scene-lstm: a Model for Human Trajectory Prediction. arXiv preprint arXiv:1808.04018.

Milan, A., Leal-Taixé, L., Reid, I., Roth, S., Schindler, K., 2016. Mot16: A Benchmark for Multi-object Tracking. arXiv preprint arXiv:1603.00831.

Mizoguchi, F., Yoshizawa, A., Iwasaki, H., 2017. Common-sense approach to avoiding near-miss incidents of pedestrians suddenly crossing narrow roads. 2017 IEEE 16th International Conference on Cognitive Informatics & Cognitive Computing (ICCI*CC) 335–340.

Montufar, J., Arango, J., Porter, M., Nakagawa, S., 2007. Pedestrians' normal walking speed and speed when crossing a street. Transp. Res. Rec. 2002 (1), 90–97.

Naphade, M., Tang, Z., Chang, M.-C., Anastasiu, D.C., Sharma, A., Chellappa, R., Wang, S., Chakraborty, P., Huang, T., Hwang, J.-N., 2019. The 2019 ai city challenge. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops 452–460.

Nhtsa, U.S.D.O.T., 2019a. Pedestrian Safety.

Nhtsa, U.S.D.O.T., 2019b. Traffic Safety Facts.

Ni, Y., Wang, M., Sun, J., Li, K., 2016. Evaluation of pedestrian safety at intersections: a theoretical framework based on pedestrian-vehicle interaction patterns. Accid. Anal. Prev. 96, 118–129.

Obeid, H., Abkarian, H., Abou-Zeid, M., Kaysi, I., 2017. Analyzing driver-pedestrian interaction in a mixed-street environment using a driving simulator. Accid. Anal. Prev. 108, 56–65.

Pedregosa, F., Ga, #235, Varoquaux, L., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., #201, Duchesnay, D., 2011. Scikit-learn: machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.

Qidian, L., Dasmehdix, Bensloane, Hyunmoahn, Chaplinhao, 2020. Real-time Multi-person Tracker Using Yolo v3 and deep_sort With Tensorflow. Github.

Radwan, E., Darius, B., Wu, J., Abou-Senna, H., 2016. Simulation of pedestrian safety surrogate measures. In: ARRB Conference, 27th, 2016. Melbourne, Victoria, Australia.

Redmon, J., Farhadi, A., 2018. Yolov3: An Incremental Improvement.

Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 779–788.

Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: towards real-time object detection with region proposal networks. In: Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1. MIT Press, Montreal, Canada, pp. 91–99.

Salamati, K., Schroeder, B., Rouphail, N.M., Cunningham, C., Long, R., Barlow, J., 2011. Development and implementation of conflict-based assessment of pedestrian safety to evaluate accessibility of complex intersections. Transp. Res. Rec. 2264 (1), 148–155.

Schneemann, F., Gohl, I., 2016. Analyzing driver-pedestrian interaction at crosswalks: a contribution to autonomous driving in urban environments. 2016 IEEE Intelligent Vehicles Symposium (IV) 38–43.

Španhel, J., Bartl, V., Juránek, R., Herout, A., 2019. Vehicle re-identification and multi-camera tracking in challenging city-scale environment. Proc. CVPR Workshops.

Szeliski, R., 2010. Computer Vision: Algorithms and Applications Springer Science & Business Media.

Tageldin, A., Zaki, M.H., Sayed, T., 2017. Examining pedestrian evasive actions as a potential indicator for traffic conflicts. IET Intell. Transp. Syst. 11 (5), 282–289.

Tang, Z., Naphade, M., Liu, M.-Y., Yang, X., Birchfield, S., Wang, S., Kumar, R., Anastasiu, D., Hwang, J.-N., 2019. Cityflow: a city-scale benchmark for multi-target multi-camera vehicle tracking and re-identification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition 8797–8806.

Tarko, A., A. Davis, G., Saunier, N., Sayed, T., 2009. Surrogate Measures of Safety.

Vogel, K., 2002. What characterizes a "free vehicle" in an urban area? Transp. Res. Part F Traffic Psychol. Behav. 5 (1), 15–29.

Wang, T., Cardone, G., Corradi, A., Torresani, L., T. Campbell, A., 2012. Walksafe: a Pedestrian Safety App for Mobile Phone Users Who Walk and Talk While Crossing Roads.
Wilson, T.B., Butler, W., Mcgehee, D.V., Dingus, T.A., 1997. Forward-looking collision warning system performance guidelines. SAE Trans. 701–725.
Wojke, N., Bewley, A., 2018. Deep Cosine Metric Learning for Person Re-identification.
Wojke, N., Bewley, A., Paulus, D., 2017. Simple online and realtime tracking with a deep association metric. 2017 IEEE International Conference on Image Processing (ICIP) 3645–3649.
Wu, Y., Abdel-Aty, M., Zheng, O., Cai, Q., Yue, L., 2019. Developing a crash warning system for the bike lane area at intersections with connected vehicle technology. Transp. Res. Rec., 0361198119840617
Yuan, J., Abdel-Aty, M., Gong, Y., Cai, Q., 2019. Real-time crash risk prediction using long short-term memory recurrent neural network. Transp. Res. Rec. 2673 (4), 314–326.
Yue, L., Abdel-Aty, M., Wu, Y., Zheng, O., Yuan, J., 2020. In-depth approach for identifying crash causation patterns and its implications for pedestrian crash prevention. J. Safety Res. 73, 119–132.
Zaki, M.H., Sayed, T., Tageldin, A., Hussein, M., 2013. Application of computer vision to diagnosis of pedestrian safety issues. Transp. Res. Rec. 2393 (1), 75–84.
Zhang, S., Abdel-Aty, M., Yuan, J., Li, P., 2020. Prediction of pedestrian crossing intentions at intersections based on long short-term memory recurrent neural network. Transp. Res. Rec., 0361198120912422