


Prediction of Pedestrian Crossing Intentions at Intersections Based on Long Short-Term Memory Recurrent Neural Network

Shile Zhang¹, Mohamed Abdel-Aty¹, Jinghui Yuan¹, and Pei Li¹

Transportation Research Record
2020, Vol. 2674(4) 57–65
© National Academy of Sciences:
Transportation Research Board 2020
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0361198120912422
journals.sagepub.com/home/trr


Abstract

Traffic violations of pedestrians at intersections are major causes of road crashes involving pedestrians, especially red-light crossing behaviors. To predict the pedestrians' red-light crossing intentions, video data from real traffic scenes are collected. Using detection and tracking techniques in computer vision, some pedestrians' characteristics, including location information, are generated. A long short-term memory neural network is established and trained to predict pedestrians' red-light crossing intentions. The experimental results show that the model has an accuracy rate of 91.6% based on internal testing at one signalized crosswalk. This model can be further implemented in the vehicle-to-infrastructure communication environment and prevent crashes because of the pedestrians' red-light crossing behaviors.

Pedestrian safety plays an important role in traffic safety. Each year, approximately 5,000 pedestrian deaths are caused by traffic accidents in the U.S.A. (1). Furthermore, pedestrians' violations at intersections, such as crossing outside of crosswalks and crossing during red-light periods, can be major causes of such accidents (2).

To investigate pedestrians' jaywalking behaviors, especially red-light crossing, behavior models such as the theory of planned behavior (TPB) model and statistical models were used (3–6). It was found that pedestrians' characteristics, such as age, gender, grouping behavior, pedestrian volume, and safety awareness were significant factors on pedestrians' red-light violations (7, 8). At signalized intersections, pedestrians' time of arrival, signal design, and types of land use at the crossing intersections, could influence pedestrians' waiting time, thus influencing pedestrians' crossing attempts (5, 9). The pedestrians' crossing intention involving red-light violation was highly time-dependent. The intention of violation increased when the pedestrians' waiting time lasted longer (7, 10).

With various built-in sensors, the advanced driver assistance system (ADAS) could alert drivers of pedestrians' unexpected dangerous crossing behaviors. Some studies used on-board sensors to estimate and predict pedestrians' paths (11, 12). Keller and Gavrila used Kalman Filter and its extension to predict whether the pedestrian would cross the street for a proactive pedestrians' safety system (11). The error of prediction result

was 10–50 cm at the prediction time horizon of 0–0.77 s. However, on-board sensors have limited field-of-view (FOV) (13). It could be too late for drivers to take evasive actions when the pedestrian was already in the FOV. Thus, more work should be done to capture and predict pedestrians' jaywalking intentions (14, 15).

As there are many irregularities in pedestrians' movements, video data could be employed to capture and analyze pedestrians' characteristics from a microscopic view. Video data were used in previous work to investigate pedestrians' crossing behaviors (2, 6, 7, 11, 15, 16). Zaki and Sayed investigated pedestrians' spatial traffic violations, that is, the pedestrian was not walking in the designated region, and temporal violations, that is, the pedestrian crossed during an improper traffic signal phase (2). It was found that pedestrians with illegal crossing behavior had higher velocity and caused more traffic conflicts. Ka et al. predicted pedestrians' crossing intentions during the red-light periods based on their characteristics (age, gender, head orientation, etc.) (15). Using Region-based Convolutional Neural Network (RCNN) object detection and Simple Online and Realtime

¹Department of Civil, Environmental & Construction Engineering, University of Central Florida, Orlando, FL

Corresponding Author:

Shile Zhang, shirleyzhang@knights.ucf.edu

Tracking (SORT) tracking technique, the system could identify potential critical events from videos caused by the pedestrians' red-light crossing behaviors, and alert right-turn vehicles (17).

With the development of deep learning, it could be used to better solve transportation problems. Traditional deep learning neural networks were less effective at capturing the relationships in sequential data for future predictions. Thus, recurrent neural network (RNN) was proposed to mitigate this defect by feeding back the output from a time window to the next time window in the same layer. But RNN had difficulty to connect the information when the time span between input and output units was long (18, 19). A particular implementation of the RNN was long short-term memory (LSTM) neural network model, which could capture long-term dependencies of time series data (20). In transportation fields, LSTM neural networks were used to predict vehicle travel time or traffic speed on highway links as well as urban arterials (21–23). They were also used for driving behavior classification and real-time crash risk prediction (24–26). Through these implementations, LSTM models proved their good performances on sequential traffic data. LSTM neural network brings the possibility to better predict pedestrians' movements such as trajectory predictions (27, 28). Besides, Alahi et al. used an LSTM neural network to predict pedestrians' movements based on their interaction between each other in crowded spaces (29). It is promising to predict pedestrians' unexpected crossings using LSTM neural network.

To address the above-mentioned research gaps, this study is intended to predict the pedestrians' crossing intentions at the signalized crosswalk based on an LSTM neural network model. With video data collected from real traffic scenes, it was found that pedestrians that crossed during the red-light were more in danger of being struck by a vehicle, from the perspective of surrogate safety measures (SSMs). Pedestrians' characteristics were generated using computer vision techniques. And the crossing intentions were labeled by observing their behaviors before they started to cross. The experiment results showed that the LSTM model achieves 91.6% accuracy based on internal testing at one signalized crosswalk. Drivers can get more prepared for jaywalking pedestrians near intersections if the model is to be implemented in the vehicle-to-infrastructure (V2I) communication system.

Data Collection

Study Site

To analyze pedestrian safety at intersections, a crosswalk with a relatively high volume of pedestrians at the University of Central Florida (UCF) was selected as the

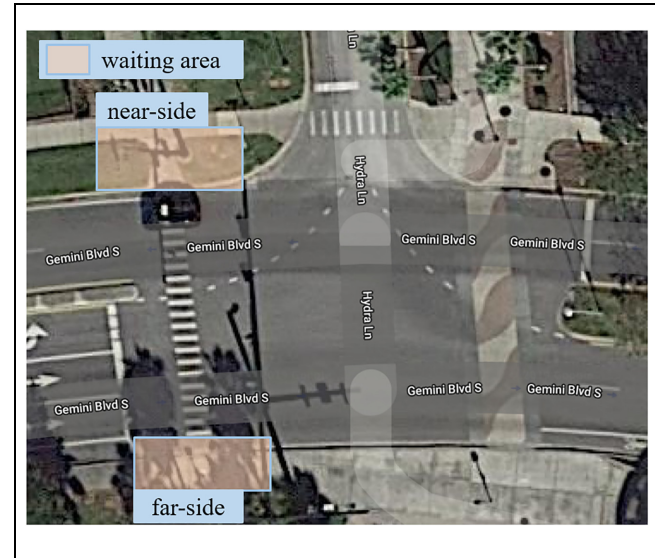


Figure 1. Spatial map of the studied site (30).

study site. This intersection was a key intersection with a total entering volume of about 200veh/h. Video data of 122 pedestrians were collected during the afternoon peak hours. The spatial map of the site is shown in Figure 1. The camera (GoPro HERO7) was set at the near-side of the studied crosswalk at the height of 6.56 ft. The data of 122 pedestrians who entered the two waiting areas during red-light periods were collected for further study.

Evaluation of the Pedestrian Safety at the Study Site

In this work, one SSM, post-encroachment time (PET) was used as an indicator of pedestrian safety (31, 32). It is defined as the time difference between the moment when the first road user left the potential collision area and the moment when the second user reached it. This indicator was good for describing “near miss” situations. As shown in Equation 1, t_2 and t_1 were the moments when the vehicle/pedestrian left/reached the same area accordingly. And the absolute value of the difference was the PET value. The PET threshold was set to be 6 s according to the literature to determine if there was a dangerous condition for the pedestrian (33, 34).

$$PET = |t_2 - t_1| \quad (1)$$

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

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

Traffic conflict data were manually collected to ensure analysis accuracy. Among 122 pedestrians, 43 pedestrians crossed during the red-light phases. Another 79

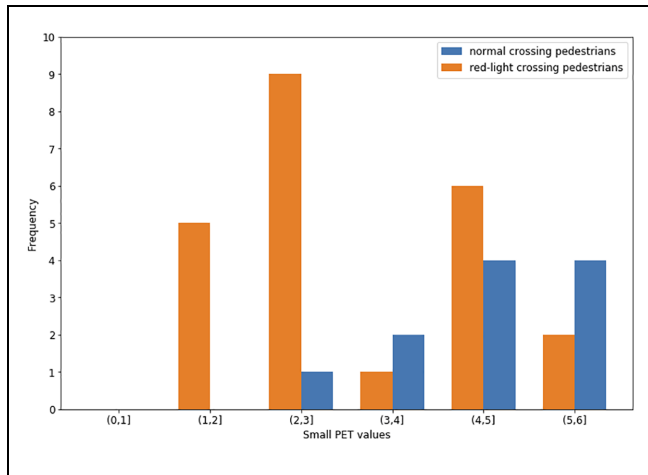


Figure 2. Frequency of small, post-encroachment times (0–6 s) of red-light crossing pedestrians and normal crossing pedestrians.

pedestrians waited for the pedestrian signal then crossed. However, among all the situations with small PET values (smaller than the threshold), 23 cases were caused by red-light crossing pedestrians, while only 11 cases were caused by normal crossing pedestrians, as shown in Figure 2. Thus, red-light crossing pedestrians could be more dangerous than other pedestrians at the intersection. This is consistent with the findings in the literature (2).

Video Processing

To extract moving trajectories of pedestrians, computer vision techniques including object detection, object tracking, and perspective transformation were used.

Object Detection. In this work, the object detection was done using YOLOv3 algorithm (35). Different from other approaches, it could apply a single neural network

to the full image, dividing and predicting anchor boxes at the same time. This was an effective detection technique running in near real-time.

Object Tracking. Multiple object tracking (MOT) techniques were widely used for following objects' movements. In this work, object tracking was done using Deep SORT algorithm (36, 37). Deep SORT was a tracker with good performance on the MOT16 Challenge benchmark, which was a standardized benchmark for evaluating the performances of different MOT algorithms (38).

Figure 3 shows a snapshot of the automated video processing procedure. For the studied crosswalk, the YOLOv3 algorithm first detected and classified objects as pedestrians (shown as blue bounding boxes). Then Deep SORT algorithm could take the detection boxes as initial input, and track the movements of each object (shown as white bounding boxes). With a unique ID (shown as green number) assigned to each tracked pedestrian, movements of the pedestrians were extracted and further analyzed.

Perspective Transformation. The world coordinate system has the relationship with image coordinate system as shown in Equation 2. The world coordinate (X_i, Y_j) could be mapped to the image coordinate (u_i, v_j) through the matrix h . h matrix was composed of nine values, from h_1 to h_9 . To obtain h matrix, linear least-squares method was used; the formula is shown in Equation 3. Ten correspondences of points in the traffic scene were extracted from images and Google Maps using the OpenCV package (39). Each of the point correspondence could form two rows of matrix A . Singular value decomposition (SVD) could be employed to obtain solution of this formula (40–42). After getting matrix h , the image coordinates could be converted to the world coordinates through the inverse matrix of h .

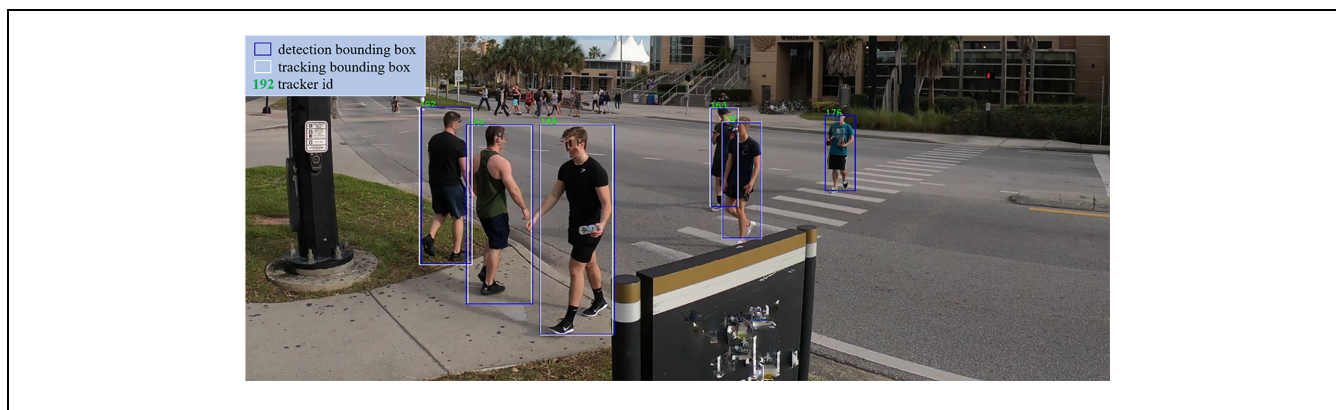


Figure 3. Pedestrian detection and tracking at the studied crosswalk.

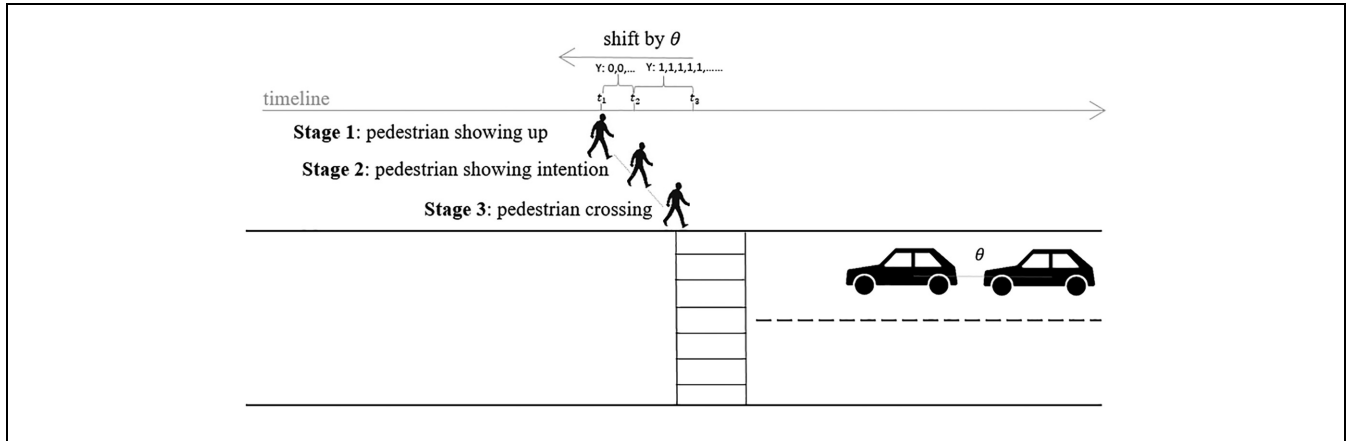


Figure 4. Pedestrian's crossing model.

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = h \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} \quad (2)$$

$$A * h = \begin{bmatrix} 0 & 0 & 0 & -X_1 & -Y_1 & 1 & v_1 X_1 & v_1 Y_1 & v_1 \\ X_1 & Y_1 & 1 & 0 & 0 & 0 & -u_1 X_1 & -u_1 Y_1 & -u_1 \\ 0 & 0 & 0 & -X_2 & -Y_2 & 1 & v_2 X_2 & v_2 Y_2 & v_2 \\ X_2 & Y_2 & 1 & 0 & 0 & 0 & -u_2 X_2 & -u_2 Y_2 & -u_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix}$$

$$\begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \\ h_7 \\ h_8 \\ h_9 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (3)$$

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

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

$$h = (h_1, h_2, h_3, h_4, h_5, h_6, h_7, h_8, h_9)^T, h = 1$$

Through the video processing process, features such as pedestrian locations were generated as input variables at a frequency of 10 Hz (environment: NVIDIA GTX 1080Ti 11G GPU).

Pedestrian Crossing Intention Labeling

After the steps above, a time series dataset was generated from pedestrians' trajectories. To label the dependent variable Y , denoting whether the pedestrian would cross during red-light, a pedestrian crossing model describing pedestrians' behaviors in the studied areas (Figure 1) was established as shown in Figure 4. Pedestrians' behaviors

could be divided into three stages, 1) pedestrian showing up, 2) pedestrian showing crossing intention (such as turning heads to watch for vehicles), 3) pedestrian starting to cross, which were observed frame by frame by one of the authors. Thus, for jaywalking pedestrians, from t_2 to t_3 , Y was labeled as positive ("1"). This time interval was when the pedestrian was observing surrounding areas and starting to cross, that is, showing crossing intentions.

For the prediction purpose, suppose the driver would take evasive actions after capturing the pedestrian's crossing intentions after reaction time θ . In this paper, the reaction time θ was taken as 1.5 s according to the literature (43, 44). Labels were shifted ahead θ units by timestamp of each pedestrian for prediction purpose, as shown in Figure 4. Thus, the pedestrian's red-light crossing intention is being predicted 1.5 s ahead.

With the crossing intention of the pedestrian during the red-light period as the dependent variable Y , the input variables used in this study included: pedestrian's gender, pedestrian's walking direction, whether the pedestrian was walking in a group, and pedestrians' locations. The walking directions of pedestrians were denoted by 1 ("towards near-sided crosswalk") and 0 ("towards far-sided crosswalk"), which can be found in Figure 1. However, pedestrians' features, such as gender, whether the pedestrian was walking in the group (yes/no), and pedestrians' walking directions were not easy to generate automatically from video. These features were labeled manually. It is believed that in the near future these features could also be detected using computer vision techniques. The summary of the descriptive statistics of all variables is shown in Table 1.

Methodology

Based on the above discussion, a methodology of pedestrian crossing intention prediction based on the pedestrians' characteristics at the signalized crosswalk is carried out using an LSTM neural network (20). As

Table 1. Summary of Variable Descriptive Statistics

| Variables | Description | Details |
|---|--|--------------------|
| Gender ^a | Gender of pedestrian | “male” or “female” |
| Direction ^a | Pedestrian’s walking direction | “1” or “0” |
| Grouping ^a | Whether the pedestrian is walking in a group | “yes” or “no” |
| Locations ^b | Pedestrian’s location | (X, Y) |
| Crossing intention (dependent variable) | Whether the pedestrian will cross during red-light | “1” or “0” |

^aFeatures manually labeled.

^bFeatures generated from video.

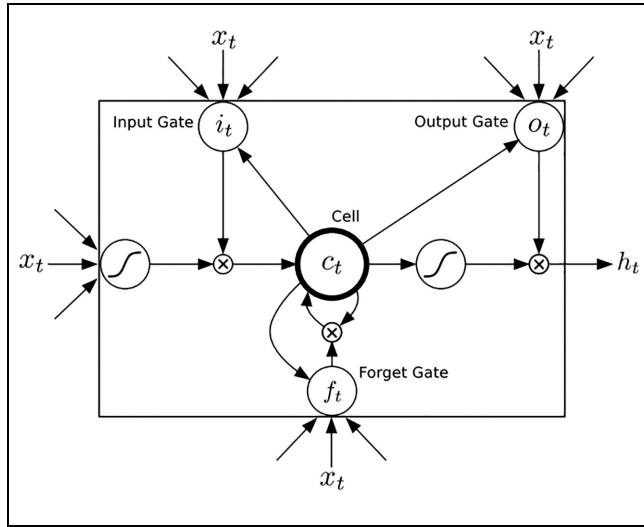


Figure 5. Schematic of long short-term memory neural network (45).

pedestrians’ trajectories are time-series data, LSTM neural network model can better capture the temporal dependencies of them.

LSTM neural network is one kind of RNN. In general, the mathematical formulations of the unit of RNN are shown in Equation 4 and Equation 5. At time T , with \mathbf{x} as the input vector, W as weight matrices, b_y as bias term, the RNN computes the hidden vector \mathbf{h}_T and the output vector \mathbf{y}_T . Then, \mathbf{y}_T is fed into the next layer of the RNN.

$$\mathbf{h}_T = \sigma(W_1 \mathbf{x}_T + W_{ih} \mathbf{h}_{T-1} + W_{hh} c_{T-1} + b_y) \quad (4)$$

$$\mathbf{y}_T = W_{hy} \mathbf{h}_T + b_y \quad (5)$$

$\mathbf{x} = (x_1, x_2, \dots, x_T)$: input vector

$\mathbf{h} = (h_1, h_2, \dots, h_T)$: hidden vector

$\mathbf{y} = (y_1, y_2, \dots, y_T)$: output vector

W_i : weight matrix

b_y : bias term

However, RNNs are not capable of learning long-term dependencies from time-series data (18). To overcome it, LSTM neural network is proposed with a purpose-built memory unit to store information (45). As shown in Figure 5, a single unit from the hidden layer of the LSTM neural network is composed of an input gate i_t , a forget gate f_t , an output gate O_t . These three gates control information flow in each unit of the neural network. C_t is the memory cell, and h_t is the hidden layer output. Given the number of time windows T , the input vector $\mathbf{x} = (x_1, x_2, \dots, x_T)$ is computed by Equations 6–11 to generate the output y_t , which is a vector of probabilities, iterated from $t = 1$ to T .

$$i_t = \sigma(W_{xi} x_t + W_{ih} h_{t-1} + W_{ic} c_{t-1} + b_i) \quad (6)$$

$$f_t = \sigma(W_{fx} x_t + W_{hf} h_{t-1} + W_{fc} c_{t-1} + b_f) \quad (7)$$

$$o_t = \sigma(W_{xo} x_t + W_{ho} h_{t-1} + W_{co} c_t + b_o) \quad (8)$$

$$c_t = f_t \cdot c_{t-1} + i_t \cdot \tanh(W_{cx} x_t + W_{hc} h_{t-1} + b_c) \quad (9)$$

$$h_t = o_t \cdot \tanh(c_t) \quad (10)$$

$$y_t = W_{hy} h_{t-1} + b_y \quad (11)$$

σ : sigmoid function

\cdot : elementwise product of the vectors

The model architecture used in the study is illustrated in Figure 6. The features from three time-slices are stacked as input to predict the result of the next time slice. The model contains one input layer, one stacked-LSTM layer, a dense (fully connected) layer, and an output neuron denoting the classification result. Besides, the dropout layer is added to prevent over-fitting. The Sigmoid function is used as the activation function to generate the output. Adam function is used as the optimization function (46). The model is implemented in Keras framework (47).

Experiments and Results

Based on video collected at the studied site, the pedestrian trajectory data are split into training set (75%) and

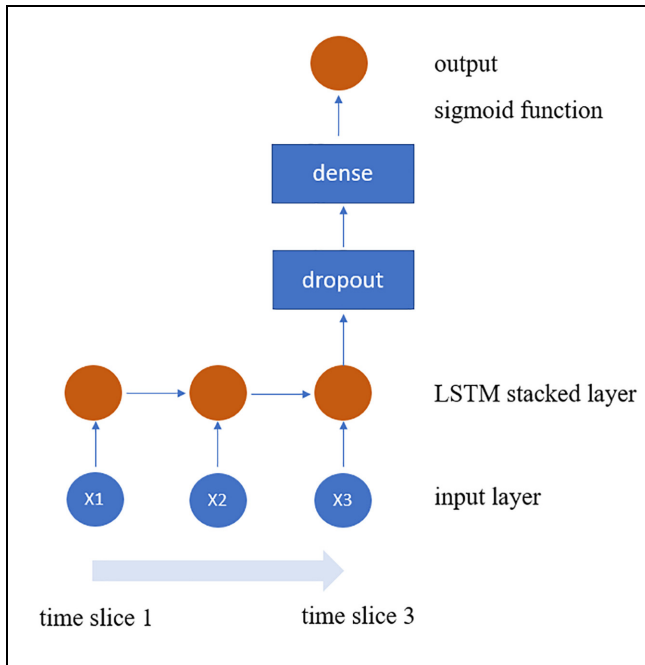


Figure 6. Model architecture.
Note: LSTM = long short-term memory.

test set (25%) for training and validating the performance of the proposed LSTM model.

As stated above, among all the 122 pedestrians, 43 pedestrians made red-light crossing behavior. The crossing intention labeling procedure generates a small number of positive samples with the dependent variable Y equals 1. For instance, for the training dataset, there are totally 1,939 positive samples and 58,882 negative samples. The ratio is around 1:30, indicating the data are highly imbalanced. As an imbalanced dataset will bring difficulty to train a model with good performance, synthetic minority over-sampling technique (SMOTE) is used to increase the number of positive samples (48). SMOTE is a popular over-sampling method, which can generate new minority class instances by interpolating between several minority class examples that lie together. It should be noted that SMOTE is only employed on the training dataset, while the test dataset is the original data with 661 positive samples and 19,613 negative samples.

After tuning different combinations of hyperparameters, the hyperparameter values are selected as below: learning rate is 0.0005, the batch size is 1,000, and the epoch number is 5. Quantitative variables, which are pedestrians' coordinates, (X, Y) , are normalized to feed into the model.

To evaluate the experiment results, the diagram for metrics calculation used is shown in Table 2. "Positive" denotes that pedestrian conducts a red-light crossing behavior. "Negative" denotes that pedestrian doesn't have red-light crossing behavior. Metrics such as

Table 2. Diagram for Metrics Calculation

| Prediction result | Ground truth | |
|-------------------|---------------------|---------------------|
| | Positive | Negative |
| Positive | True positive (TP) | False positive (FP) |
| Negative | False negative (FN) | True negative (TN) |

sensitivity, specificity, and accuracy are calculated as shown in Equations 12–14. Sensitivity measures show how good the model is among all the positives, that is, the proportion of actual positives that are correctly identified by the model. Specificity measures the proportion of actual negatives that are correctly identified by the model. Accuracy value measures the proportion of true positives and negatives in all detected results.

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (12)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (13)$$

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (14)$$

The receiver operating characteristic (ROC) curve is used as a comprehensive metric to evaluate the model's performance. This curve plots two parameters, true positive rate (sensitivity) and false positive rate (1-specificity), as shown in Equation 15 at different classification thresholds. Area under the ROC curve (AUC) value is used as an indicator of accuracy.

$$\text{False Positive Rate} = \frac{FP}{FP + TN} \quad (15)$$

The prediction results of the model are listed in Table 3. The sensitivity value is 92.4%, and the false positive rate is 20.3%. Overall, the model shows a prediction accuracy of 91.6% at the studied crosswalk.

Conclusion and Discussion

This paper uses video data to predict pedestrians' red-light crossing intentions at signalized intersections with a stacked LSTM neural network. As pedestrians' red-light crossing behaviors are one of the causal factors of traffic conflicts, with real traffic data collected at the studied site, pedestrians' location features are generated using automated video analysis. Other features such as gender, walking direction, and grouping behavior, which are important factors influencing pedestrian's traffic violation intentions, are also used to feed into the LSTM model. And the red-light crossing intentions of pedestrians are labeled after analyzing the interaction between

Table 3. Prediction Results

| Training set | Test set | | | |
|--------------|-------------|---------------------|----------|-------|
| | Sensitivity | False positive rate | Accuracy | AUC |
| AUC | | | | |
| 0.943 | 0.924 | 0.203 | 0.916 | 0.938 |

Note: AUC = area under the ROC curve; ROC = receiver operating characteristic.

red-light crossing pedestrians and vehicles. An LSTM model is proposed to predict pedestrians' crossing intentions at 1.5 s ahead. The experiment result shows that the model reaches an accuracy of 91.6% based on internal testing at one signalized crosswalk.

Compared with previous studies, this work analyzes the pedestrians' unexpected crossing intention from a more microscopic view, instead of treating it as binary outcome. The dependent variables, red-light crossing intentions, are labeled using trajectory data generated from videos. And the LSTM neural network with stacked layer proved to perform well on the dataset.

However, there are still some improvements to be made. The proposed model has a relatively high false positive rate, which means it is more likely to treat normal-walking pedestrians as jaywalking pedestrians. More features related to the pedestrian's mobility information such as walking speed and acceleration should be extracted as input for the model to overcome this problem.

The proposed model can be further implemented at intersections to alert drivers of pedestrians with unexpected crossing behaviors, thus preventing collisions between pedestrians and vehicles.

Author Contributions

The authors confirm contribution to the paper as follows: study conception and design: S. Zhang, M. Abdel-Aty, J. Yuan; data collection: S. Zhang, P. Li; analysis and interpretation of results: S. Zhang, M. Abdel-Aty, J. Yuan; draft manuscript preparation: S. Zhang, M. Abdel-Aty, J. Yuan, P. Li. All authors reviewed the results and approved the final version of the manuscript.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study was sponsored by the Florida Department of Transportation (FDOT).

References

1. FHWA, U.S. Department of Transportation. *Pedestrian & Bicycle Safety*. https://safety.fhwa.dot.gov/ped_bike/. Accessed June 16, 2019.
2. Zaki, M. H., and T. Sayed. Automated Analysis of Pedestrians' Nonconforming Behavior and Data Collection at an Urban Crossing. *Transportation Research Record: Journal of the Transportation Research Board*, 2014. 2443: 123–133.
3. Ajzen, I. The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes*, Vol. 50, No. 2, 1991, pp. 179–211.
4. Evans, D., and P. Norman. Understanding Pedestrians' Road Crossing Decisions: An Application of the Theory of Planned Behaviour. *Health Education Research*, Vol. 13, No. 4, 1998, pp. 481–489.
5. Brosseau, M., S. Zangenehpour, N. Saunier, and L. Miranda-Moreno. The Impact of Waiting Time and Other Factors on Dangerous Pedestrian Crossings and Violations at Signalized Intersections: A Case Study in Montreal. *Transportation Research Part F: Traffic Psychology and Behaviour*, Vol. 21, 2013, pp. 159–172.
6. Hashimoto, Y., G. Yanlei, L.-T. Hsu, and K. Shunsuke. A Probabilistic Model for the Estimation of Pedestrian Crossing Behavior at Signalized Intersections. *Proc., 2015 IEEE 18th International Conference on Intelligent Transportation Systems*, Las Palmas, 2015, pp. 1520–1526.
7. Guo, H., Z. Gao, X. Yang, and X. Jiang. Modeling Pedestrian Violation Behavior at Signalized Crosswalks in China: A Hazards-Based Duration Approach. *Traffic Injury Prevention*, Vol. 12, No. 1, 2011, pp. 96–103.
8. Hamed, M. M. Analysis of Pedestrians' Behavior at Pedestrian Crossings. *Safety Science*, Vol. 38, No. 1, 2001, pp. 63–82.
9. Cinnamon, J., N. Schuurman, and S. M. Hameed. Pedestrian Injury and Human Behaviour: Observing Road-Rule Violations at High-Incident Intersections. *PLoS One*, Vol. 6, No. 6, 2011, p. e21063.
10. Keegan, O., and M. O'Mahony. Modifying Pedestrian Behaviour. *Transportation Research Part A: Policy and Practice*, Vol. 37, No. 10, 2003, pp. 889–901.
11. Keller, C. G., and D. M. Gavrilu. Will the Pedestrian Cross? A Study on Pedestrian Path Prediction. *IEEE Transactions on Intelligent Transportation Systems*, Vol. 15, No. 2, 2014, pp. 494–506.
12. Ellis, D., E. Sommerlade, and I. Reid. Modelling Pedestrian Trajectory Patterns with Gaussian Processes. *Proc., 2009 IEEE 12th International Conference on Computer*

- Vision Workshops, ICCV Workshops*, Kyoto, Japan, IEEE, New York, 2009, pp. 1229–1234.
13. Basma, F., Y. Tachwali, and H. H. Refai. Intersection Collision Avoidance System using Infrastructure Communication. *Proc., 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, Washington, D.C., IEEE, New York, 2011, pp. 422–427.
 14. Kotte, J., C. Schmeichel, A. Zlocki, H. Gathmann, and L. Eckstein. Concept of an Enhanced V2X Pedestrian Collision Avoidance System with a Cost Function-Based Pedestrian Model. *Traffic Injury Prevention*, Vol. 18, 2017, pp. S37–S43.
 15. Ka, D., D. Lee, S. Kim, and H. Yeo. Study On the Framework of Intersection Pedestrian Collision Warning System Considering Pedestrian Characteristics. *Transportation Research Record: Journal of the Transportation Research Board*, 2019. 2673(5): 747–758.
 16. Ismail, K., T. Sayed, and N. Saunier. Automated Analysis of Pedestrian–Vehicle Conflicts: Context for Before-and-After Studies. *Transportation Research Record: Journal of the Transportation Research Board*, 2010. 2198: 52–64.
 17. Girshick, R., J. Donahue, T. Darrell, and J. Malik. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. *Proc., 2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, Ohio, IEEE, New York, 2014.
 18. Bengio, Y., P. Simard, and P. Frasconi. Learning Long-Term Dependencies with Gradient Descent is Difficult. *IEEE Transactions on Neural Networks*, Vol. 5, No. 2, 1994, pp. 157–166.
 19. Pascanu, R., T. Mikolov, and Y. Bengio. On the Difficulty of Training Recurrent Neural Networks. *Proc., International Conference on Machine Learning*, Atlanta, GA, 2013. pp. 1310–1318.
 20. Hochreiter, S., and J. Schmidhuber. Long Short-Term Memory. *Neural Computation*, Vol. 9, No. 8, 1997, pp. 1735–1780.
 21. Yanjie, D., L. Yisheng, and W. Fei-Yue. Travel Time Prediction with LSTM Neural Network. *Proc., 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, Rio de Janeiro, Brazil, 2016, pp. 1053–1058.
 22. Alché, F., and A. de L. Fortelle. An LSTM Network for Highway Trajectory Prediction. *Proc., 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, Yokohama, 2017. pp. 353–359.
 23. Ma, X., Z. Tao, Y. Wang, H. Yu, and Y. Wang. Long Short-Term Memory Neural Network for Traffic Speed Prediction using Remote Microwave Sensor Data. *Transportation Research Part C: Emerging Technologies*, Vol. 54, 2015, pp. 187–197.
 24. Saleh, K., M. Hossny, and S. Nahavandi. Driving Behavior Classification Based on Sensor Data Fusion using LSTM Recurrent Neural Networks. *Proc., 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC)*, Yokohama, Japan, IEEE, 2017, pp. 1–6.
 25. Li, P., M. Abdel-Aty, and J. Yuan. Real-Time Crash Risk Prediction on Arterials Based on LSTM-CNN. *Accident Analysis & Prevention*, Vol. 135, 2020, p. 105371.
 26. Yuan, J., M. Abdel-Aty, Y. Gong, and Q. Cai. Real-Time Crash Risk Prediction using Long Short-Term Memory Recurrent Neural Network. *Transportation Research Record: Journal of the Transportation Research Board*, 2019. 2673: 314–326.
 27. Xue, H., D. Q. Huynh, and M. Reynolds. SS-LSTM: A Hierarchical LSTM Model for Pedestrian Trajectory Prediction. *Proc., 2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Tahoe, NV, 2018. pp. 1186–1194.
 28. Manh, H., and G. Alaghband. Scene-LSTM: A Model for Human Trajectory Prediction. *arXiv preprint arXiv:1808.04018*, 2018.
 29. Alahi, A., K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese. Social LSTM: Human Trajectory Prediction in Crowded Spaces. *Proc., IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 2016, pp. 961–971.
 30. Google. University of Central Florida. <https://www.google.com/maps/@28.5965396,-81.1993853,87m/data=!3m1!1e3>. Accessed July 19, 2019.
 31. Tarko, A., G. A. Davis, N. Saunier, T. Sayed, and S. Washington. Surrogate Measures of Safety. *Safe Mobility: Challenges, Methodology and Solutions*. April 28, 2009. pp. 383–405.
 32. Allen, B. L., B. T. Shin, and P. J. Cooper. Analysis of Traffic Conflicts and Collisions. *Transportation Research Record: Journal of the Transportation Research Board*, 1978. 667: 67–74.
 33. Radwan, E., B. Darius, J. Wu, and H. Abou-Senna. Simulation of Pedestrian Safety Surrogate Measures. *Proc., ARRB Conference*, Melbourne, Victoria, Australia, 2016.
 34. Formosa, N., M. Quddus, S. Ison, M. Abdel-Aty, and J. Yuan. Predicting Real-Time Traffic Conflicts using Deep Learning. *Accident Analysis & Prevention*, Vol. 136, 2020, p. 105429.
 35. Redmon, J., and A. Farhadi. YOLOv3: An Incremental Improvement. *arXiv preprint arXiv:1804.02767*, 2018.
 36. Wojke, N., A. Bewley, and D. Paulus. Simple Online and Realtime Tracking with a Deep Association Metric. *Proc., 2017 IEEE International Conference on Image Processing (ICIP)*, Beijing, 2017. pp. 3645–3649.
 37. Wojke, N., and A. Bewley. Deep Cosine Metric Learning for Person Re-Identification. *Proc., 2018 IEEE Winter Conference on Applications of Computer Vision*, Lake Tahoe, 2018, pp. 748–756.
 38. Milan, A., L. Leal-Taixé, I. Reid, S. Roth, and K. Schindler. MOT16: A Benchmark for Multi-Object Tracking. *arXiv preprint arXiv:1603.00831*, 2016.
 39. Bradski, G. *OpenCV (Open Source Computer Vision Library)*. <https://opencv.org/>. Accessed July 15, 2019.
 40. Naphade, M., Z. Tang, M.-C. Chang, D. C. Anastasiu, A. Sharma, R. Chellappa, S. Wang, P. Chakraborty, T. Huang, J.-N. Hwang, and S. Lyu. The 2019 AI City Challenge. *Proc., IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Long Beach, 2019, pp. 452–460.
 41. Španhel, J., V. Bartl, R. Juránek, and A. Herout. Vehicle Re-Identification and Multi-Camera Tracking in Challenging

- City-Scale Environment. *Proc., CVPR Workshops*, Long Beach, 2019.
42. Tang, Z., M. Naphade, M.-Y. Liu, X. Yang, S. Birchfield, S. Wang, R. Kumar, D. Anastasiu, and J.-N. Hwang. City-flow: A City-Scale Benchmark for Multi-Target Multi-Camera Vehicle Tracking and Re-Identification. *Proc., IEEE Conference on Computer Vision and Pattern Recognition*, Long Beach, 2019, pp. 8797–8806.
 43. Wilson, T. B., W. Butler, D. V. McGehee, and T. A. Dingus. *Forward-Looking Collision Warning System Performance Guidelines*. SAE Technical Paper 970456, 1997, pp. 701–725.
 44. Rahman, M. H., M. Abdel-Aty, J. Lee, and M. S. Rahman. Enhancing Traffic Safety at School Zones by Operation and Engineering Countermeasures: A Microscopic Simulation Approach. *Simulation Modelling Practice and Theory*, Vol. 94, 2019, pp. 334–348.
 45. Graves, A., A.-r. Mohamed, and G. Hinton. Speech Recognition with Deep Recurrent Neural Networks. *Proc., 2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, Vancouver, BC, Canada, IEEE, New York, 2013, pp. 6645–6649.
 46. Kingma, D. P., and J. Ba. Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980*, 2014.
 47. Chollet, F. *Keras*. GitHub. <https://github.com/fchollet/keras>. Accessed July 18, 2019.
 48. Chawla, N. V., K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. SMOTE: Synthetic Minority Over-Sampling Technique. *International Journal of Artificial Intelligence Research*, Vol. 16, No. 1, 2002, pp. 321–357.

All results and opinions are those of the authors only and do not reflect the opinion or position of FDOT.