

Electronic Theses and Dissertations, 2020-

2020

Prediction of Pedestrians' Red Light Violations Using Deep Learning

Shile Zhang
University of Central Florida

 Part of the [Civil Engineering Commons](#), and the [Transportation Engineering Commons](#)
Find similar works at: <https://stars.library.ucf.edu/etd2020>
University of Central Florida Libraries <http://library.ucf.edu>

This Masters Thesis (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Electronic Theses and Dissertations, 2020- by an authorized administrator of STARS. For more information, please contact STARS@ucf.edu.

STARS Citation

Zhang, Shile, "Prediction of Pedestrians' Red Light Violations Using Deep Learning" (2020). *Electronic Theses and Dissertations, 2020-*. 160.
<https://stars.library.ucf.edu/etd2020/160>

PREDICTION OF PEDESTRIANS' RED-LIGHT VIOLATIONS USING DEEP LEARNING

by

SHILE ZHANG

B.Sc. Southeast University, 2016

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science
in the Department of Civil, Environmental and Construction Engineering
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Spring Term

2020

Major Professor: Mohamed Abdel-Aty

© 2020 Shile Zhang

ABSTRACT

Pedestrians are regarded as Vulnerable Road Users (VRUs). Each year, thousands of pedestrians' deaths are caused by traffic crashes, which take up 16% of the total road fatalities and injuries in the U.S. (FHWA, 2018). Crashes can happen if there are interactions between VRUs and motorized transportation. And pedestrians' unexpected crossings, such as red-light violations at the signalized intersections, would expose them to motorized transportation and cause potential collisions.

This thesis is intended to predict the pedestrians' red-light violation behaviors at the signalized crosswalks based on an LSTM (Long Short-term Memory) neural network. With video data collected from real traffic scenes, it is found that pedestrians that crossed during the red-light periods are more in danger of being struck by vehicles, from the perspective of Surrogate Safety Measures (SSMs). Pedestrians' features are generated using computer vision techniques. An LSTM model is used to predict pedestrians' red-light violations using these features. The experiment results at one signalized intersection show that the LSTM model achieves the accuracy of 91.6%. Drivers can be more prepared for these unexpected crossing pedestrians if the model is to be implemented in the vehicle-to-infrastructure (V2I) communication system.

Keywords: pedestrian safety, red-light violations, deep learning

ACKNOWLEDGMENTS

I would like to convey my heartiest gratitude to my honorable supervisor Dr. Mohamed Abdel-Aty for his excellent supervision and constant support in this thesis. I would also like to acknowledge the support and encouragement from my family and friends.

TABLE OF CONTENTS

LIST OF FIGURES	VII
LIST OF TABLES	VIII
LIST OF ACRONYMS/ABBREVIATIONS	IX
CHAPTER 1 INTRODUCTION	- 1 -
1.1 Introduction	- 1 -
1.2 Thesis Contributions	- 1 -
1.3 The Objectives of the Thesis	- 2 -
1.4 Thesis Organization	- 2 -
CHAPTER 2 LITERATURE REVIEW	- 3 -
2.1 Pedestrians' Violation Behaviors	- 3 -
2.2 Video Data Applications in Transportation	- 4 -
2.3 LSTM Neural Network	- 6 -
CHAPTER 3 DATA COLLECTION	- 8 -
3.1 Experiment setup	- 8 -
3.2 Evaluation of the Pedestrian Safety at the Study Site	- 8 -
3.3 Video Processing	- 10 -
3.3.1 Object Detection	- 11 -
3.3.2 Object Tracking	- 14 -
3.3.3 Perspective Transformation	- 17 -
3.3.4 Other Features	- 18 -
CHAPTER 4 METHODOLOGY	- 19 -
4.1 Pedestrian Crossing Modeling	- 19 -
4.2 LSTM Neural Network Model	- 20 -
4.2.1 LSTM Model	- 20 -
4.2.2 Model Architecture	- 22 -
4.3 Experiments and Results	- 23 -
4.3.1 Data Set Overview	- 23 -
4.3.2 Oversampling	- 24 -

4.3.3 Experiment Results	- 25 -
CHAPTER 5 CONCLUSIONS	- 27 -
REFERENCES	- 28 -

LIST OF FIGURES

Figure 1 Spatial map of the studied site (Maps, 2020)	- 8 -
Figure 2 Frequency of small PETs (0-6s) of red-light crossing pedestrians & normal crossing pedestrians.....	- 10 -
Figure 3 YOLO detection model (Redmon et al., 2016)	- 12 -
Figure 4 Snapshot of automated video processing	- 16 -
Figure 5 Pedestrians' trajectories at the studied site.....	- 16 -
Figure 6 Perspective transformation	- 18 -
Figure 7 Pedestrian's crossing model	- 20 -
Figure 8 Schematic of LSTM unit (adapted from (A. Graves, A. Mohamed, & G. Hinton, 2013; Kang, Lv, & Chen, 2017))	- 21 -
Figure 9 Model architecture.....	- 23 -

LIST OF TABLES

Table 1 Evaluation of the Deep SORT tracking model (N. Wojke et al., 2017).	- 15 -
Table 2 Summary of variable descriptive statistics	- 24 -
Table 3 Diagram for metrics calculation	- 25 -
Table 4 Prediction results.....	- 26 -

LIST OF ACRONYMS/ABBREVIATIONS

ADAS: Advanced Driver Assistance System

AUC: Area under curve

CNN: Convolutional Neural Network

DST: Deceleration-to-safety time

FPR: False Positive Rate

GT: Gap time

LSTM: Long-short term memory

MOT: Multiple Object Tracking

PET: Post-encroachment-time

ROC: Receiver operating characteristics

SMOTE: Synthetic minority over-sampling technique

SSMs: Surrogate Safety Measures

TTC: Time-to-collision

VRU: Vulnerable Road Users

V2I: Vehicle-To-Infrastructure

CHAPTER 1 INTRODUCTION

1.1 Introduction

Pedestrian safety plays an important role in traffic safety. In 2018, there was 3% increase in the pedestrian deaths, ranked the highest since 1990 (NHTSA, 2019).

Intersections usually have relatively high pedestrian volumes. However, the pedestrians' violation behaviors, especially red-light violations at the signalized intersections, can expose pedestrians to motorized transportations and cause potential crashes.

With the development of Connected Vehicle (CV) technologies, it is possible to warn drivers of these unexpected pedestrians. This thesis predicted pedestrians' red-light violations at the signalized intersection using video data. The proposed LSTM neural network (Long-short term memory neural network) achieves an accuracy of 91.6% at one crosswalk.

1.2 Thesis Contributions

This thesis has made a few contributions to pedestrian safety:

- Traditional studies treated red-light crossings as binary outcomes. This thesis analyzes pedestrians' crossing intentions and labels pedestrians' red-light violations in time series.
- Traditional models are less effective in capturing the sequential information lying in the time series data. The LSTM neural network has special architecture that can better learn features from the previous time windows.

1.3 The Objectives of the Thesis

The objectives of this thesis are:

- To analyze pedestrians' crossing behaviors at the signalized intersections,
- To investigate factors influencing pedestrians' violation behaviors,
- To predict pedestrians' red-light violations using the LSTM neural network.

1.4 Thesis Organization

The rest of the thesis is organized as follows: Chapter 2 provides the literature review. Chapter 3 provides the procedures of data collection. Chapter 4 describes the proposed methodologies to predict pedestrians' red-light violations, and the experimental results. Chapter 5 describes the summary based on the results obtained in the previous chapters.

CHAPTER 2 LITERATURE REVIEW

Pedestrian safety plays an important role in traffic safety. Compared with vehicles, pedestrians are not protected by metal shells. They are regarded as Vulnerable Road Users (VRUs). Each year, approximately 5,000 pedestrian deaths are caused by traffic crashes in the U.S. (FHWA, 2018).

2.1 Pedestrians' Violation Behaviors

Among all the crashes between vehicles and pedestrians, pedestrians' violation behaviors can be one of the causes. At the signalized intersections, normally, pedestrians are sequentially separated from vehicles because of traffic signals. However, pedestrians' violation behaviors especially red-light violations, will expose them to vehicles and cause potential crashes. Besides, pedestrians' abrupt movements, such as suddenly walking out of road curvature, can make it hard for drivers to take evasive actions (Yue, Abdel-Aty, Wu, Zheng, and Yuan (2020)).

Some previous studies investigated the relationships between pedestrians' characteristics with their red-light violations. Behavior models such as the theory of planned behavior (TPB) model (Ajzen, 1991; Evans & Norman, 1998), and some statistical models such as discrete choice models (Brosseau, Zangenehpour, Saunier, & Miranda-Moreno, 2013; Hashimoto, Yanlei, Hsu, & Shunsuke, 2015) were used. Hamed (2001) investigated the factors influencing pedestrians' waiting time and crossing attempts. It was found that pedestrians' characteristics, such as age, gender, number of people in groups, were significant

factors. Pedestrians' volume, pedestrians' time of arrival, and safety awareness were also significant factors for pedestrians' red-light violations (Brosseau et al., 2013; Guo, Gao, Yang, & Jiang, 2011; Hamed, 2001).

Nevertheless, some studies investigated factors of external factors such as road geometry design (Gitelman, Balasha, Carmel, Hendel, & Pesahov, 2012). At the signalized intersections, types of land use (Cinnamon, Schuurman, & Hameed, 2011), and signal design (Brosseau et al., 2013) could influence pedestrians' waiting time, thus influencing pedestrians' crossing attempts. Traffic conditions (Guo et al., 2011), numbers of the central refuges (Hamed, 2001) were also significant factors. It was found that pedestrians' red-light violations were highly time dependent. The pedestrians' violation intentions increased when the pedestrians' waiting time elapsed longer (Guo et al., 2011; Keegan & O'Mahony, 2003). Thus, it is necessary to analyze pedestrians' red-light violations in time series, to make the prediction more effective.

2.2 Video Data Applications in Transportation

Video data could be used to analyze the behaviors of road users. As there were high irregularities in the pedestrians' movements, video data could be employed to capture pedestrians' characteristics from a more microscopic view. Video data were used in previous work (Formosa, Quddus, Ison, Abdel-Aty, & Yuan, 2020; Fu, Miranda-Moreno, & Saunier, 2017; Guo et al., 2011; Hashimoto et al., 2015; Ismail, Sayed, & Saunier, 2010; Ka, Lee, Kim, & Yeo, 2019; Zaki & Sayed, 2014) to investigate pedestrians' crossing behaviors. Zaki and Sayed (2014) investigated pedestrians' spatial violations, i.e., the pedestrians were not

walking in the designated region, and temporal violations, i.e., the pedestrians crossed during improper traffic signal phases. Ka et al. (2019) proposed a system used at intersections to predict pedestrians' red-light crossing intentions based on their characteristics, such as age, gender, head orientation, etc. Using R-CNN (Region-based Convolutional Neural Network) object detection model (Girshick, Donahue, Darrell, & Malik, 2014) and SORT tracking model (Nicolai Wojke, Alex Bewley, & Dietrich Paulus, 2017), the system could identify potential dangerous events caused by the pedestrians' violation behaviors. However, as the four models used in this study were machine learning models, they were less effective at capturing the relationships in sequential data for future predictions.

Moreover, with the development of CV technologies, more and more video applications could be used to alert drivers of potential dangerous events of pedestrians. Heng (2008) investigated a traffic signal device that composed of a transmitter, a receiver and a storage medium, where the transmitter could broadcast signals to the receiver. When a collision happened, the facilitation of collision impact data was recorded to ease responsibility determination. Wolterman (2008) designed a traffic signal system to mitigate collisions caused by drivers' traffic violations. Another application "SAFESPOT" was composed of hazard warnings and speed alerts. It was implemented at the black spots of crash on the road networks (Bonnetfoi, Bellotti, Scendzielorz, & Visintainer, 2007). M. Rahman, Islam, Calhoun, and Chowdhury (2019) proposed a system using cameras that outperformed DSRC-enabled devices from the perspective of localization accuracy. It could satisfy the latency requirements with the processing speed of 100ms/frame. The system used a camera to

detect the pedestrians' presences. With TTC (Time-to-collision) as the indicator of safety, the drivers could receive warnings when there were potentially dangerous situations. Another system could recognize pedestrians' crossing intentions, and carry out vehicle evasive maneuvers automatically (Köhler et al., 2013).

Some studies used onboard video sensors to predict the pedestrians' positions and moving paths for proactive pedestrian protection systems (Møgelmoose, Trivedi, & Moeslund, 2015; Schmidt & Färber, 2009). These systems, together with Advanced Driver Assistance System (ADAS) technologies, such as Forward Collision Warning (FCW), were found in the literature to significantly reduce crashes (Yue, Abdel-Aty, & Wu, 2019; Yue, Abdel-Aty, Wu, & Wang, 2018; Yue, Abdel-Aty, Wu, & Farid, 2019). More work should be done to predict pedestrians' red-light violations (Ka et al., 2019; Kotte, Schmeichel, Zlocki, Gathmann, & Eckstein, 2017).

2.3 LSTM Neural Network

Traditional 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. A particular implementation of the Recurrent Neural Network is LSTM (Long Short-Term Memory) neural network model (Hochreiter & Schmidhuber, 1997), which could capture long-term dependencies of time series data. In transportation field, LSTM neural networks were used to predict vehicle travel time or traffic speed on highway links (Altché & Fortelle, 2017; Yanjie, Yisheng, & Fei-Yue, 2016) as well as urban

arterials (Ma, Tao, Wang, Yu, & Wang, 2015). They were also used for driving behavior classifications (Saleh, Hossny, & Nahavandi, 2017) and real-time crash risk predictions (Li, Abdel-Aty, & Yuan, 2020; Yuan, Abdel-Aty, Gong, & Cai, 2019). 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 (Manh & Alaghband, 2018; Xue, Huynh, & Reynolds, 2018). Besides, Alahi et al. (2016) used an LSTM neural network to predict pedestrians' movements based on pedestrians' interactions between each other in crowded spaces. It is promising to predict pedestrians' red-light violation behaviors using LSTM neural network.

CHAPTER 3 DATA COLLECTION

3.1 Experiment setup

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 study site. This intersection was a key intersection with a total entering volume of about 200veh/h. The data of 122 pedestrians who entered the two waiting areas during red-light periods were collected for this study. The spatial map of this site is shown in Figure 1. The camera was GoPro HERO7 camera, setting on a tripod about 6.56 feet high.

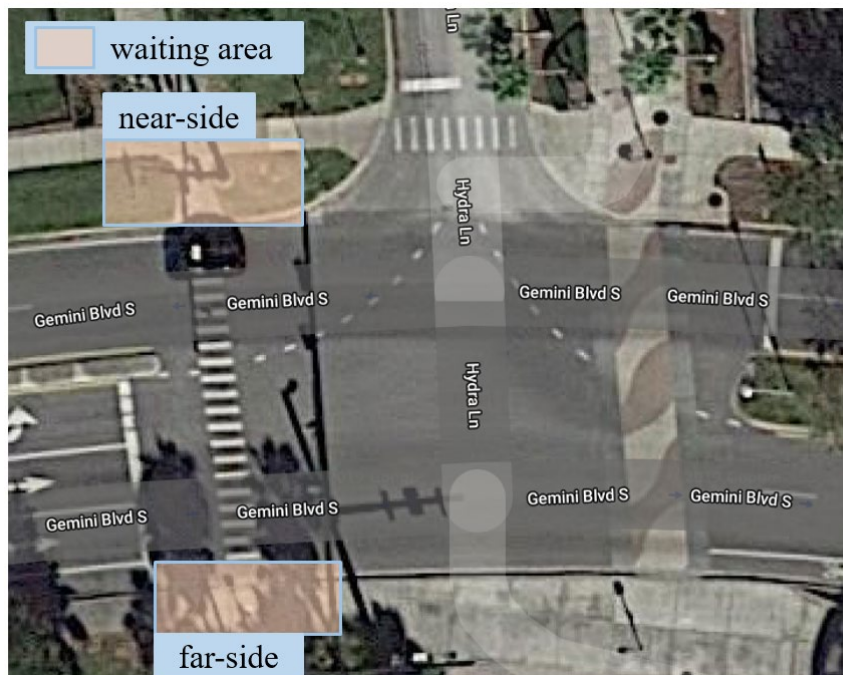


Figure 1 Spatial map of the studied site (Maps, 2020)

3.2 Evaluation of the Pedestrian Safety at the Study Site

Traditional research mainly used crash data to investigate safety. However, crash data suffer from defects listed as follows:

- (1)Crashes are highly rare events. Thus, crash data usually have small sample sizes.
- (2)Crash-based safety evaluation is a reactive method to investigate crashes (Tarko, A. Davis, Saunier, & Sayed, 2009). It is hard to reveal the happening mechanisms of crashes using the information from crash reports, as the process records sometimes come from the experiences and the expectations of police officers (Svensson & Hydén, 2006).

Given above considerations, Surrogate Safety Measures (SSMs), are proposed to investigate the happening mechanisms of crashes (Tarko et al., 2009). SSMs, also called traffic conflict techniques (TCT), are suitable to measure the near-miss situations, i.e., when there are proximities of crashes. Indicators of traffic conflicts include TTC (Time-to-collision, Minderhoud, Bovy, and Prevention (2001)), PET (Post-encroachment-time, Allen, Shin, and Cooper (1978)), DST (Deceleration-to-safety time, Hupfer (1997)), GT (Gap time), etc.

In this work, Post-encroachment time (PET) is used as the indicator of pedestrian safety. PET is defined as the time difference between the moment when the first road user leaves the potential collision area and the moment when the second user reaches it. As shown in Equation 3.1, t_2 and t_1 is the time for different road users to reach the same point accordingly. And PET is the absolute value of the difference. The PET threshold is set to be 6s according to the literature (Radwan, Darius, Wu, & Abou-Senna, 2016) to determine if there is a dangerous condition for the pedestrian.

$$PET = |t_2 - t_1| \qquad \text{Equation 3.1}$$

The PET values are manually collected to ensure analysis accuracy. Among 122 pedestrians, 43 pedestrians start to cross the intersection during the red-light phase. As Figure 2 depicts, among all the situations with small PET values (smaller than the threshold), 23 cases are caused by red-light crossing pedestrians, while 11 cases are caused by normal-crossing pedestrians. Thus, red-light violation pedestrians are more dangerous than others at the intersection. This is consistent with the findings in the literature.

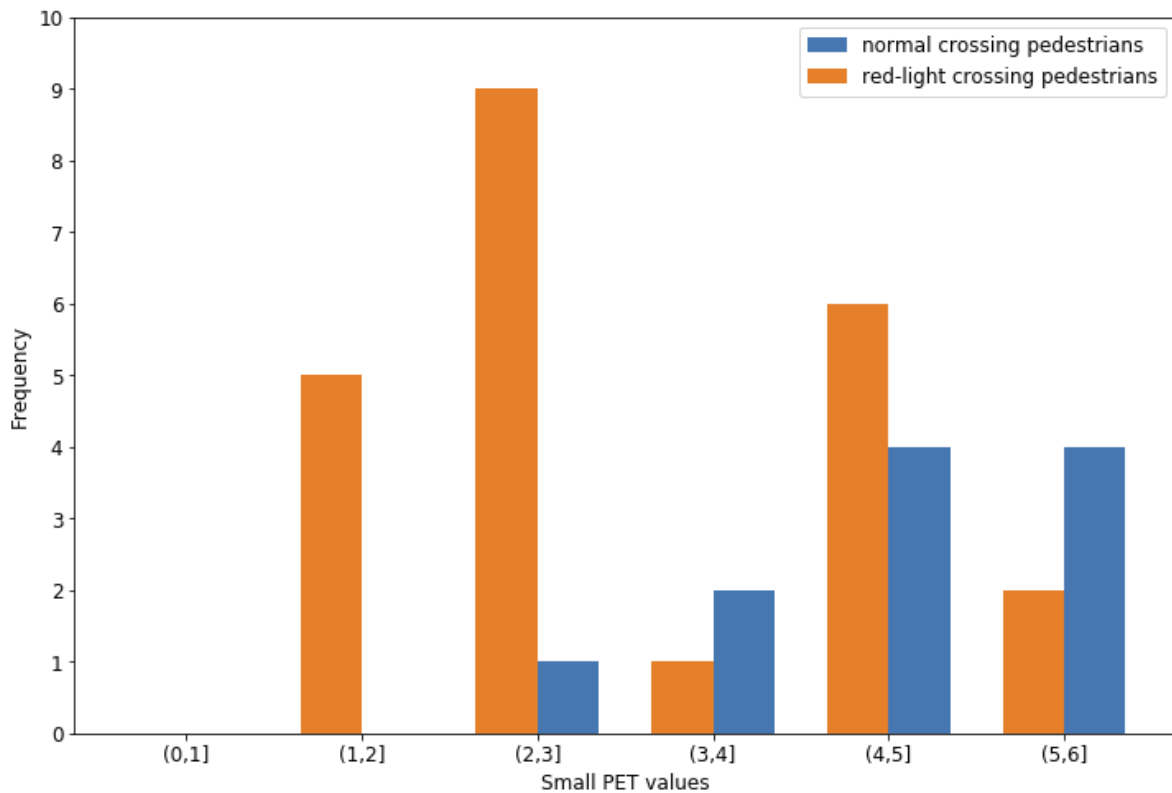


Figure 2 Frequency of small PETs (0-6s) of red-light crossing pedestrians & normal crossing pedestrians

3.3 Video Processing

To extract moving trajectories of pedestrians, computer vision techniques including camera calibration, object detection, and object tracking are used.

3.3.1 Object Detection

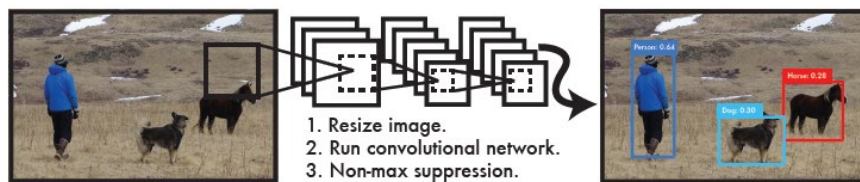
Automated object detection models can identify different kinds of objects in videos, such as pedestrians and vehicles.

For traditional object detection models, background subtraction methods, feature-based methods, frame differencing and motion based methods were used (R. Hadi, G. Sulong, & L. George, 2014) . Motion-based detection methods were typically difficult to detect low-speed moving objects. On the other hand, neural networks were proposed with higher accuracy. Neural networks, especially CNNs (Convolutional Neural Networks), could be used to identify different objects from videos by telling objects' features in pixel level, thus classifying objects into different categories.

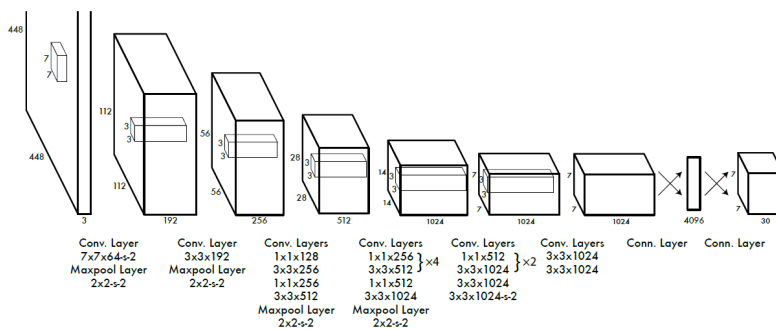
Neural networks models can be divided into two-stage approaches and one-stage approaches. Two-stage approaches, such as Faster R-CNN (Girshick et al., 2014), first divide image into different crops. Then, the CNN is applied to these crops of image to classify different objects. Thus, two-stage approaches are usually computationally expensive and hard to conduct in real time.

One-stage approaches, on the other hand, are more computational effective. One-stage approaches, such as Single Shot Multi-Box Detector (SSD) (Liu et al., 2016) and YOLO (Redmon, Divvala, Girshick, & Farhadi, 2016; Redmon & Farhadi, 2018), can combine the two procedures of cropping and classifying together. Take YOLO for example. The YOLO model applies the single neural network to the full image, dividing the crops and conducting classifications at the same time. As Figure 3 (a) shows, it first resizes the input image to the

resolution of 448×448 (pixel) and runs CNN on the image. CNN is used for predicting bounding boxes, as well as object classes in these boxes. Figure 3(b) shows the architecture of CNN. This CNN has 24 convolutional layers followed by 2 fully connected layers. CNN will then output classes and the confidence scores of the predicted objects.



(a) Detection procedure



(b) Neural network architecture

Figure 3 YOLO detection model (Redmon et al., 2016)

Compared with other detection models, YOLO is an effective detection model for pedestrian detection. It was used in the previous work to detect pedestrians from videos to solve transportation problems (Jana, Biswas, & Mohana, 2018; Ka et al., 2019; J. Lin & Sun, 2018).

In this thesis, a YOLOv3 model is implemented in Keras framework (allanzelener, 2017; qqwwee, 2018). YOLOv3 model improves the original model from the perspective of accuracy, by using multi-scale images, data augmentation, and batch normalization during the

training procedure (Redmon & Farhadi, 2018). OpenCV package is also used to process videos (Bradski, 2019).

YOLOv3 model is evaluated on COCO data set. COCO is a large-scale standardized data set (T.-Y. Lin et al., 2014). It has 330,000 images and 80 categories of objects. It uses mean Average Precision (mAP) for measuring the performance of various object detection models. As COCO data set has 80 categories, including common objects such as person, car, truck, cup, et.al. For each category, the Average Precision value is calculated using the precision-recall curve (Zhu, 2004). Every point of the curve is made up with precision and recall values at different confidence levels. As shown in Equation 3.2, $p_{interp}(r)$ is the maximum precision value over all recall values greater than the recall value r . Then the Equation 3.3 calculated the area under the curve with all the $p_{interp}(r)$ values on the curve.

$$p_{interp}(r) = \max_{\tilde{r}:\tilde{r}\geq r} p(\tilde{r}) \quad \text{Equation 3.2}$$

$$\text{Average Precision} = \int_0^1 p_{interp}(r) dr \quad \text{Equation 3.3}$$

The mean Average Precision (mAP) value is calculated by taking the mean value of Average Precision over 80 categories, as shown in Equation 3.4. The mean Average Precision reaches 57.9% on the test data set (test-dev) of COCO data set.

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i \quad \text{Equation 3.4}$$

3.3.2 Object Tracking

Object tracking is the process of taking the initial sets of object detections (bounding boxes' coordinates), and tracking the objects as they move in the continuous frames. Thus, the trajectories of different objects can be generated from video. The calculations of most SSM indicators, such as TTC, PET, GT, require the trajectories of both parties during their interaction course of the involving road users (Kathuria & Vedagiri, 2020).

As R. A. Hadi, G. Sulong, and L. E. J. a. p. a. George (2014) demonstrated, there were four traditional tracking models available for object tracking, including region based tracking methods, contour tracking methods, 3D model based tracking methods, and feature based tracking methods. Neural networks are used in the state-of-the-art tracking models, especially Multiple Object Tracking (MOT) models. Multiple Object Tracking emerged in these years for tracking multiple objects' movements in the scenes at the same time. In this thesis, tracking is conducted using Deep SORT algorithm (Wojke & Bewley, 2018; N. Wojke, A. Bewley, & D. Paulus, 2017).

Deep SORT improves the original SORT model (Bewley, Ge, Ott, Ramos, & Upcroft, 2016) by using integrated appearance information. It utilizes recursive Kalman filtering and frame-by-frame data association as the original SORT model does. A pre-trained CNN is used to compute the bounding boxes of the tracked objects.

Deep SORT achieves good performance on the MOT16 Challenge benchmark (Milan, Leal-Taixé, Reid, Roth, & Schindler, 2016). MOT16 Challenge Benchmark is a standardized benchmark for measuring the performances of Multiple Object Tracking algorithms. It is a

very important guide for related studies. The accuracy of the tracking model (Deep SORT) in this paper, is measured based on that. According to N. Wojke et al. (2017), as shown in Table 1, the previous models are in the red dotted box, and Deep SORT model is in the blue dotted box. The Deep SORT model outperformed previous models by increasing MOTA score (Multi-object tracking accuracy), and reducing FN (false negatives), etc. This shows this tracking model is very effective. And the MOTA score of the model is 61.4.

Table 1 Evaluation of the Deep SORT tracking model (N. Wojke et al., 2017).

		MOTA ↑	MOTP ↑	MT ↑	ML ↓	ID ↓	FM ↓	FP ↓	FN ↓	Runtime ↑
KDNT [16]*	BATCH	68.2	79.4	41.0%	19.0%	933	1093	11479	45605	0.7 Hz
LMP_p [17]*	BATCH	71.0	80.2	46.9%	21.9%	434	587	7880	44564	0.5 Hz
MCMOT_HDM [18]	BATCH	62.4	78.3	31.5%	24.2%	1394	1318	9855	57257	35 Hz
NOMTwSDP16 [19]	BATCH	62.2	79.6	32.5%	31.1%	406	642	5119	63352	3 Hz
EAMTT [20]	ONLINE	52.5	78.8	19.0%	34.9%	910	1321	4407	81223	12 Hz
POI [16]*	ONLINE	66.1	79.5	34.0%	20.8%	805	3093	5061	55914	10 Hz
SORT [12]*	ONLINE	59.8	79.6	25.4%	22.7%	1423	1835	8698	63245	60 Hz
Deep SORT (Ours)*	ONLINE	61.4	79.1	32.8%	18.2%	781	2008	12852	56668	20 Hz

Figure 4 shows a snapshot of automated video processing process. Road users are assigned with tracker ID and bounding boxes. The blue boxes are generated from the detection model. And the white bounding boxes are generated from the tracking model. The green number is the tracker ID. Through the video processing process, features such as pedestrian locations were generated as input variables at a frequency of 15Hz (environment: NVIDIA GTX 1080Ti 11G GPU).

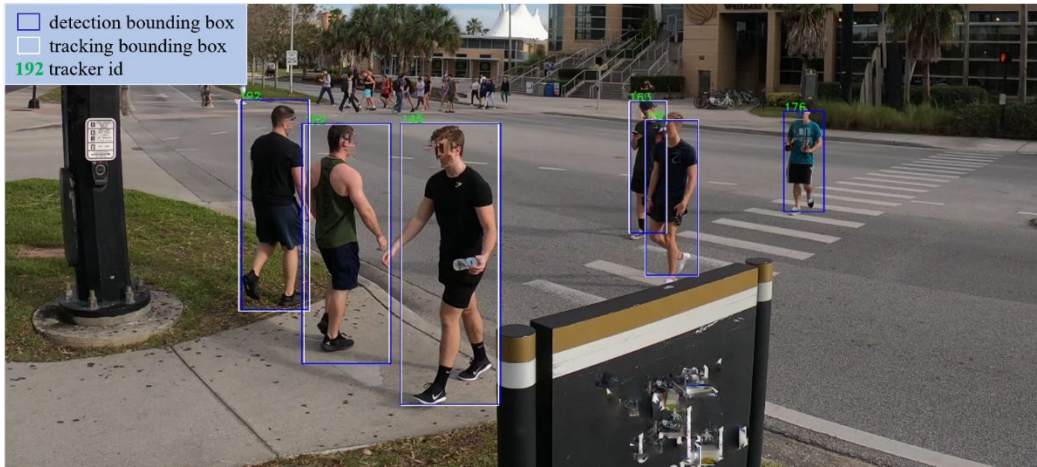


Figure 4 Snapshot of automated video processing

With a unique tracker ID assigned to each pedestrian detected by detection algorithm, movements of the pedestrians are extracted and further analyzed. As Figure 5 shows, trajectories of pedestrians are assigned with different colors, which means they own different ID numbers. And most people use the crosswalks to cross this three-lag intersection, while some walk out of the crosswalk, i.e., they cross the street randomly. It should be noted that only one crosswalk is regarded as the studied area in this work under the consideration of accuracy.

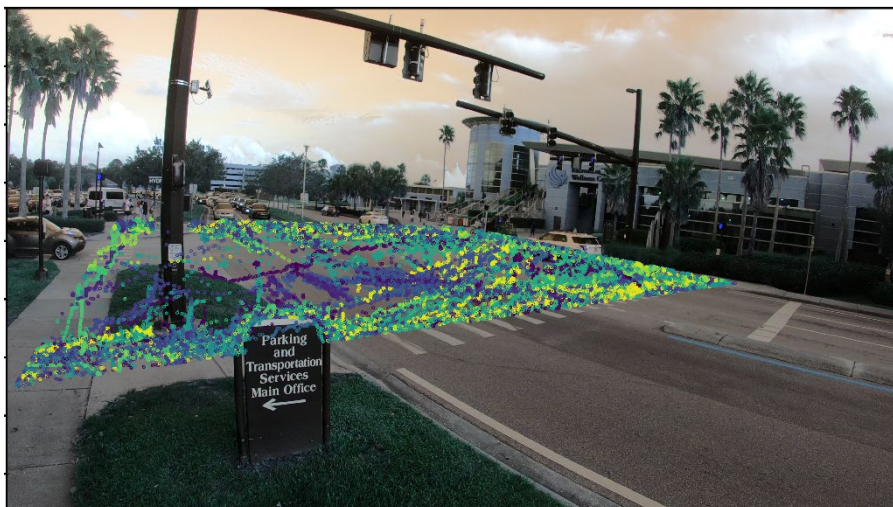


Figure 5 Pedestrians' trajectories at the studied site

3.3.3 Perspective Transformation

As the camera is usually thought to have distortions over objects, and the coordinate system of image is in 2D plane, a homograph matrix \mathbf{h} is used to transform the trajectory extracted from image plane to world plane (Naphade et al., 2019; Španhel, Bartl, Juránek, & Herout, 2019; Tang et al., 2019), as shown in Equation 3.5. (u, v) is the point on image plane, and (X, Y) is the point on world plane.

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \mathbf{h} \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} \quad \text{Equation 3.5}$$

The procedure to obtain \mathbf{h} matrix is demonstrated in Equation 3.6. \mathbf{h} matrix contains 9 values in total, from h_1 to h_9 , while the last one can be 1. Thus, to get the other 8 values, both point coordinates in image plane, denoting by (u_i, v_i) , and point coordinates in real world plane, denoting by (X_i, Y_i) , are gathered. Ten pairs of points in the traffic scene, which are not collinear, are extracted from image and Google Maps[©] to obtain \mathbf{h} matrix. Each pair of points forms two rows of matrix A. And the right side of the function is a zero vector. This function has a linear least-squares solution. And singular value decomposition (SVD) can be used. Namely, the SVD will find the solution for minimizing the value of $\|\mathbf{A} * \mathbf{h}\|$, with $\|\mathbf{h}\|$ equals 1.

$$\mathbf{A} * \mathbf{h} = \begin{bmatrix} 0 & 0 & 0 & -X_1 & -Y_1 & 1 & v_1 X_1 & v_1 Y_1 & v_1 \\ X_1 & Y_1 & 1 & 0 & 0 & 0 & -u_1 X_1 & -u_1 Y_1 & -u_1 \\ 0 & 0 & 0 & -X_2 & -Y_2 & 1 & v_2 X_2 & v_2 Y_2 & v_2 \\ X_2 & Y_2 & 1 & 0 & 0 & 0 & -u_2 X_2 & -u_2 Y_2 & -u_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \\ h_7 \\ h_8 \\ h_9 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad \text{Equation 3.6}$$

The Figure 6 shows the result of calibration. To validate the accuracy of the camera calibration, the back-projection error is used. The back-projection error is defined as Euclidean distance in pixel level between a projected point and a measured one. As Figure 7 shows, the blue dots show the input points (u_i, v_i) . And the red dots (\hat{u}_i, \hat{v}_i) are the corresponding points back projected from (X_i, Y_i) using \mathbf{h} matrix. The average back-projection error is calculated to be 3.008 in this work, as shown in Equation 3.7.

Average back – projection error

$$= \frac{\sum_{i=0}^n \text{distance}((u_i, v_i), (\hat{u}_i, \hat{v}_i))}{n} \quad \text{Equation 3.7}$$



Figure 6 Perspective transformation

After acquiring \mathbf{h} , all pixel coordinates generated from video (images) are converted to real world coordinates further through the inverse matrix of \mathbf{h} .

3.3.4 Other Features

Other independent variables used in this study, which are manually labeled, include: pedestrian's gender, pedestrian's walking direction, and whether the pedestrian is walking in a group. The walking direction of pedestrians are denoted by 1 ("towards near-sided crosswalk") and 0 ("towards far-sided crosswalk").

CHAPTER 4 METHODOLOGY

4.1 Pedestrian Crossing Modeling

After above steps, a time series dataset, which is composed of pedestrians' trajectories, is generated from videos. To label the dependent variable Y , denoting whether the pedestrian has red-light violations, a pedestrian's crossing model is established as shown in Figure 7. Suppose the driver will take evasive actions after capturing the pedestrian's crossing intention after the reaction time θ . In this study, the reaction time θ is taken as 1.5s according to the literature (M. H. Rahman, Abdel-Aty, Lee, & Rahman, 2019; Wilson, Butler, McGehee, & Dingus, 1997). On the other hand, pedestrians' crossing behavior can be divided into three stages, (1) pedestrian showing up, (2) pedestrian showing crossing intention, (3) pedestrian starting to cross. The behaviors are observed frame by frame by the author. Thus, for jaywalking pedestrians, from t_2 to t_3 , the dependent predictor Y is labeled as positive ("1"). This time interval is when the pedestrian is observing surrounding areas and starting to cross, i.e., showing crossing intentions. And the time intervals between t_2 to t_3 are different for different persons. Labels are shifted ahead θ units by the timestamp of each pedestrian for prediction purpose. Thus, we are basically predicting the pedestrian's red-light crossing intention 1.5 seconds ahead.

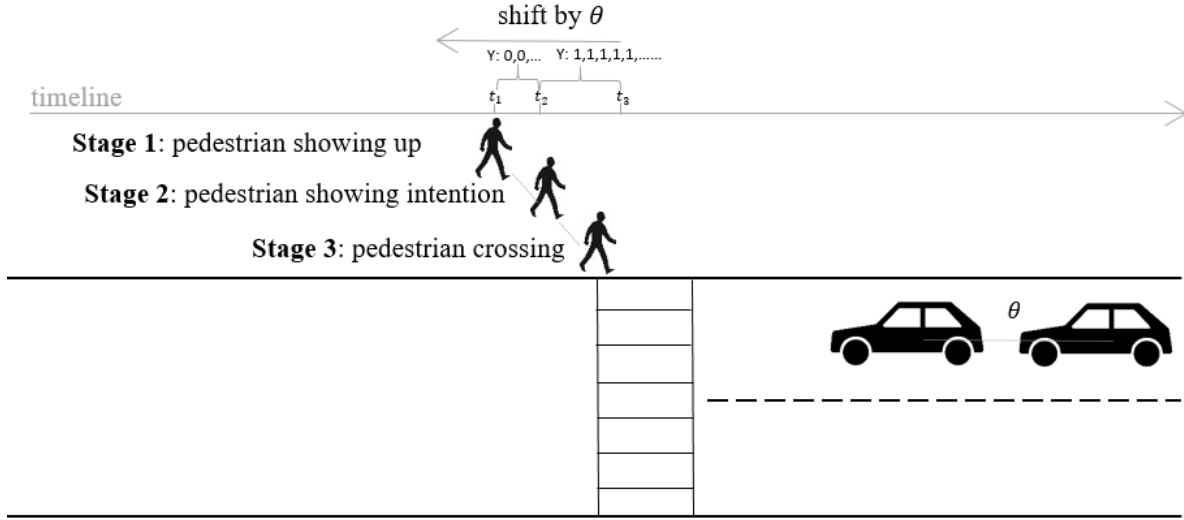


Figure 7 Pedestrian's crossing model

4.2 LSTM Neural Network Model

4.2.1 LSTM Model

Based on the above discussion, a methodology of pedestrian crossing intention prediction at the signalized crosswalk is carried out using an LSTM neural network (Hochreiter & Schmidhuber, 1997). As pedestrians' trajectories are time-series data, LSTM neural network model can better capture the temporal evolution of time sequences.

LSTM neural network is a special kind of RNN (Recurrent Neural Network). The basic formulation of RNN is shown in Equation 4.1 and Equation 4.2. With \mathbf{x} is the input vector, \mathbf{y} is the output vector, the output of the last hidden layer is fed into the input of the next layer in the time sequence. W is weight matrix. And b_y is bias term.

$$\mathbf{h}_T = \sigma(W_1 \mathbf{x}_t + W_{ih} \mathbf{h}_{T-1} + W_{hh} c_{t-1} + b_y) \quad \text{Equation 4.1}$$

$$\mathbf{y}_T = W_{hy} \mathbf{h}_T + b_y \quad \text{Equation 4.2}$$

However, RNNs suffer from learning long-term dependency from time-series data, which will cause the problem of vanishing gradients. To overcome it, LSTM neural network is proposed with a purpose-built memory unit to store information (A. Graves, A.-r. Mohamed, & G. Hinton, 2013). As shown in Figure 8, 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 sequence x_t is computed by Equation 4.3-Equation 4.8 to generate the output y_t , which is a vector of probabilities, iterated from $t=1$ to T . σ denotes logistics sigmoid function. And \otimes denotes elementwise product of the vectors. \emptyset is the activation function \tanh .

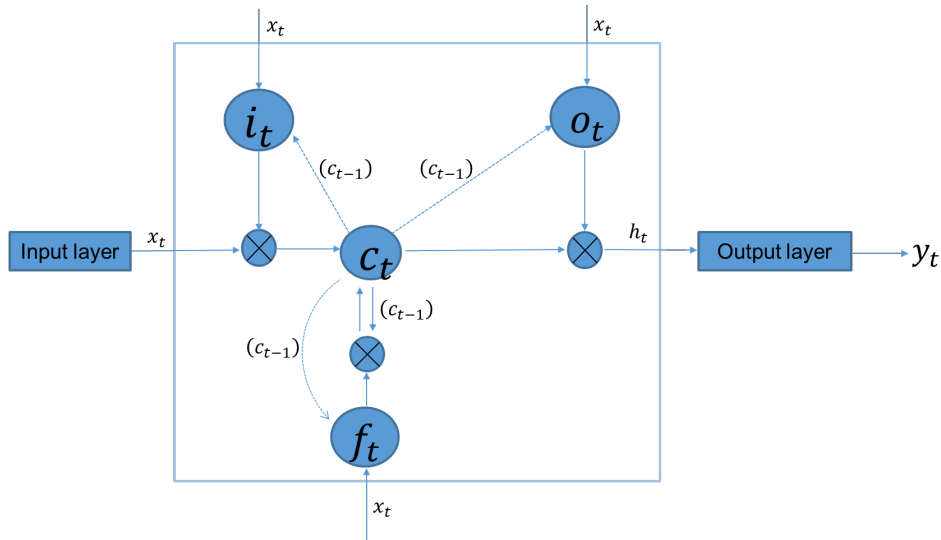


Figure 8 Schematic of LSTM unit (adapted from (A. Graves, A. Mohamed, & G. Hinton, 2013; Kang, Lv, & Chen, 2017))

$$i_t = \sigma(W_{xi}x_t + W_{ih}h_{t-1} + W_{ic}c_{t-1} + b_i) \quad \text{Equation 4.3}$$

$$f_t = \sigma(W_{fx}x_t + W_{fh}h_{t-1} + W_{fc}c_{t-1} + b_f) \quad \text{Equation 4.4}$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad \text{Equation 4.5}$$

$$c_t = f_t \otimes c_{t-1} + i_t \otimes \phi(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad \text{Equation 4.6}$$

$$h_t = o_t \otimes \phi(c_t) \quad \text{Equation 4.7}$$

$$y_t = W_{hy}h_t + b_y \quad \text{Equation 4.8}$$

4.2.2 Model Architecture

The model architecture used in the study is illustrated in Figure 9. The features from three time-slices are stacked as input to predict the output 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, dropout layer is added to prevent overfitting. The Sigmoid function is used as the activation function to generate the output. Adam function is used as the optimization function (Kingma & Ba, 2014). The model is implemented in Keras framework (Chollet, 2015). After testing different combinations of hyperparameters, the hyperparameter values are selected as below: learning rate is 0.0005, batch size is 1000, and epoch number is 5.

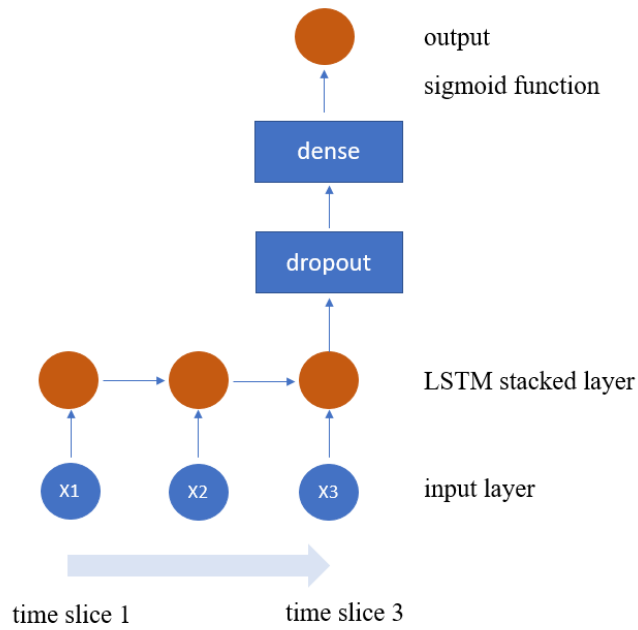


Figure 9 Model architecture

4.3 Experiments and Results

4.3.1 Data Set Overview

The summary of the descriptive statistics of all variables is shown in Table 2. The bold marked row is the dependent predictor. The location feature is generated from videos automatically, other features are manually labelled.

Table 2 Summary of variable descriptive statistics

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

* Note that the first 3 features are manually labeled. Locations are generated from video.

4.3.2 Oversampling

After splitting the data set into training and test data sets (0.75:0.25), the training data set has 60,821 observations and the test data set has 20,274 observations. In the training data set, there were 58,882 normal crossing samples and 1,939 jaywalking samples. The ratio is around 30:1, indicating the data are highly imbalanced. A highly imbalanced data set will result in a bad model. Thus, the oversampling strategy is used to generate a balanced data set on the training data set.

Synthetic Minority Over-Sampling Technique (SMOTE) is an oversampling strategy used to increase the number of positive samples (Chawla, Bowyer, Hall, & Kegelmeyer, 2002). SMOTE is a popular over-sampling method, which can create new instances by interpolating between several minority class examples that lie together.

4.3.3 Experiment Results

To evaluate the experiment results, the diagram for metrics calculation used is shown in Table 3. “Positive” denotes that pedestrians conducted the red-light violations. “Negative” denotes that pedestrian doesn’t conduct the red-light violations. Metrics such as sensitivity, specificity, and accuracy are calculated as shown in Equation 4.9- Equation 4.11. Sensitivity measures show how good the model is among all the positives, i.e., 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.

Table 3 Diagram for metrics calculation

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

$$Sensitivity = \frac{TP}{TP+FN} \quad \text{Equation 4.9}$$

$$Specificity = \frac{TN}{TN+FP} \quad \text{Equation 4.10}$$

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \quad \text{Equation 4.11}$$

The ROC curve (Receiver Operating Characteristic 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 4.12 at

different classification thresholds. AUC (Area Under the ROC Curve) value is used as an accuracy indicator.

$$\text{False Positive Rate} = \frac{FP}{FP+TN} \quad \text{Equation 4.12}$$

The results of the prediction with the proposed model are listed in Table 4. The proposed model achieved the sensitivity value of 92.4% on positive samples, which means it can predict 92.4% red-light violation behaviors successfully. However, there are still some room for improvements, since the model has a relatively high False Positive Rate. Overall, the model shows a prediction accuracy of 91.6% on all classes. And the AUC values achieved on the training data set and test data set are 0.943 and 0.938.

Table 4 Prediction results

Training set	Test set			
AUC	Sensitivity	False Positive Rate	Accuracy	AUC
0.943	0.924	0.203	0.916	0.938

It should be noted that the model is trained and tested using the data collected at the same location, which is also regarded as internal testing. The future study can be to apply the trained model to a different location, to prove the model's generalization.

CHAPTER 5 CONCLUSIONS

This study uses video data to predict pedestrians' red-light violations at a signalized intersection with a stacked LSTM neural network. 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 are also used to feed into the LSTM model. And the red-light crossing intentions of pedestrians are labeled after analyzing the interaction between red-light crossing pedestrians and vehicles. An LSTM model is proposed to predict pedestrians' red-light violation behaviors at 1.5 seconds ahead. The experiment result shows that the model reaches the accuracy of 91.6% at one signalized crosswalk. The related work has been published (Zhang, Abdel-Aty, Yuan, & Li, 2020).

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 the normal crossing pedestrians as pedestrians with red-light violations. More features related to 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 more intersections to alert drivers of the pedestrians with unexpected crossing behaviors, thus preventing collisions between pedestrians and vehicles.

REFERENCES

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179-211. Retrieved from <http://www.sciencedirect.com/science/article/pii/074959789190020T>. doi:[https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L., & Savarese, S. (2016). Social lstm: Human trajectory prediction in crowded spaces. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- allanzelener, A. G., Guillaume Ardaud. (2017). YAD2K: Yet Another Darknet 2 Keras. Retrieved from <https://github.com/allanzelener/YAD2K>
- Allen, B. L., Shin, B. T., & Cooper, P. J. (1978). Analysis of traffic conflicts and collisions (0361-1981). Retrieved from
- Altché, F., & Fortelle, A. d. L. (2017, 16-19 Oct. 2017). An LSTM network for highway trajectory prediction. Paper presented at the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC).
- Bewley, A., Ge, Z., Ott, L., Ramos, F., & Upcroft, B. (2016). Simple online and realtime tracking. Paper presented at the 2016 IEEE International Conference on Image Processing (ICIP).
- Bonnefoi, F., Bellotti, F., Scendzielorz, T., & Visintainer, F. (2007). SAFESPOT Applications for Infrastructurebased Co-operative Road Safety. Paper presented at the 14th World Congress and Exhibition on Intelligent Transport Systems and Services.
- Bradski, G. (2019). Open source computer vision. Retrieved from <https://opencv.org/>
- Brosseau, M., Zangenehpour, S., Saunier, N., & Miranda-Moreno, L. (2013). 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*, 21, 159-172. Retrieved from <http://www.sciencedirect.com/science/article/pii/S1369847813000855>. doi:<https://doi.org/10.1016/j.trf.2013.09.010>
- 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.
- Chollet, F. (2015). Keras. Retrieved from <https://github.com/fchollet/keras>
- Cinnamon, J., Schuurman, N., & Hameed, S. M. (2011). Pedestrian Injury and Human Behaviour: Observing Road-Rule Violations at High-Incident Intersections. *PLOS ONE*, 6(6), e21063. Retrieved from <https://doi.org/10.1371/journal.pone.0021063>. doi:10.1371/journal.pone.0021063
- Evans, D., & Norman, P. (1998). Understanding pedestrians' road crossing decisions: an application of the theory of planned behaviour. *Health Education Research*, 13(4), 481-489. Retrieved from <https://doi.org/10.1093/her/13.4.481-a>. doi:10.1093/her/13.4.481-a
- FHWA, U. S. D. o. T. (2018). FHWA Home, Safety, Pedestrian & Bicycle Safety. Retrieved from https://safety.fhwa.dot.gov/ped_bike/

- Formosa, N., Quddus, M., Ison, S., Abdel-Aty, M., & Yuan, J. (2020). Predicting real-time traffic conflicts using deep learning. *Accident Analysis & Prevention*, 136, 105429. Retrieved from <http://www.sciencedirect.com/science/article/pii/S000145751930973X>. doi:<https://doi.org/10.1016/j.aap.2019.105429>
- Fu, T., Miranda-Moreno, L., & Saunier, N. (2017). A novel framework to evaluate pedestrian safety at non-signalized locations (Vol. 111).
- Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. Paper presented at the Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition.
- Gitelman, V., Balasha, D., Carmel, R., Hendel, L., & Pesahov, F. (2012). Characterization of pedestrian accidents and an examination of infrastructure measures to improve pedestrian safety in Israel. *Accident Analysis & Prevention*, 44(1), 63-73. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0001457510003611>. doi:<https://doi.org/10.1016/j.aap.2010.11.017>
- Graves, A., Mohamed, A.-r., & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. Paper presented at the 2013 IEEE international conference on acoustics, speech and signal processing.
- Graves, A., Mohamed, A., & Hinton, G. (2013, 26-31 May 2013). Speech recognition with deep recurrent neural networks. Paper presented at the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing.
- Guo, H., Gao, Z., Yang, X., & Jiang, X. (2011). Modeling Pedestrian Violation Behavior at Signalized Crosswalks in China: A Hazards-Based Duration Approach. *Traffic Injury Prevention*, 12(1), 96-103. Retrieved from <https://www.tandfonline.com/doi/abs/10.1080/15389588.2010.518652>. doi:10.1080/15389588.2010.518652
- Hadi, R., Sulong, G., & George, L. (2014). Vehicle Detection and Tracking Techniques: A Concise Review (Vol. 5).
- Hadi, R. A., Sulong, G., & George, L. E. J. a. p. a. (2014). Vehicle detection and tracking techniques: a concise review.
- Hamed, M. M. (2001). Analysis of pedestrians' behavior at pedestrian crossings. *Safety Science*, 38(1), 63-82. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0925753500000588>. doi:[https://doi.org/10.1016/S0925-7535\(00\)00058-8](https://doi.org/10.1016/S0925-7535(00)00058-8)
- Hashimoto, Y., Yanlei, G., Hsu, L., & Shunsuke, K. (2015, 15-18 Sept. 2015). A Probabilistic Model for the Estimation of Pedestrian Crossing Behavior at Signalized Intersections. Paper presented at the 2015 IEEE 18th International Conference on Intelligent Transportation Systems.
- Heng, S. T. (2008). Pedestrian road safety system. In: Google Patents.
- Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.

- Hupfer, C. (1997). Deceleration to safety time (DST)-a useful figure to evaluate traffic safety. Paper presented at the ICTCT Conference Proceedings of Seminar.
- Ismail, K., Sayed, T., & Saunier, N. (2010). Automated Analysis of Pedestrian–Vehicle Conflicts: Context for Before-and-After Studies. *Transportation Research Record*, 2198(1), 52-64. Retrieved from <https://doi.org/10.3141/2198-07>. doi:10.3141/2198-07
- Jana, A. P., Biswas, A., & Mohana. (2018, 18-19 May 2018). YOLO based Detection and Classification of Objects in video records. Paper presented at the 2018 3rd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT).
- Ka, D., Lee, D., Kim, S., & Yeo, H. (2019). Study On the Framework of Intersection Pedestrian Collision Warning System Considering Pedestrian Characteristics.
- Kang, D., Lv, Y., & Chen, Y. (2017, 16-19 Oct. 2017). Short-term traffic flow prediction with LSTM recurrent neural network. Paper presented at the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC).
- Kathuria, A., & Vedagiri, P. (2020). Evaluating pedestrian vehicle interaction dynamics at un-signalized intersections: a proactive approach for safety analysis. *Accident Analysis & Prevention*, 134, 105316.
- Keegan, O., & O'Mahony, M. (2003). Modifying pedestrian behaviour. *Transportation Research Part A: Policy and Practice*, 37(10), 889-901. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0965856403000612>. doi:[https://doi.org/10.1016/S0965-8564\(03\)00061-2](https://doi.org/10.1016/S0965-8564(03)00061-2)
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- Köhler, S., Schreiner, B., Ronalter, S., Doll, K., Brunsmann, U., & Zindler, K. (2013, 23-26 June 2013). Autonomous evasive maneuvers triggered by infrastructure-based detection of pedestrian intentions. Paper presented at the 2013 IEEE Intelligent Vehicles Symposium (IV).
- Kotte, J., Schmeichel, C., Zlocki, A., Gathmann, H., & Eckstein, L. (2017). Concept of an enhanced V2X pedestrian collision avoidance system with a cost function-based pedestrian model (Vol. 18).
- Li, P., Abdel-Aty, M., & Yuan, J. (2020). Real-time crash risk prediction on arterials based on LSTM-CNN. *Accident Analysis & Prevention*, 135, 105371.
- Lin, J., & Sun, M. (2018, 30 Nov.-2 Dec. 2018). A YOLO-Based Traffic Counting System. Paper presented at the 2018 Conference on Technologies and Applications of Artificial Intelligence (TAAI).
- Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., . . . Zitnick, C. L. (2014). Microsoft coco: Common objects in context. Paper presented at the European conference on computer vision.
- Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.-Y., & C. Berg, A. (2016). SSD: Single Shot MultiBox Detector (Vol. 9905).
- Ma, X., Tao, Z., Wang, Y., Yu, H., & Wang, Y. (2015). Long short-term memory neural network for traffic speed prediction using remote microwave sensor data.

- Transportation Research Part C: Emerging Technologies, 54, 187-197. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0968090X15000935>.
doi:<https://doi.org/10.1016/j.trc.2015.03.014>
- Manh, H., & Alaghband, G. (2018). Scene-lstm: A model for human trajectory prediction. arXiv preprint arXiv:1808.04018.
- Maps, G. (2020). University of Central Florida. Retrieved from <https://www.google.com/maps/@28.5965396,-81.1993853,87m/data=!3m1!1e3>
- Milan, A., Leal-Taixé, L., Reid, I., Roth, S., & Schindler, K. (2016). MOT16: A benchmark for multi-object tracking. arXiv preprint arXiv:1603.00831.
- Minderhoud, M. M., Bovy, P. H. J. A. A., & Prevention. (2001). Extended time-to-collision measures for road traffic safety assessment. 33(1), 89-97.
- Møgelmoose, A., Trivedi, M. M., & Moeslund, T. B. (2015). Trajectory analysis and prediction for improved pedestrian safety: Integrated framework and evaluations. Paper presented at the 2015 IEEE Intelligent Vehicles Symposium (IV).
- Naphade, M., Tang, Z., Chang, M.-C., Anastasiu, D. C., Sharma, A., Chellappa, R., . . . Hwang, J.-N. (2019). The 2019 AI City Challenge. Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops.
- NHTSA, N. H. T. S. A. (2019). Pedestrian Safety. Retrieved from <https://www.nhtsa.gov/road-safety/pedestrian-safety>
- qqwweee, t., dleam, b02902131, stefanbo92, philtrade, jiaowoboshao. (2018). A Keras implementation of YOLOv3 (Tensorflow backend) Retrieved from <https://github.com/qqwweee/keras-yolo3>
- Radwan, E., Darius, B., Wu, J., & Abou-Senna, H. (2016). Simulation of pedestrian safety surrogate measures. Paper presented at the ARRB Conference, 27th, 2016, Melbourne, Victoria, Australia.
- Rahman, M., Islam, M., Calhoun, J., & Chowdhury, M. (2019). Real-Time Pedestrian Detection Approach with an Efficient Data Communication Bandwidth Strategy. Transportation Research Record, 0361198119843255.
- Rahman, M. H., Abdel-Aty, M., Lee, J., & Rahman, M. S. (2019). Enhancing traffic safety at school zones by operation and engineering countermeasures: A microscopic simulation approach. Simulation Modelling Practice and Theory, 94, 334-348. Retrieved from <http://www.sciencedirect.com/science/article/pii/S1569190X19300358>.
doi:<https://doi.org/10.1016/j.simpat.2019.04.001>
- Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection.
- Redmon, J., & Farhadi, A. (2018). YOLOv3: An Incremental Improvement.
- Saleh, K., Hossny, M., & Nahavandi, S. (2017, 16-19 Oct. 2017). Driving behavior classification based on sensor data fusion using LSTM recurrent neural networks. Paper presented at the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC).

- Schmidt, S., & Färber, B. (2009). Pedestrians at the kerb – Recognising the action intentions of humans. *Transportation Research Part F: Traffic Psychology and Behaviour*, 12(4), 300-310. Retrieved from <http://www.sciencedirect.com/science/article/pii/S1369847809000102>. doi:<https://doi.org/10.1016/j.trf.2009.02.003>
- Španhel, J., Bartl, V., Juránek, R., & Herout, A. (2019). Vehicle Re-Identification and Multi-Camera Tracking in Challenging City-Scale Environment. Paper presented at the Proc. CVPR Workshops.
- Svensson, Å., & Hydén, C. (2006). Estimating the severity of safety related behaviour. *Accident; analysis and prevention*, 38, 379-385. doi:10.1016/j.aap.2005.10.009
- Tang, Z., Naphade, M., Liu, M.-Y., Yang, X., Birchfield, S., Wang, S., . . . Hwang, J.-N. (2019). Cityflow: A city-scale benchmark for multi-target multi-camera vehicle tracking and re-identification. Paper presented at the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.
- Tarko, A., A. Davis, G., Saunier, N., & Sayed, T. (2009). Surrogate Measures of Safety.
- Wilson, T. B., Butler, W., McGehee, D. V., & Dingus, T. A. (1997). Forward-looking collision warning system performance guidelines. *SAE Transactions*, 701-725.
- Wojke, N., & Bewley, A. (2018). Deep Cosine Metric Learning for Person Re-Identification.
- Wojke, N., Bewley, A., & Paulus, D. (2017, 17-20 Sept. 2017). Simple online and realtime tracking with a deep association metric. Paper presented at the 2017 IEEE International Conference on Image Processing (ICIP).
- Wojke, N., Bewley, A., & Paulus, D. (2017). Simple online and realtime tracking with a deep association metric.
- Wolterman, M. (2008). Infrastructure-based collision warning using artificial intelligence. In: Google Patents.
- Xue, H., Huynh, D. Q., & Reynolds, M. (2018, 12-15 March 2018). SS-LSTM: A Hierarchical LSTM Model for Pedestrian Trajectory Prediction. Paper presented at the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV).
- Yanjie, D., Yisheng, L., & Fei-Yue, W. (2016, 1-4 Nov. 2016). Travel time prediction with LSTM neural network. Paper presented at the 2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC).
- Yuan, J., Abdel-Aty, M., Gong, Y., & Cai, Q. (2019). Real-Time Crash Risk Prediction using Long Short-Term Memory Recurrent Neural Network. *Transportation Research Record*, 2673(4), 314-326. Retrieved from <https://doi.org/10.1177/0361198119840611>. doi:10.1177/0361198119840611
- Yue, L., Abdel-Aty, M., & Wu, Y. (2019). The Crash Avoidance Effectiveness of Advanced Driver Assistance Systems in Real-World Environment. Paper presented at the International Conference on Transportation and Development 2019: Smarter and Safer Mobility and Cities.
- Yue, L., Abdel-Aty, M., Wu, Y., & Wang, L. (2018). Assessment of the safety benefits of vehicles' advanced driver assistance, connectivity and low level automation systems. *Accident Analysis & Prevention*, 117, 55-64. Retrieved from

- <http://www.sciencedirect.com/science/article/pii/S0001457518301404>.
doi:<https://doi.org/10.1016/j.aap.2018.04.002>
- 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. *Journal of Safety Research*, 73, 119-132. Retrieved from <http://www.sciencedirect.com/science/article/pii/S002243752030027X>. doi:<https://doi.org/10.1016/j.jsr.2020.02.020>
- Yue, L., Abdel-Aty, M. A., Wu, Y., & Farid, A. (2019). The Practical Effectiveness of Advanced Driver Assistance Systems at Different Roadway Facilities: System Limitation, Adoption, and Usage. *IEEE Transactions on Intelligent Transportation Systems*.
- Zaki, M. H., & Sayed, T. (2014). Automated Analysis of Pedestrians' Nonconforming Behavior and Data Collection at an Urban Crossing. *Transportation Research Record*, 2443(1), 123-133. Retrieved from <https://doi.org/10.3141/2443-14>. doi:10.3141/2443-14
- 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. *Transportation Research Record*, 0361198120912422. Retrieved from <https://doi.org/10.1177/0361198120912422>. doi:10.1177/0361198120912422
- Zhu, M. (2004). Recall, precision and average precision. *Department of Statistics and Actuarial Science, University of Waterloo, Waterloo*, 2, 30.