Research on Dynamic Decision-making Hybrid Model of Pedestrian Flow Based on Deep Learning

Shuxu Zhao, Shiyao Men, Baohua Zhang

School of Electronic & Information Engineering, Lanzhou Jiaotong University, Lanzhou, China

e-mail: 2986375972@qq.com

ABSTRACT. With the progress and development of China's transportation system, pedestrian travel behavior and safety have received increasing attention. At the same time, controlling the population density and population movement in densely populated areas plays an important role in ensuring national security for the public security departments. Therefore, how to properly adjust the switching time of traffic signals has become an urgent problem. In view of the above requirements, this paper proposes a signal switching model based on deep learning for dynamic regulation of pedestrian traffic. The hybrid model is divided into three parts, named the real-time data acquisition part, the historical data analysis and prediction part and the decision model part. Firstly, the LSTM model is used for the analysis and prediction of historical traffic data with time series characteristics. Then, the real-time data acquisition adopts the lightweight and high-performance target detection model MobileNet-SSD proposed by Google. Finally, the signal switching decision model is proposed to analyze and determine the data provided by the above model, and two adjustment factors are defined to adjust the proportion of historical data and real-time data to the impact of decision results.

CCS Concepts
•Computing methods→Artificial intelligence→Computer vision →computer vision tasks.

1. INTRODUCTION
With the development and progress of China's transportation system, the convenience and safety of pedestrian travel has received increasing attention. Controlling the flow of people in densely populated areas plays an important role in ensuring people's safety in public security and other related departments [1]. At the same time, in order to effectively reduce traffic hazards caused by illegal activities such as red light in urban road traffic [2]. We must reasonably adjust the transit time of road signal lights, control the pedestrian flow rate, and reduce the occurrence of hidden dangers. In the current traffic signal control, there are two main types. The first type is our traditional fixed duration method, and the second type is the way for pedestrians to actively switch signals. The first method is fixed in a single way, and the time utilization rate is low. In the second way, due to irregular switching of pedestrians, the congestion of the vehicle may occur. Therefore, how to control the conversion time of traffic lights has become an urgent problem we need to solve.

In recent years, deep learning has become a mainstream solution in image and natural language processing, signal processing and other issues. Similarly, it has been greatly applied in the exploration of many problems such as target detection for computer vision. All kinds of target detection algorithm based on depth of learning in the research field of people, such as R-CNN, Fast-RCNN, YOLO, SSD, and so on. For R-CNN, Faster R-CNN, the detection principle is to first obtain the candidate box
through CNN, and then carry out classification and regression, and YOLO and SSD can complete the
detection directly in one step [3]. Compared to YOLO, SSD uses CNN for direct detection, rather than
YOLO-like detection after the fully connected layer. The disadvantage of the YOLO algorithm is that
it is difficult to detect small targets and the positioning is not accurate, but the improvement on the
SSD overcomes these disadvantages to some extent.

At the same time, using the deep learning model to analyze the regression problem becomes the
mainstream method. The LSTM model is mostly used to solve regression problems based on time
series data sets [4]. In terms of traffic flow prediction, many scholars have made effective predictions
based on traffic history data [5]. Therefore, we propose a decision-making method based on the deep
learning hybrid model to combine the advantages of the two models to make decisions.

2. URBAN ROAD FLOW DATA CONTROL ANALYSIS

2.1 Periodic Characteristics of Urban Road Flow
Pedestrian flow at any chosen intersection varies periodically [6, 7]. There are many reasons that cause
this cyclical changes, such as fixed living people, fixed schedules, and by the fixed requirements (such
as supermarket, attractions, school, etc.) caused by the behavior. By understanding the cyclical
changes in the flow of people on the road, we can better understand the characteristics of travel
behavior of people in a certain area [8]. Therefore, we consider using the analysis of historical data to
obtain the periodic characteristics of the human flow.

2.2 Real-time Characteristics of Urban Road Flow
In addition to the fixed cyclical changes caused by the above reasons, uncertain factors such as holiday
activities and sudden events around the road will have a dynamic impact on pedestrian traffic. We call
this impact the real-time nature of urban road flow. This kind of real-time nature requires us to follow
the macro-periodic rule of the flow of people, and timely adjust the dynamic control of the
corresponding traffic signals to cope with the real-time changes. Therefore, we need to obtain real-
time dynamic data through the target detection model [9, 10].

3. ALGORITHM MODEL
The algorithm model consists of three parts: the LSTM model, the object detection model, and the
decision model. The overall model is shown in the following figure:

![Figure 1 Overall model structure](image)

3.1 LSTM Model
Long- and short-term memory networks are often referred to as "LSTM", [11] is a special kind of
RNN that can learn long-term dependencies [12]. They were proposed by Hochreiter & S chmidhuber
(1997). And it has been developed and applied in solving the regression problem of data with timing
characteristics. In the original RNN, its network structure is too simple. As shown below:
Such a structure makes RNN not able to pass the previous data information to a position far enough back. Therefore, LSTM is designed to overcome this shortcoming (Figure 3 is the model structure of LSTM).

The core of LSTM lies in the transfer and update of unit states, which is the biggest difference between RNN and LSTM. The LSTM retains the valid information of each unit through the cell state and passes it to the next unit. Its working principle is as follows ($h_{t-1}$ indicates the output of the previous unit, $x_t$ for the input of new information, $x_t$ for the current unit, $\sigma$ for the activate function of the sigmoid):

1. The forget gate layer determines that we will discard the useless information data from the cell state.
   \[
   f_t = \sigma(W_f[h_{t-1}, x_t] + b_f)
   \] (1)

2. The input gate layer determines that we will store valid new information data in the unit state.
   \[
   i_t = \sigma(W_i[h_{t-1}, x_t] + b_i)
   \]
   \[
   \tilde{C}_t = \tanh(W_c[h_{t-1}, x_t] + b_c)
   \] (2)

Figure 4 The get gate layer in LSTM

\[
C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t
\]

\[
h_t = \sigma(W_h[C_t] + b_h)
\] (3)
(3) The current state combines past and current recorded information data:
\[ C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \]  

Figure 5 input gate layer in LSTM

(4) output gate layer: The output gate selects the result to be output.
\[ O_t = \sigma(W_o[h_{t-1}, x_t] + h_t) \]
\[ h_t = O_t * \tanh(C_t) \]

Figure 6 the current state in LSTM

(5) output gate layer: The output gate selects the result to be output.
\[ \sigma(tanh(C_t)) \]

Figure 7 output gate layer in LSTM

3.2 MobileNet-SSD Object Detection Model

We use the MobileNet-SSD network model with higher recognition efficiency for object detection. There are two main reasons for this:

1) Object detection has made important progress in recent years. The mainstream algorithms are mainly divided into two types: (1) two-stage method: such as the R-CNN series method, the main idea is to first generate candidate frames through selective search. Then, the candidate frames are classified and the frame is returned, and only one type of image of the background is discarded, and the image output of the target is included, and the target candidate frame is selected by the non-maximum suppression method. Although this method has high precision, it is difficult and slow to train. (2) End to end methods: such as YOLO and SSD, the main idea is to divide the image into several grids and target. Each grid is labeled separately. Different scales and aspect ratios can be used for sampling, and then the features are extracted and classified by CNN. The whole process only needs one step and the speed is faster. Therefore, this type of approach is adopted in engineering.

2) MobileNet is an efficient model recently proposed by Google for mobile terminals and embedded devices [13, 14]. MobileNet proposes a new convolution method called depthwise separable convolutions to construct a lightweight deep neural network [14]. The new convolution method extracts the planar features and channel features of the data by decomposing the standard convolution operations into two new convolution operations, as shown in Figure 8 below:
In addition, two adjustment parameters are introduced to adjust the size of the model and the number of parameters to achieve a balanced state in terms of accuracy and efficiency. By comparing it with multiple parameter quantities and compared with other advanced models in ImageNet classification task, it shows powerful performance. It also verifies the excellent performance of the model in object detection, face recognition, large-scale geolocation and so on. Among them, in the object detection task shown in figure 9, The MobileNet-SSD model is based on an improved detection algorithm model based on MobileNet and SSD, which also has better performance, as shown in Table 1 below:

| Framework Resolution | Model          | MAP   | Billion Mult-Adds | Million Parameters |
|----------------------|----------------|-------|-------------------|--------------------|
| SSD 300              | deeplab-VGG    | 21.2% | 34.9              | 33.1               |
|                      | Inception-V2   | 22.0% | 3.8               | 13.7               |
|                      | MobileNet      | 19.3% | 1.2               | 6.8                |
| Faster-RCNN300       | VGG            | 22.9% | 64.3              | 138.5              |
|                      | Inception V2   | 15.4% | 118.2             | 13.3               |
|                      | MobileNet      | 16.4% | 25.2              | 6.1                |
| Faster-RCNN600       | VGG            | 25.7% | 149.6             | 138.5              |
|                      | Inception V2   | 21.9% | 129.6             | 13.3               |
|                      | MobileNet      | 19.8% | 30.5              | 6.1                |
The MobileNet-SSD network has a smaller amount of computation and fewer parameters with little difference in accuracy [15]. This feature makes it more deployable on the mobile side and performs target detection tasks locally without the need for networking functions such as cloud services, which will reduce the computing power of the entire road network system. Figure 10 is an example of acquiring real-time data of pedestrian traffic using the model.

After removing the global average pooling, full connection layer and Softmax layer of MobileNet v1, and combined with SSD model structure, the improved MobileNet-SSD model structure is shown in the figure below:
3.3 Decision Control Model
First, in order to facilitate the discussion of the impact of pedestrian traffic on traffic signal decisions, we exclude vehicle factors by default. We define arbitrary standard unit time length $T$, the traffic threshold is $r$. Expected to control the flow rate $V$. The pedestrian flow result predicted by the LSTM model during this period is $m$, the required signal passing time $t = \frac{m}{v}$ Meanwhile, if the real-time pedestrian flow of object detection is $n$ and the required signal passage time is $t2$ , when $|nT - m| > r$, start the calculation of the new signal switching time. If $nT > m$, then $t2 = \frac{nT - m}{v}$, the total length of time is $t = t1 - t2$. To adjust the magnitude of the impact of historical and real-time data on our decision-making results, we define two adjustment factors $\alpha$ and $\beta$. So there is $t = \alpha \cdot t1 ± \beta \cdot t2$ Because During the $T$ time period, If we need $a$ times signal adjustments, the unit time of each adjustment signal is $s = \frac{t}{a}$

The algorithm flow is as follows:

Step1: Select the time period $T$, set the traffic threshold $r$.
Step2: Predict the traffic result $m$ through the LSTM model of the time period. Get real-time traffic by object detection.
Step3: When $|nT - m| > r$ is satisfied. The switching time calculation will be started: if the condition is not satisfied, the operations of step 1-step 2 will be repeated at fixed time intervals.
Step4: Enter the desired traffic $v$. Judge the size relationship between $nT$ and $m$, and respectively calculate $t1$, $t2$, $t$: $(t = \alpha \cdot t1 ± \beta \cdot t2)$
Step5: Through the given signal adjustment times $a$, the unit time length of each signal adjustment will be calculated, then we can set the signal switching time uniformly over the time period $T$.

![Decision Model](image)

Figure 12 Decision Model

4. ANALYSIS OF EXPERIMENT RESULTS
In this section, we present experimental results to demonstrate the effectiveness of the decision system performance.

4.1 Experimental Environment and Experimental Data

4.1.1 Experimental environment
The experimental environment uses the processor CPU: Intel (R) Core (TM) i5-4590 CPU @3.30GHz 3.30GHz, running memory RAM 8.00GB, operating system Windows10, 64-bit, python 3.6.5;

4.1.2 Experimental data
This data is taken from a traffic intersection in the commercial area near Lanzhou Jiaotong University, Anning District, Lanzhou City, Gansu Province, China. As it is located at the intersection of Lanzhou west railway station, it is an important transportation hub in Lanzhou, and the most prosperous
shopping center and commercial office building in Lanzhou, which attracts numerous pedestrians. Therefore, it has research value. By using statistical methods, the flow data of the same time period on Saturdays during consecutive weeks from June to August 2019 was recorded as a data set of the LSTM model. And take pictures and video as the data set of object detection model at the same time.

4.2 Experimental Results

4.2.1 LSTM prediction results

From June to August of 2019, the flow data of the same time period on Saturday for several consecutive weeks is taken as the data set of the LSTM model. The collected data is distributed as shown:

![Real traffic data distribution map](image13)

The data in figure 13 is divided into two parts of the training set and the test set according to the ratio of roughly 7:3, and then predicted by using the LSTM model, and the obtained result is as shown in the following figure. The yellow polyline in figure 14 is the prediction result of the training data, and the green polyline is the test result of the test set. In order to measure the validity of the prediction results, the regression evaluation indicators are calculated as follows:

![LSTM prediction results](image14)

- Root mean square error: 5.97
- Average absolute percentage error: 6.113516%
- Average absolute error: 4.758786
- Mean square error: 35.654878

By comparing image and evaluation data with existing mainstream prediction models [16], we can determine that the model can effectively predict the general trend of pedestrian flow. However, there is still some error, so we need to correct it by using the results of the target real-time detection model below.
4.2.2 MobileNet-SSD Detection Model Results
We obtained real-time images of the road through the camera, and detected and counted the crowds in the screen through the MobileNet-SSD model. The real-time recognition result is 13 person as shown below:

![Figure 15 target test results](image)

Through this model, we will obtain real-time real data, and then correct the prediction results of the LSTM model in the final decision model.

4.2.3 Decision Control Model
After obtaining the prediction of historical data and the results of real-time data, we need to enter the expected value \( v \) (pedestrian flow) through the decision model window, two adjustment factors \( \alpha, \beta \); and the number of signal adjustments \( a \) in the current time period, and then click the calculate button to calculate, to get the final decision result. The figure below is the result of our experiment in 15 minutes:

1) Adjust the desired flow rate to be greater than the current flow rate \( v \), set the adjustment factors to 0.65 and 0.35, respectively:

![Figure 16 Decision Model Calculation Interface (v=15)](image)

![Figure 17 decision results (v = 15)](image)

2) Adjust the desired flow rate to be less than the current flow rate \( v \), and set the adjustment factors to 0.5 and 0.5 respectively:
4.3 Analysis of Algorithm Results
In the above experiment, we conducted the experiment at the intersection for 15 minutes. The actual situation of the intersection is 7 signal conversions, and the transit time is about 40s. Through experiments, we adjusted the threshold of the desired flow rate to 10 person/second and 15 person/second. And adjust the unit passing time to 1.39 minutes and 0.6 minutes respectively. Under the premise of meeting the requirements of pedestrian flow, the single signal switching time is effectively adjusted.

5. CONCLUSION
This paper proposes a hybrid pedestrian traffic signal decision control model based on LSTM and MobileNet-SSD. The model is divided into three parts, named the processing of historical data, the processing of real-time data, and the final comprehensive decision processing. The LSTM model is used to analyze the historical data, and the real-time situation is obtained through the object detection model MobileNet-SSD. We combine the results of the two models and adjust the current signal switching interval through the decision model. So as to optimize the road traffic congestion phenomenon. The feasibility and effectiveness of the decision model are illustrated by experiments. In the following work, we will consider improving the accuracy and effectiveness of the experiment from the following aspects: 1) Improve the recognition efficiency of the target detection model for people and reduce the occurrence of missed detection. 2) Improve the LSTM model to improve the accuracy of prediction. 3) Add statistical-related methods to the decision-making model to improve the decision-making effect. 4) Take into account the factors of the vehicle's decision-making results. 5) Associate the influence relationship between the upper and lower sections and take the spatial factors between roads into account. 6) Deploy at actual traffic intersections, collecting more data for tuning, and collecting more exposed shortcomings for improvement.

REFERENCES
[1] Tao Zhiying. Research on some problems in public security group incident decision analysis and early warning system based on human flow [d]. 2014.
[2] Zhou Zhina, Shi Zhongke, Li Yingfeng. Decision Model of Pedestrian Group Red Light Behavior[j]. Systems Engineering - Theory & Practice, 2009, 29(11): 177-182.
[3] Chen Bingqu, Deng Tao. Research on Target Vehicle Detection Based on Improved Ssd Algorithm[J]. Journal of Chongqing University of Technology (Natural Science), 2019, 33(01): 64-69+135

[4] Ahmed, Mohammed S., and Allen R. Cook. Analysis of freeway traffic time-series data by using Box-Jenkins techniques. No. 722. 1979.

[5] Stephanedes Y J, Michalopoulos P G. Improved estimation of traffic flow for real-time control //Transportation Research Record 795. Washington DC: Transportation Research Board, 1981: 28–39

[6] Ledoux C. An urban traffic flow model integrating neural networks. Transportation Research Part C: Emerging Technologies, 1997, 5(5): 287–300.

[7] Yuan J, Fan B Q. VHSSA Model for Predicting Short-term Traffic Flow of Urban Road[J]. Journal of Highway & Transportation Research & Development, 2014, 8(4):90-96.

[8] Zhang Y, Zhang Y, Haghani A. A hybrid short-term traffic flow forecasting method based on spectral analysis and statistical volatility model[J]. Transportation Research Part C: Emerging Technologies, 2014, 43:65-78.

[9] Qian Wei, Yang Jinli, Yang Huihui, et al. Short-term traffic flow prediction based on combined model[j]. Computer Simulation, 2015, 32(2): 175-178.

[10] Wang Xiangxue, Xu Lunhui. Research on Short-Term Traffic Flow Prediction Based on Deep Learning[J]. Journal of Transportation Systems Engineering and Engineering, 2018, 18(1):81-88. WANG Xiang-xue, XU Lun-hui. Short-term Traffic Flow Prediction Based on Deep Learning [J]. Journal of Transportation Systems Engineering and Information Technology, 2018, 18(1): 81-88.

[11] Gang Y, Yunpeng W, Haiyang Y, et al. Short-Term Traffic State Prediction Based on the Spatiotemporal Features of Critical Road Sections[J]. Sensors, 2018, 18(7):2287-.

[12] Tan Miaomiao. Short-term traffic flow prediction based on lstm and grey model integration [d].2017.

[13] Canziani A, Paszke A, Culurciello E. An Analysis of Deep Neural Network Models for Practical Applications[J]. 2016.

[14] Howard A G, Zhu M, Chen B, et al. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications[J]. 2017.

[15] Tong Xing, Zhang Ji. Target Detection of ROS Platform Based on SSD_MobileNet Model[J]. Journal of Computer Systems, 2019, 28(01): 96-101.

[16] Qiao Songlin, Sun Rencheng, Liu Ji. Short-term traffic flow prediction based on deep learning[j]. Journal of Qingdao University (Natural Science), 2017, 30(4).