Campus safety monitoring system based on deep learning

Qianying Zhang ,Jin Hang,Boxuan Zhang,Mengzhong Wei
Faculty of Computer & Software Engineering , Huaiyin Institute of Technology ,Huai'an,China
1532202289@qq.com

Quanyin Zhu*,Xiang Li
Faculty of Computer & Software Engineering , Huaiyin Institute of Technology ,Huai'an,China
*Corresponding author's email:hyitzqy@126.com

Abstract— **In view of the safety hazards existing on campus, a campus safety detection system based on deep learning behavior recognition, license plate recognition and speed detection is proposed and designed. The design divides the campus safety monitoring system into three functional modules: hazard behavior identification, license plate tracking and vehicle speed measurement, and on this basis, designs the database of storage information and the warning system of speed too fast to realize the monitoring function of campus safety.**

Keywords ***Deep learning; Campus safety; Behavior recognition***

INTRODUCTION

In recent years, with the national attention to campus safety, campus security management has been increasing. At present, the safety and security work of our campus mainly depends on security personnel, although play a certain effect, but there are also many problems, such as in the course of campus safety management, there are security personnel can not find the students' dangerous behavior in the first time, can not identify the speed. There is a security risk. Therefore, the traditional supervision of campus safety through manpower is not enough. Therefore, it is necessary to develop a safety monitoring system for the identification of dangerous behavior and speed detection on campus.

This project is a campus safety monitoring system that integrates hazardous behavior identification, license plate tracking, and vehicle speed measurement.

RELATED WORK

A. Vehicle speed detection algorithm based on virtual coil

The vehicle speed detection [1] algorithm based on virtual coil is mainly divided into several processes. (1) Read traffic video: read the video using MATLAB platform [2].

(2) Set virtual coil: use the manually set virtual coil. Because this coil is flexible, it can be easily displayed on the first frame image.

(3) Calculate the gray difference between virtual coils VL1 and VL2: calculate the gray values of initial virtual coils VL1 and VL2, which are L1 and L2 respectively; Then calculate the gray values of virtual coils vl3 and vl4 in the next frame, which are L3 and L4 respectively. Thus, the VL gray difference m and N between frames can be determined, that is:

$$M = L3 - L1,\ N = L4 - L2.$$

Then, two sets of values can be obtained by using the cyclic frame difference, which are stored in an array, and the absolute value of the difference is used for statistics. Therefore, the method of difference between each subsequent frame and the first frame can be used. (4)Virtual coil control: read in each traffic video, draw the two groups of values and the ordinal number of the video frame into a two-dimensional image and analyze it. According to the jump change of the gray difference, get the corresponding gray threshold, calculate the time length of the vehicle passing through the virtual coil, and finally calculate the speed of the vehicle.

B. License plate detection

Efficient Det is used as the vehicle detection network[3], because the detection speed and accuracy of the network are well performed on the ImageNet data set.

Considering the different imaging angles of license plate in real environment, a distortion corrected license plate detection network is proposed. The network can detect different degrees of distorted license plates in natural scenes, and correct the distorted license plates to a rectangular shape similar to the front view. In order to locate the license plate more accurately, a network based on residual block is designed to extract the license plate features, and the convolution layer in the network is set to 3x3. In order to prevent the license plate feature information from disappearing after pooling, only four steps of 2 are used × 2, which reduces the input dimension by 16 times. Finally, as shown in the figure, the detection module has two parallel convolution layers: (1) one is used to embed the probability, and the function is activated as softmax [4] (2) The other is used to regress affine parameters. In order to extract the features on the deformed license plate, first consider a hypothetical square whose size is fixed at the center of the license plate (x, y). If the target probability of the unit is greater than the given detection threshold, the affine matrix is established by using the regression parameters to convert the virtual square into the license plate area[5]. Therefore, the license plate can be easily extended to horizontally and vertically aligned objects. Use Mi = [Xi, Yi] t (where I = 1, 2, 3, 4) to represent the four corners of the annotated license plate, which are marked clockwise from the upper left corner; n1=[-0.5， 0.5]T， n2=[0.5， -0.5]T， n3=[0.5, 0.5]T， n4=[-0.5，0.5]T represent the corresponding vertex of the standard unit centered on the origin. The height h, width W

2473-3636/21/$31.00 ©2021 IEEE
DOI 10.1109/DCABES52998.2021.00036

and network step of the input image are given through Ns = 24 (four maximum pool layers), and the network output characteristic diagram is given by P × Q It consists of 8 volumes, where p = H / N and q = w / NS. For each point unit in the characteristic graph, there are 8 values to be estimated: the first 2 values (V1-V2) are the probability of object / non object, and the last 6 values (v3-v8) are used to establish the local affine transformation TXY, which is obtained by equation (1):

$$T_{xy}(n) = \begin{bmatrix} \max(v_3, 0) & v_4 \\ v_5 & \max(v_6, 0) \end{bmatrix} n + \begin{bmatrix} v_4 \\ v_5 \end{bmatrix} \quad (1)$$

Txy(n) uses the learned parameters, the bounding box representing the predicted license plate position is affine transformed. The maximum functions V3 and V6 are used to ensure that the diagonal is positive (avoiding unnecessary mirroring or excessive rotation). In order to match the output resolution of the network, the MX points are rescaled according to the reciprocal of the network step, and re centered according to each (x, y) in the feature map. By using the standardized function (2):

$$A_{xy} = \frac{1}{\beta}\left(\frac{1}{N_s} \mathrm{m} - \begin{bmatrix} x \\ y \end{bmatrix}\right) \quad (2)$$

Among them, β Is a scaling constant that represents the edges of a virtual square. Set up β= 8.15 is the mean point between the maximum and minimum dimensions of the candidate license plate in the augmented training data segmented by the network stride [6]. In order to maintain a good compromise between accuracy and processing time, the maximum dimension 508 and the minimum dimension 128 are selected in this paper. Assuming that there is an object (license plate) at point (x, y), the first part of the loss function considers the error between the regular square of a distorted version and the standardized marking points of the license plate, which is given by equation (3):

$$f_{affine}(x, y) = \sum_{k=1}^{4} \|T_{xy}(n_k) \quad A_{xy}(m_k)\| \quad (3)$$

The license plates with tilt or even distortion is corrected to the front license plate by the distortion correction network, and the output license plate is used as the input of license plate recognition. In this way, the corrected license plate makes license plate recognition extremely easy, without considering the external influence such as angle, distortion or illumination, which improves the speed and accuracy of license plate recognition.

C. Extraction and filtering of behavioral feature

The first prerequisite for using the campus monitoring system to identify dangerous behavior of campus personnel is to be able to extract the behavior characteristics of the tested personnel, and then send the extracted behavior characteristics into the trained model to classify, obtain the basic classification results, and determine whether it is dangerous behavior. However, influenced by the dynamic change, lighting change, background clutter, time change and so on, it is the first problem to be solved to extract the behavior characteristics of the person being tested from the complex background. Unlike analytical images, video analytics have a larger amount of data and more complex conditions to analyze. Feature extraction can reduce the data dimension at the same time, analyze effective information, filter invalid information. There are several kinds of behavioral analysis feature extraction algorithms: shape feature, motion feature, space-time feature and mixed feature. The problem of features is mainly the features of time domain and frequency domain, which are extracted by sliding window. The sliding window contains two variables：size（size of window）、step（slip step）。Index times of window size 2 are generally selected in practical applications，$size = 2^{ceil(log2(2*f))}$, where f represents the sampling frequency of the sensor, and ai represents the combined acceleration. The calculation method is given by formula (1).

$$a_i = \sqrt{(a_i^x)^2 + (a_i^y)^2 + (a_i^z)^2} \quad (1)$$

In the time domain characteristics, $mean = \frac{1}{n}\sum_{n=1}^{n}(a_i - mean)^2$ represents the mean， $std = \sqrt{\frac{1}{n-1}}$ represents standard deviation， mod represents the mode， $max = \max(a_i)$ represents the majority， $min = \min(a_i)$ represents a minimum, $range = |max - min|$ represents the rang， $above\_mean = \sum_{n=1}^{n}\prod(a_i > mean)$ represents the mean point number of data, represents correlation coefficient，given by formula (2)：

$$\rho_{x,\ y} = \frac{\cos(x, y)}{\sigma_x \sigma_y} \quad (2)$$

SMA Represents the signal amplitude area, given by formula (3)：

$$SMA = \frac{\int_0^t |x(t)|dt + \int_0^t |y(t)|dt + \int_0^t |z(t)|dt}{t} \quad (3)$$

Because the scene of behavioral feature extraction often contains crowded people, resulting in a large number of noise that affects the recognition, therefore, a feature filtering method based on cover[7]is adopted. The construction of the mask is shown in Figure 1:

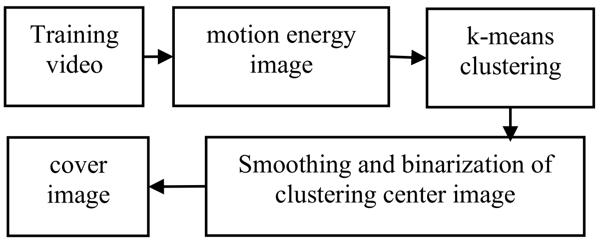

Figure 1. Construction process of shield

D. Classification of mask behavior

After feature filtering, behavior recognition becomes easier. The process of behavior recognition is as follows:
(1) The dangerous behavior images are input and the data are marked in the network model training. The location

behavior information of the detected personnel predicted by the model is used as the input of the recognition model to train the recognition model.

(2) The minimum circumscribed rectangle of the image to be judged is expanded in the left and right directions. The open window time length is T (representing the continuous time of the behavior). The video motion image MEI, motion history image MHI and optical flow energy image FEI are calculated in the window.

(3) The matching mask image is found from the behavioral mask of the training results, and the mask is placed on the detection window, MHI and FEI are retained, and the non-coverage area is removed to realize feature filtering.

(4) The behavioral characteristics are classified by the method based on trainers, so as to judge the behavioral characteristics.

Feature classification has two methods: direct classification and time domain state space fusion model. Dynamic Time Planning Adjustment (DTW) [8]is similar to the randomness of speech signal. The action behavior of different people in different scenes in video is different, and there is no exactly the same time length. Therefore, the time axis should be distorted and bent unevenly when matching with the established recognition model. Align video features with template features. DTW is essentially a time warping function that satisfies certain conditions to describe the time correspondence between the input model and the training model, so as to find the minimum cumulative corresponding warping function when two templates match.

REFERENCES

[1]Seda Kul, Isabek Tashiev, Ali Sentas, Ahmet Sayar:Event-Based Microservices With Apache Kafka Streams: A Real-Time Vehicle Detection System Based on Type, Color, and Speed Attributes. IEEE Access 9: 83137-83148 (2021)

[2]Liyan Zhao, Aize She:Design of Virtual Experiment Platform Based on MATLAB Simulation Technology. DPTA 2019: 335-341

[3]Riadh Ayachi, Mouna Afif, Yahia Said, Abdessalem Ben Abdelali:Drivers Fatigue Detection Using EfficientDet In Advanced Driver Assistance Systems. SSD 2021: 738-742

[4]Zhang Fan, Wang Xiaodong, Hao Xianpeng.Intelligent vehicle character recognition based on edge features [J].Automation and instrumentation, 2020 (6) : 11-4,20.

[5]Hae-Chan Noh, Jae-Pil Heo:Mutually Orthogonal Softmax Axes for Cross-Domain Retrieval. IEEE Access 8: 56491-56500 (2020)

[6] Caicai Li, Hao Jin, Huihui Bai:Ratio Test for Time-Varying Variance of Mean Change Point Based on Big Data Analysis. CSIA (1) 2021: 369-376

[7]Guo Ping.Video-based human behavior analysis.2012 Beijing Jiaotong University, PhD paper.

[8]Li Dongdong, Zhang Limin, Deng Xiangyang, Jiang Jie. Dynamic gesture recognition based on multi-feature fusion [J]. Computer applications and software, 2021,38 (08) : 214-219.