ISSN: 2688-9323

Research on Road Target Image Processing and Target Detection Method

Lu Wang[1, 2]

[1]Department of Electrical Automation, Shanghai Maritime University, Shanghai, 201306, China

[2]Department of Image and Network Investigation, Railway Police College, Zhengzhou, 450053, China

Abstract

With the increasing complexity and diversification of the current road traffic environment, the road traffic safety situation is particularly serious. As an important means of road traffic monitoring, the effective detection of road traffic target has become an important research topic. Based on this background, in this paper, it first discusses the image graying and image filtering involved in the methods of image pre-processing, and on this basis, it further analyzes the related content of target detection method of the deep neural network , so as to provide a certain theoretical basis for the accurate detection of road target image.

Keywords

Road Target; Image Graying; Image Filtering; Candidate Window; Deep Neural Network.

1. Introduction

With the further development and progress of social economy, people's living standard has been greatly improved. As a more frequent and convenient means of transportation, vehicles has become an important part of people's life and social production. However, with the popularization and use of vehicles, the number of them is also increasing day by day, which brings a certain degree of safety risks to the current road traffic safety[1,2,3]. The contradiction among pedestrians, vehicles and limited road traffic resources is gradually increasing with the increasing complexity of road traffic environment. Under the background of such road traffic situation, how to ensure the personal and property safety of each participant is an important research content in the field of road traffic. Therefore, the accurate detection of road target image has become a research hotspot under this background[4,5,6].

With the development of computer technology and artificial intelligence technology, target detection technology is gradually applied to road traffic applications[7,8,9,10]. Especially in the road traffic monitoring system with video and image as the main content, the target detection of road traffic image has become an important research direction. However, the current road traffic environment is complex and various, and the road traffic images obtained by the relevant monitoring equipment also have various kinds of noise interference. Thus ,how to accurately extract the road traffic target from these complex road traffic images is a difficult point in target detection of road traffic image.

In this paper, aiming at the difficulties of accurate detection of road traffic targets, and taking the current related technologies of image processing and target detection as the background, it analyzes and discusses the relevant technical methods of image data preprocessing, and based on the deep convolution neural network, it studies the basic network structure model and the common detection framework which can be applied to the road traffic target detection, and

ISSN: 2688-9323

then it could provide some theoretical basis for the next step of accurate detection for road traffic target.

2. Image Data Preprocessing

The data preprocessing of road target image is to achieve better effect of classification, prediction, recognition and image target detection for road target. Therefore, preprocessing and operation are needed before road target image recognition and detection. The preprocessing of image data is mainly because of the massive data in the current natural world, many of which cannot be directly analyzed and mined. In the data acquisition process, these data are affected by many interference factors, making the data itself often with noise, meanwhile, the data quality is not high, which affects the training of network model and cannot achieve satisfactory results. In addition, the data entered into the network model for training need to have the same data format. Therefore, for the selected data set, it is more necessary to preprocess the data first, remove the noise in the data, and then convert the data into a unified data format that can be easily used by the network. Thus, it can make the data itself have better availability, and provide rich data basis for the effective extraction, classification, recognition and detection of data features in the network model.

2.1. Image Graying

In a color image, the color of each pixel is composed of three component: red, green and blue (namely, R, G and B), and the process of transforming a color image with R, G and B into a gray image with only one component is called image graying. In the actual representation, each component can take one of the 256 gray values, and the range of these values is 0-255, in which 0 is black, and its color is the darkest, while 255 is white, and its color is the brightest. and the value between 0 and 255 is the transition representation from black to white. The gray image is a special representation of the color image, therefore, in the process of digital image processing, while the color image with three components is converted into a gray image with only one component, the computation, the time and space complexity of image processing are relatively simplified. In addition, there is no difference between the gray image and the color image in describing the image features, and the gray image itself can still reflect the overall and local color and brightness characteristics of the whole image. Therefore, in many image processing scenes, the image graying will be operated first. As shown in Formula (1), it represents the pixel information of a color image, and each pixel contains the gray values of R, G and B color components:

$$\begin{bmatrix} 121,65,33 & 13,128,171 & 8,105,73 \\ 56,17,216 & 189,33,203 & \vdots \\ 79,135,11 & \cdots & \ddots \end{bmatrix} \tag{1}$$

Formula (2) represents the pixel information of a gray image, and it can be seen that each pixel of the gray image contains only one gray value:

$$\begin{bmatrix} 19 & 156 & 0 \\ 76 & 261 & \vdots \\ 37 & \cdots & \ddots \end{bmatrix} \tag{2}$$

It should be noted that the color images used and involved in the current information world are mainly composed of three components of RGB mode, therefore, when processing color images, the three components of RGB should be processed respectively. However, the principle of RGB

ISSN: 2688-9323

color pattern is only the combination of color based on optical characteristics, and it cannot truly reflect the morphological characteristics of color image itself. Therefore, based on the limitations of RGB, in the process of color image processing, the color image will be grayed first in the process of image preprocessing. Image graying can effectively reduce the redundant information of the image itself, but on the other hand, the graying process itself is irreversible.

To sum up, the commonly used image gray processing methods mainly include component method, average value method, maximum value method and weighted average method.

(1) component method

The component method means that only one of the three component values of R, G and B of each pixel is taken as the gray value of gray image after graying, and the expression is as follows:

$$\text{Gray} = \text{R or Gray} = \text{G or Gray} = \text{B} \tag{3}$$

(2) average value method

The average value method is to calculate the average value of the R, G and B components of each pixel in the color image, and take the value as the gray value of the gray image. The expression of the average value method is as follows:

$$\text{Gray} = (\text{R} + \text{G} + \text{B})/3 \tag{4}$$

(3) maximum value method

The maximum value method is to take the maximum value of the R, G and B components of each pixel in the color image, and take the value as the gray value of the gray image, and the expression of the maximum value method is as follows:

$$\text{Gray} = \text{Max}(\text{R}, \text{G}, \text{B}) \tag{5}$$

(4) weighted average method

The weighted average method refers to weighting the values of R, G and B of each pixel in the color image, and then taking the weighted value as the gray value of gray image after graying processing. The expression is as follows:

$$\text{Gray} = \text{W}_\text{R} * \text{R} + \text{W}_\text{G} * \text{G} + \text{W}_\text{B} * \text{B} \tag{6}$$

where W_R, W_G and W_B is the weight coefficient of R, G and B respectively.

According to a large number of related studies, when the values of W_R, W_G and W_B are 0.299, 0.587 and 0.114 respectively, the gray-scale image converted by the color image is the most consistent with human color perception, thus, the expression can be expressed as follows:

$$G\text{ray} = 0.299 * \text{R} + 0.587 * \text{G} + 0.114 * \text{B} \tag{7}$$

2.2. Image Filtering

In the current acquired video images, there are some noises inevitably among them, and the key to the performance of target detection is to eliminate or suppress these noises, and then improve the overall performance of target detection by improving the expression quality of the image itself. The process of eliminating or suppressing image noise can be called image filtering.

ISSN: 2688-9323

The commonly used image filtering methods include mean filtering[11], median filtering[12] and Gaussian filtering[13].

(1) Mean filtering method

The mean filtering method is a filtering method to smooth the image in the local area of the image space, and the operation essence is to get the average of all the pixel values in a local area of the image space, and then get the new image gray value. The commonly used 3×3 means filter can be expressed as follows:

$$\frac{1}{9}\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix} \tag{8}$$

The mean filter is also known as eight neighborhood mean filter, and according to the different selection requirements of mean filter, in addition to 3×3 filter, the commonly used filters are also include 5×5 mean filter, 7×7 mean filter, etc. The mean filtering of image can effectively remove the noise and make the gray value of the filtered image more uniform than that before filtering.

(2) Median filtering method

Median filtering is also an effective filtering method to remove image noise. Different from mean filtering, median filtering does not take the average value of all pixels in a local space area of the image as the new gray value, but arranges all the pixel values in the region window from large to small (or from small to large), and selects the pixel value with the most middle value in the sorted pixel value list as the new gray value. The median filtering can effectively eliminate the image noise and retain more details in the image. The commonly used 3×3 median filter can be expressed as follows:

$$mid\begin{bmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{bmatrix} \tag{9}$$

(3) Gaussian filtering method

As a smoothing filtering method, Gaussian filtering is widely used in image denoising. Especially for the scene with complex image noise such as road traffic image, Gaussian filter also has its application space. In the specific implementation, the Gaussian filter uses the Gaussian filter kernel, the Gaussian filter is used to weighted the local spatial pixels in the image, and the value of each filtered pixel is obtained by weighted average between itself and the related pixels in the neighborhood. Similar to mean filter and median filter, Gaussian filter kernel is usually used to select the filter template with the number of odd. The commonly used 3×3 Gaussian filter and 5×5 Gaussian filter can be expressed as follows:

$$\frac{1}{16}\begin{bmatrix}1 & 2 & 1\\ 2 & 4 & 2\\ 1 & 2 & 1\end{bmatrix} \tag{10}$$

$$\frac{1}{273}\begin{bmatrix}1 & 4 & 7 & 4 & 1\\ 4 & 16 & 26 & 16 & 4\\ 7 & 26 & 41 & 26 & 7\\ 4 & 16 & 26 & 16 & 4\\ 1 & 4 & 7 & 4 & 1\end{bmatrix} \tag{11}$$

ISSN: 2688-9323

3. Target Detection Method

3.1. Deep Convolution Neural Network Model

The target detection methods in road traffic scene mainly include traditional target detection methods based on manual feature extraction and target detection methods of deep learning based on depth feature extraction. There are mainly Histogram of Oriented Gradient(HOG)[14], Haar[15], Scale-Invariant Feature Transform(SIFT)[16], Deformable Parts Model(DPM)[17] and so on. Combined with SVM[18], Softmax[19] and other classifiers, the extracted features can be further put into these classifiers to classify the detected target categories. In the process of feature extraction, these traditional target detection methods mainly rely on manually designed feature extraction methods, so the process of feature extraction is more complex. With the continuous development of deep learning technology, deep feature extraction method based on deep convolution neural network has become a research hotspot in the field of road traffic target detection. Therefore, in this section, it mainly takes the deep convolution neural network model as an example to introduce in detail.

The use of deep convolution neural network is rising with the deep network training model explored by Professor Geoffrey Hinton in 2006. In 2012, it won the champion of ImageNet through the constructed deep convolution neural network[20]. Since then, the research of deep convolution neural network in the field of image has become a hot spot.

The basic structure of deep convolution neural network is shown in Figure 1.

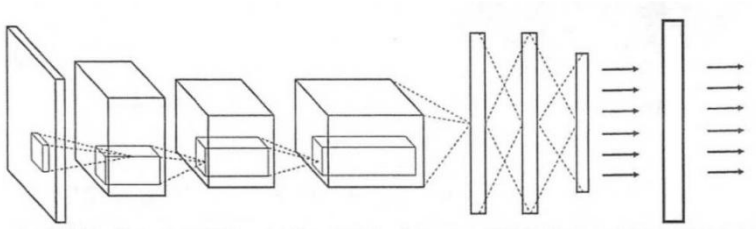

Figure 1. the basic structure of deep convolution neural network

In the deep convolution neural network model, the feature extraction is realized mainly by the operation of convolution and pooling.

In the process of convolution operation, convolution kernel is used to slide the convolution feature map with a specific step size, and then a new feature map is constructed. The feature map of each layer is convoluted in this way, which forms the structure of deep convolution neural network. The commonly used convolution kernel size is 3×3, and the two-dimensional convolution diagram is shown in Figure 2, which shows the convolution operation using 3×3 convolution.

$$\begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} & x_{15} \\ x_{21} & x_{22} & x_{23} & x_{24} & x_{25} \\ x_{31} & x_{32} & x_{33} & x_{34} & x_{35} \\ x_{41} & x_{42} & x_{43} & x_{44} & x_{45} \\ x_{51} & x_{52} & x_{53} & x_{54} & x_{55} \end{bmatrix} * \begin{bmatrix} w_{11} & w_{12} & w_{13} \\ w_{21} & w_{22} & w_{23} \\ w_{31} & w_{32} & w_{33} \end{bmatrix} = \begin{bmatrix} y_{11} & y_{12} & y_{13} \\ y_{21} & y_{22} & y_{23} \\ y_{31} & y_{32} & y_{33} \end{bmatrix}$$

Figure 2. two dimensional convolution diagram

ISSN: 2688-9323

According to the operation method of convolution kernel in convolution neural network, the convolution expression is expressed as follows:

$$y_{ij} = \sum_{m=1}^{3} \sum_{n=1}^{3} (x_{i+m-1,j+n-1} \times w_{mn}) \quad (12)$$

Pooling is a method to reduce the dimension of the convoluted feature map. Through pooling, convolution features are compressed and extracted properly, which can reduce redundancy and extract image features more effectively. There are two common pooling methods: max pooling and average pooling. Among them, max pooling refers to selecting the maximum value of the feature map as the output, and average pooling is to calculate the average value of all the values in the feature map and take it as the output. Pooling operation can effectively reduce the number of feature pixels; thus, it can effectively reduce the computational complexity of the network, and it can also speed up the learning and training of the network.

3.2. Deep Convolution Neural Network-based Target Detection Framework

In the current application of target detection in road traffic scene, the deep convolution neural network is mainly one-stage detection method and two-stage detection method. Among them, one-stage detection, mainly represented by the detection methods of Yolo, from Yolo V1 to Yolo V5, its core idea is to take the whole image as the input of the network, and it directly regress the position and category of bounding box in the output layer. Different from the one-stage detection method, the two-stage detection method is mainly represented by RCNN series of detection methods, mainly including RCNN, Fast RCNN, Faster RCNN, Mask RCNN, Cascade RCNN, and so on. These methods divide the detection results into two parts: object classification and object position regression.

Compared with the above two kinds of target detection methods, the one-stage detection runs fast, but its detection effect is not as good as the two-stage detection method; while the two-stage detection has higher detection accuracy, but its running time is also obviously not as good as the one-stage detection method. Therefore, it is a hot issue in the field of road traffic target detection to design a target detection model with deep neural network which can meet the requirements of both running time and detection accuracy by combining the advantages of the above two methods.

4. Conclusion

Under the current complex road traffic environment, the diversity and complexity of road traffic targets have brought severe challenges to road traffic safety. Therefore, how to detect various kinds of road traffic targets efficiently in complex road traffic scenes has become an important research direction in the field of current road traffic. In this context, taking the current road traffic image as the starting point, in this paper, it analyzes the related content of image preprocessing, and deeply discusses the image graying and image filtering involved in it. On this basis, the related target detection methods are discussed, including the related technical contents of the target detection methods represented by deep convolution neural network, which provides a theoretical reference for the further efficient detection of road target images.

Acknowledgements

This work is supported by Ministry of Public Security Technology Research Project of China (2019JSYJC25), Fundamental Research Funds for the Central Universities of China (2019TJJBKY012, 2020TJJBKY001), Henan Province Educational Commission Key Scientific Research Project of China (21B580005).

ISSN: 2688-9323

References

[1] Nogal M, Honfi D. Assessment of road traffic resilience assuming stochastic user behavior[J], Reliability Engineering & System Safety, 2019, 185(5):72-83.

[2] Garg K, Ramakrishnan N, Prakash A, et al. Rapid and Robust Background Modeling Technique for Low-Cost Road Traffic Surveillance Systems[J], IEEE Transactions on Intelligent Transportation Systems, 2019, PP(99):1-12.

[3] Zhu Q. Research on Road Traffic Situation Awareness System Based on Image Big Data[J], IEEE Intelligent Systems, 2020, 35(1):18-26.

[4] Song J G, WU Y. Improved Road Target Detection Algorithm Based on YOLOv2 Model[J], Software Guide, 2019, 18(12):126-129.

[5] LI H G, LU C Y, QI L. Road Target Detection Based on Otsu Multi-Threshold Segmentation[J], Mechanical Engineering and Control Systems, 2016: 265-269.

[6] Zhang H Y, Qin H B. FMCW Radar Moving Road Target Detection System with GPRS Communication[J], Technology of IoT & AI, 2018, 1(01):31-34+44.

[7] Prakash I, Neves O, Cumbe E, et al. The Financial Burden of Road Traffic Injuries in Mozambique: A Hospital-Related Cost-of-Illness Study of Maputo Central Hospital, World Journal of Surgery, 2019, 43(12):2959-2966.

[8] Zhao J W, Hua J, Liu Y T, et al. Research on Road Traffic Accident Information Collection Technology and Its Application, Journal of Chongqing University of Technology, 2019, 33(07):28-36.

[9] Sivaraman S, Trivedi M M. A General Active-Learning Framework for On-Road Vehicle Recognition and Tracking, IEEE Transactions on Intelligent Transportation System, 2016, 11(2):267-276.

[10] Cheon M, Lee W, Yoon C, et al. Vision-Based Vehicle Detection System With Consideration of the Detection Location, IEEE Transaction on Intelligent Transportation System, 2012, 13(3):1243-1252.

[11] Malek-Mohammadi M, Rojas C R, Wahlberg B. A Class of Nonconvex Penalties Preserving Overall Convexity in Optimization-Based Mean Filtering[J]. IEEE Transactions on Signal Processing, 2016, 64(24):6650-6664.

[12] Chen C, Ni J, Huang J. Blind Detection of Median Filtering in Digital Images: A Difference Domain Based Approach[J]. IEEE Transactions on Image Processing, 2013, 22(12):4699-4710.

[13] Manuel W, Sebastian T, Cristina G C, et al. A new perspective and extension of the Gaussian Filter[J]. International Journal of Robotics Research, 2017, 35(14): 1731-1749.

[14] Dalal N, Triggs B. Histograms of Oriented Gradients for Human Detection[C]. Computer Vision and Pattern Recognition, 2005 CVPR, 2005, San Diego, CA, USA. IEEE:886-993.

[15] Viola P, Jones M. Rapid object detection using a boosted cascade of simple features[C]. Computer Vision and Pattern Recognition, 2001 CVPR, 2001, Kauai, HI, USA, IEEE:1-11.

[16] Lowe D G. Object Recognition from Local Scale-Invariant Feature[C]. Computer Vision, 2002 The Proceedings of the Seventh IEEE International Conference, Kerkyra, Greece. IEEE,2:1150-1157.

[17] Felzenszwalb, Pedro, F, et al. Object Detection with Discriminatively Trained Part-Based Models.[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2010, 32(9):1627-1645.

[18] Ukil A. Support Vector Machine[J]. Computer Science, 2002, 1(4):1-28.

[19] Bouchard G. Clustering and Classification Employing Softmax Function Including Efficient Bounds: us, 8065246[P]. 2011-11-22.

[20] Krizhevsky A, Sutskever I & Hinton G E. Imagenet classification with deep convolutional neural networks[C]. Advances in Neural Information Processing Systems (NIPS). 2012: 1097-1105.