Advances in Artificial Intelligence, Big Data and Algorithms
G. Grigoras and P. Lorenz (Eds.)

doi:10.3233/FAIA230852

Research on Improved YOLOv7 Based Vehicle Speed and Traffic Flow Detection System

An PING, Yanfeng XUE[1], Bohao YAO, Shihao XIAO, Xiang ZHANG, Liming LI
Deptof Computer & Science Technology, Lyuliang Univetsity, 033000, Shanxi, China

Abstract. The detection of speed and traffic flow is one of the important issues in the current smart city construction and intelligent transportation development process. The automatic detection of speed and traffic flow through video information in road traffic is conducive to optimising road traffic and transport management, strengthening urban management and improving the efficiency of urban transport and services. Therefore, the study of vehicle speed and traffic flow detection systems is of great practical value. To address the issue of urban road transport management, an improved YOLOv7-based vehicle speed and traffic flow detection system was designed and implemented based on traffic video data, and the system was validated and analysed based on the video data.

Keywords. Speed detection; Traffic flow detection; YOLOv7; Attentional mechanisms; Target tracking

1. Introduction

With the development of technology, China's urbanisation has accelerated considerably, but urban road congestion, traffic accidents and energy shortages are also becoming more and more prominent. The negative impact of this series of problems has seriously affected the improvement of people's living standards and the further development of the city. China spends a lot of manpower and financial resources on the transport system every year. With the increasing demand for transport, improving transport capacity and strengthening urban transport infrastructure has become an urgent problem in China's economic construction. The research and application of speed and traffic flow detection systems can effectively enhance urban transport management, improve urban planning and scheduling capabilities and further increase the efficiency of urban production and services [1]. In the urban traffic planning sector, the commonly used traditional traffic flow detection methods mainly include air duct detection technology and magnetic induction detection technology and other methods, but these methods are not robust and their detection effects are easily affected by external environmental factors, which are no longer able to adapt to the needs of the times of automation and intelligence in the context of smart cities [2].

[1] Corresponding Author: Yanfeng XUE, Deptof Computer & Science Technology, Lyuliang Univetsity; Email: m19135807310@163.com

2. YOLOv7 Improvements that add the BiFormer Attention Mechanism

2.1. Input

The Input module of YOLOv7 consists of three main parts: Mosaic data enhancement, adaptive anchor frame calculation and adaptive image scaling. The Mosaic data enhancement is performed by randomly scaling, cropping and laying out any four images selected from the dataset and stitching them together into a new image for training. Mosaic data enhancement can effectively enrich the detection target background, improve the detection of small targets and further reduce the Mini-batch, which can reduce the training cost of the network. The adaptive anchor calculation is mainly carried out by genetic algorithm and K-Means algorithm to optimise the initial anchor of the dataset, which can effectively and adaptively calculate the best anchor of the dataset and avoid the extra process of anchor calculation before the network training [3].

2.2. Head

The Head module of YOLOv7 consists mainly of a PAFPN structure. For the final output of the backbone, the 32-fold downsampled feature map C5, the number of channels is changed from 1024 to 512 after SPPCSP processing [4]. Then fuse with C4 and C3 according to top down to get P3, P4 and P5; then fuse with P4 and P5 according to bottom-up. The process is shown in Figure 1.

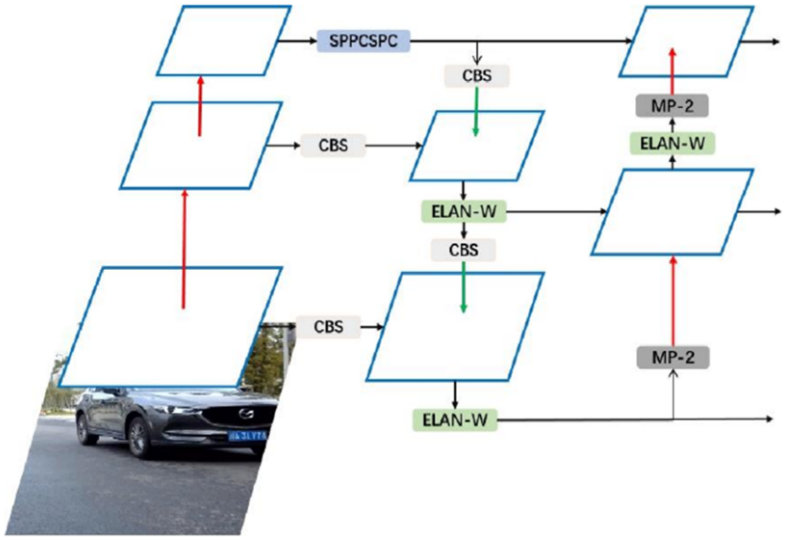

Figure 1. PAFPN flow chart

2.3. Bi-Level Routing Attention (BRA) module

To alleviate the scalability problem of Multi-Head Self-Attention (MHSA) [5], some related researchers have proposed different sparse attention mechanisms in which each query focuses on only a small number of key-value pairs instead of all of them [6]. The process is shown in Figure 2.

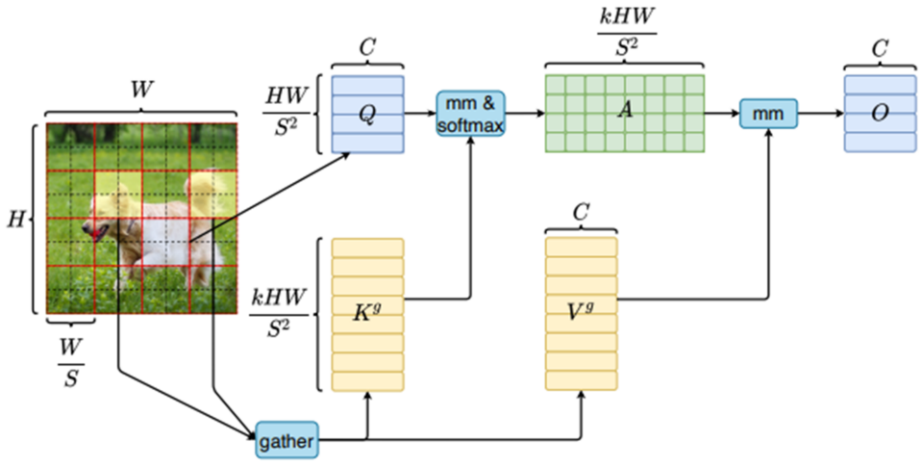

Figure 2. BRA flow chart

As shown in Figure 2, the BRA module saves the number of parameters and computation by collecting the key-value pairs in the first k relevant windows and using sparsity operations to skip the computation of the least relevant regions directly.

3. Vehicle Speed and Traffic Detection Based on DeepSORT Algorithm

3.1. Overview of the DeepSORT Multi-Objective Tracking Algorithm

The detailed flow of the DeepSORT algorithm is basic process mainly consists of two processes: Kalman filter prediction update and data association matching.

The main input of the DeepSORT algorithm is the target detection result, including the detection box BoundingBox, the confidence score and the current RGB image feature information, the confidence score is mainly used to filter the detection box, and the detection box and image feature information are used for the matching calculation with the tracker [7].

(1). Kalman filter prediction

In the prediction phase, the algorithm takes the predicted value $\hat{x}_{k-1}$ of the previous moment ($t = k-1$) to predict the state $\hat{x}_k$ of the current moment ($t = k$) and estimates the error covariance matrix P_k^- such that the error covariance matrix is minimized. The basic Kalman filter equation is shown in equations (1) and (2).

$$\hat{x}_k = A \cdot \hat{x}_{k-1} + \Gamma_{k-1} W_{k-1} \tag{1}$$

$$Z_k = H_k \cdot \hat{x}_k + v_k \tag{2}$$

Where equation (1) is the equation of state and equation (2) is the equation of measurement. A is the state transfer matrix, W_{k-1} is the process noise, Γ_{k-1} is the noise driven matrix, Z_k is the observed quantity, H_k is the observation matrix and v_k is the measurement noise.

In DeepSORT target tracking, the tracking target centre point coordinates $x_x(k)$、$x_y(k)$ and the corresponding x- and y-axis velocities $v_x(k)$、$v_y(k)$ need to be predicted. Due to the short time interval between adjacent image frames, the vehicle motion can be approximated as uniform linear motion, so the state of the vehicle target at the current moment $(t = k)$ can be defined as $\hat{x}_k^- = [x_x(k)、x_y(k)、v_x(k)、v_y(k)]$ according to these four state quantities, whose corresponding Kalman equations are shown in equations (3) to (6).

$$\hat{x}_k = A \cdot \hat{x}_{k-1} + B \cdot \mu_{k-1} + \omega_{k-1} \tag{3}$$

$$\hat{x}_k = \left(x_x(k)、x_y(k)、v_x(k)、v_y(k)\right) \tag{4}$$

$$Z_k = H_k \cdot \hat{x}_k + v_k \tag{5}$$

$$Z_k = \left(x_x(k)、x_y(k)\right)^T \tag{6}$$

where $\hat{x}_k$ is the vehicle target state vector at the current moment $(t = k)$, A and B are the state transfer matrices of the system, μ_{k-1} are the control variables of the system, ω_{k-1} is the covariance matrix of Gaussian noise with zero mean, and Z_k、H_k、v_k denotes the observation, observation matrix and measurement noise respectively. The Kalman filter corresponds to the covariance matrix update shown in equation (7).

$$P(x_k|x_{k-1}) = A \cdot P(x_{k-1}|x_{k-1})A^T + Q \tag{7}$$

where $P(x_k|x_{k-1})$ is the covariance matrix corresponding to the current moment and $P(x_{k-1}|x_{k-1})$ is the system covariance matrix for the previous moment $(t = k - 1)$.

(2) Data association matching

After obtaining the target detection frame and the Kalman filtered target prediction frame, the DeepSORT algorithm uses cascade level matching and IOU matching algorithms to correlate the data, matching the target detection frame with the prediction frame. The data association process is shown in Figure 3.

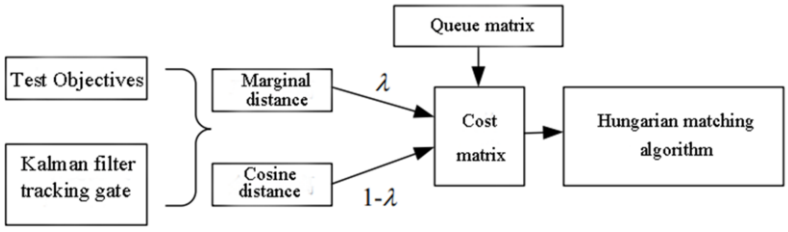

Figure 3. DeepSORT data association process

In the process of data association, DeepSORT calculates the overall similarity between the target and prediction frames in the system using the weighted sum of the martingale and cosine distances, calculated as shown in equation (8).

$$c_{ij} = \lambda d_1(i,j) + (1 - \lambda) d_2(i,j) \tag{8}$$

where c_{ij} is the overall similarity between the ith prediction frame and the jth detection frame, $d_1(i,j)$ is the corresponding martingale distance between the two target frames, λ is their corresponding weight values, and $d_2(i,j)$ is the cosine distance.

3.2. Vehicle Speed Detection Based on Multi-Objective Tracking Results

After obtaining the multi-vehicle target tracking results in the video stream, this paper implements real-time vehicle speed detection based on the perspective transformation relationship between the image pixel plane and the actual road plane. The principle of the perspective transformation is shown in Figure 4.

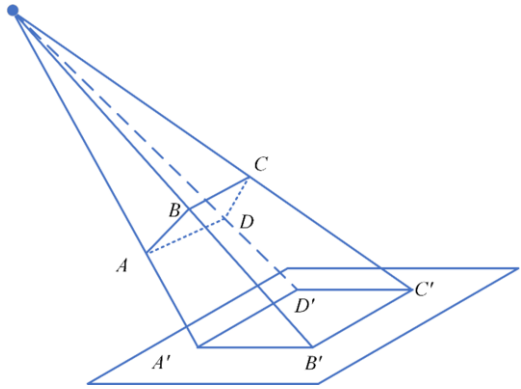

Figure 4. Perspective transformation schematic

In the figure, A, B, C and D are the coordinate points on the original image plane, A'、B'、C'、D' corresponding to the coordinate points of the target image plane. The perspective transformation relationship between a point $P(X,Y)$ on the original image plane and a point $P'(X',Y')$ on the target image plane is shown in equation (9).

$$\begin{bmatrix} X' \\ Y' \\ 1 \end{bmatrix} = H \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{pmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \tag{9}$$

As shown in the above equation, H is the perspective transformation matrix from the original plane to the target plane, also called the single-strain matrix. The single-response matrix can describe the transformation relationship between two planes in terms of rotation, scaling, shearing, translation, perspective, etc. Its degrees of freedom are 8, so that it can be made $h_{33} = 1$, indicating that scaling at any scale can be performed.

4. Conclusion

In this paper, we apply the improved YOLOv7 algorithm with BiFormer attention mechanism to detect moving vehicles and combine it with DeepSORT algorithm to achieve multi-objective tracking, vehicle speed measurement and traffic flow detection in urban transport management [8]. There are still some issues in the development of the system that need to be further addressed as follows:

The addition of attention mechanisms can be explored in part by trying to add faster attention modules and the crossover of multiple attention modules in search of better detection speed and detection accuracy [9].

For the vehicle speed measurement part, the current method relies on a conversion relationship between the pixel distance and the actual distance, and due to the varying size of the picture in terms of distance and proximity, this method can lead to objects in the distance going even faster, so a more accurate method of measuring vehicle speed is needed [10].

Acknowledgements

Innovation Project of Higher Education Teaching Reform in Shanxi (No. J20221164).

Research of Technological Important Programs in the city of Lüliang, China (No. 2022GXYF18).

References

[1] Chen Jiaqian,Jin Yanhong,Wang Wenyuan,et al. Traffic flow detection based on YOLOv3 and DeepSort[J]. Journal of Metrology,2021,42(6):718-723.

[2] Hu Yunlu. Research on video-based traffic flow and speed detection system [D]. Shaanxi:Chang'an University,2017.

[3] Li Kun. Design of a vehicle flow detection system based on magnetoresistive sensors [J]. Measurement and control technology,2019,38(1):114-116,144.

[4] DALAI, N, TRIGGS, B. Histograms of oriented gradients for human detection[C]. Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on vol.1. 2005:886-893.

[5] LIENHART, R., MAYDT, J.. An extended set of Haar-like features for rapid object detection[C]. Image Processing. 2002. Proceedings. 2002 International Conference on. 2002:I.900-I.903.

[6] TONG ZHANG. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods[J]. AI magazine: Artificial intelligence,2001,22(2):103-104.

[7] Strobl, Carolin;Malley, James;Tutz, Gerhard.An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests.[J].Psychological Methods,2009,Vol.14(4): 323-348.

[8] Freund, Yoav, Schapire, Robert E. A decision-theoretic generalization of on-line learning and an application to boosting[J]. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 1995, Vol.904(1): 23-37.

[9] Luo Huilan, Chen Hongkun. A review of deep learning-based target detection research[J]. Journal of Electronics,2020,48(6):1230-1239.

[10] Li Kequan, Chen Yan, Liu Jia Chen, et al. A review of deep learning-based target detection algorithms[J]. Computer Engineering,2022,48(7): 1-12.