International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 05 | May - 2023 Impact Factor: 8.176 ISSN: 2582-3930

Vehicle Counting and Detection

Vaishnavi Tiwari (19SCSE1010428)
Anadi Singh (19SCSE1010745)
in vision of Dr. Arvind Kumar

Abstract

Intelligent vehicle identification and computing has become an important factor in highway management. However, due to the difference between vehicles, their search remains a problem that directly affects the accuracy of vehicle counting. To solve this problem, this article presents an image-based vehicle tracking and counting system. The study published new information about highways with 57,290 descriptions in 11,129 images. Compared to the available publicly available data, the proposed data has little in-image description, providing complete data for in-depth investigation of vehicle detection.In the proposed vehicle tracking and counting system, first the road pavement in the image is removed and a new proposed road segmentation is divided into far and near areas; This method is important for the development of car detection. Then place the above two regions in the YOLOv3 network to determine the type and location of the vehicle. Finally, the vehicle trajectory is obtained with the ORB algorithm, which can be used to determine the driving direction of the vehicle and to obtain different vehicles. Several main video analysis methods are used under different conditions to analyze the proposed process. Experimental results prove that the use of the proposed segmentation method can provide more detection, especially for the detection of small objects.Also, the new concept explained in this article is very effective in travel and traffic decision. This form is very important for the control of detail scenes..

Keywords: Vehicle dataset, Image segmentation, Vehicle detection, Vehicle counting, Highway management

Introduction

Traffic detection and statistics in high-speed traffic analysis video scenes are essential for intelligent highway management. With the widespread installation of traffic surveillance cameras, large files of video images are obtained for analysis. Most of the time, the distance to the road can be determined from an elevation perspective. Vehicle objects in this view are very different in size and small objects far from the road have poor accuracy. In the face of complex camera scenes, it is important to solve the above problems well and use them more.

In this article, we propose solutions to the above problems and apply vehicle search results to various tracking and vehicle counts.

Related work on vehicle detection

Currently, vision-based tool object detection is divided into machine vision models and deep learning. Traditional machine vision techniques use the motion of a vehicle to separate it from a static background image. Such methods can be divided into three groups [1]: methods using background subtraction [2], methods using contrast images [3], and techniques using visual inspection [4]. Using the video frame differences, the difference is calculated from the pixel values of two or three consecutive video frames. In addition, the mobile front area is separated from the head [3].

Parking can also be detected using this method and reducing noise [5].

While editing the background image in the video, the background model is created using the background information [5]. After that, each image frame is compared to the background model, which can classify moving objects. The technique of using optical equipment can detect moving parts in the video. The resulting optical field represents the orientation and pixel speed of each pixel [4]. Vehicle detection methods using tool features such as scale-invariant feature transform (SIFT) and surface-invariant feature transform (SURF) methods are widely used.

For example, 3D modeling has been used for vehicle detection and task classification [6]. Using the curved lines [7] of the 3D ridges of the exterior of the car, cars are divided into three classes: sedans, SUVs, and vans.

The use of deep convolutional networks (CNNs) has been quite successful in vehicle detection. CNN is capable of learning image features and can perform many functions such as classification and bounding box regression [8]. The inspection method can be broadly divided into two groups. The two-step approach generates input objects from various algorithms and then distributes the objects through a neural network. The one-step method does not create a bounding box and directly converts the object-bounding box location problem into a regression problem for processing.

In the two-step approach, Region-CNN (R-CNN) [9] uses selective region detection [10] in images. The input image to the convolutional network must be relatively large, and the deep structure of the network requires a long training time and consumes a lot of memory. Borrowing the concept of spatial pyramid matching, SPP NET [11] allows the network to input images of various sizes and have a fixed output. R-FCN, FPN, and Mask RCNN improve the video extraction process, feature selection, and resource allocation of different communication channels. Among the single-stage methods, the most important is the Single Smuggling Multiple Box Detector (SSD) [12] and You Look Only One (YOLO) [13] basis. SSD uses MutiBox [14], Region Bid Network (RPN), and multiple representations; they use the default set of junction boxes with different aspect ratios to localize objects. YOLO [13] network splits the image into fixed rows, unlike SSD. Each grid is responsible for predicting objects whose positions are in the grid. YOLOv2 [15] adds a BN (Bulk Normalization) layer that allows the network to normalize the input of each layer and make the network converge.

YOLOv2 uses a multivariate training algorithm that randomly selects a new image size every ten packets. Our vehicle detection device uses the YOLOv3 [16] network.

The creation of YOLOv2, YOLOv3 uses logistic regression of product groups. The category loss method is a binary cross entropy loss which can handle multiple tags for the same item. And use logistic regression to regress the confidence box to determine whether the IOU of the previous box and the actual box is greater than 0.5. If there is more than one priority box that meets the conditions, only the highest priority box in the IOU is retrieved.

In the final prediction, YOLOv3 uses three variables to predict objects in the image.

Machine vision methods are always faster at detecting vehicles but do not work well in situations such as changes in bright images, moving backgrounds, slow traffic, or harsh conditions. Advanced CNNs achieved great results in object detection; however, CNNs are sensitive to changes in their detection properties [17, 18]. The one-stage method uses a grid to estimate the product, and the limitation of the grid prevents the two-stage method from achieving greater accuracy, especially for small products. The two-step method uses a region of interest integration to divide the candidate region into blocks based on constraints, and if the candidate region is smaller than the size of the given parameter, the candidate area is populated according to the size of the given parameter. In this way, the characteristic pattern of small objects is destroyed and the detection accuracy is low.Current systems do not distinguish between major or minor items in the same category. Using the same method for the same type of product will also lead to erroneous detection. Using pyramids or multi-view images can solve the above problems, but is considered too expensive.

Vehicle detection research in Europe

Image-based vehicle tracking in Europe has had great results. In [19], on the "Hofold-ing" and "Weyern" sections of the A8 motorway in Munich, Germany, the multivariate displacement (MAD) method [20] was used to analyze short-term changes in two figures. time delay Vehicles in motion are shown in the displacement diagram, which is used to estimate the vehicle's speed on the road. In [21], the A95 and A96 motorways near Munich, the A4 motorway near Dresden, and the "Mittlere Ring" in Munich were taken as test areas, and the Canny edge algorithm [22] was applied to the road image and the calculation results were obtained. . perpendicularity histogram. Then, using the k-means algorithm, the edge perpendicularity statistics are divided into three parts, and the closed car model is determined according to the height. The comparison method is used to create a color model to identify and remove shadows from vehicles [23], thereby removing noise from motion in space. Vehicle detection can be greatly improved after shadows are removed.

The experiments in [23] were carried out on Italian and French highways. In [24] HOG and Haar-like features were compared and the two concepts were combined to create a vehicle detection system that was tested on French vehicle images. However, when the above method is used for vehicle control, the vehicle type cannot be determined. Also, when the light is not enough, it is difficult to edge the car or detect the car's movement, causing the problem of the car not being clear and disturbing the further use of search results. [19, 20] used aerial images but failed to capture the features of all vehicles and created fake vehicles.

However, with the development of deep learning, CNN-based traffic detection has been successful in Europe. In [25], Fast R-CNN was used for vehicle detection in a traffic environment in Karlsruhe, Germany. Fast R-CNN uses a selective search strategy to find all matching boxes, which is time-consuming and slow to search for traffic.

In summary, research on vision-based vehicle detection is still advancing, and important problems useful for design car repair in Europe are gradually being overcome.

Related work on vehicle tracking

Advanced vehicle tracking applications such as multi-object tracking are also important for ITS projects [26]. Most material tracking methods typically use search-based tracking (DBT) and free detection (DFT) for initialization. The DBT method uses a background model to detect moving objects in the video before playback. The DFT method is required to initiate object tracking, but cannot handle joining new objects and leaving old objects. The Multi-Object Tracking algorithm should take into account the similarity between objects in the center and the coordination between adjacent objects.

The similarity between objects in a frame can be exploited using the correlation coefficient (NCC). It is used to calculate the distance between colored histograms of objects, such as the Bhattacharyya distance [27]. When joining the link, it should be taken into account that an object can appear in only one match, and the path can correspond to only one object. Now, classification at the analysis level or classification level can solve this problem. To solve the problems caused by the change of moving objects and changes in light [28], he used SIFT features to track objects even at high speeds.

The ORB feature point detection algorithm [29] is proposed for this task. ORB can get better content at a faster rate than SIFT.

In summary, it can be concluded that the method suitable for the purpose of the car is transferred from the search for traditional methods to the search for deep convolutional network method experience. Also, there are several public records for specific situations. The sensitivity of convolutional neural networks to scale change causes erroneous detection of small objects. It is difficult to monitor multiple objects and traffic using highway security cameras. As a result, our contributions include:

1. A large database of large vehicles is created, which can provide many different elements of the vehicle collected in various situations captured by cameras on the highway. This information can be used to evaluate the performance of various vehicle detection algorithms when dealing with changes in traffic.
2. A method of detecting small objects, which improves the accuracy of vehicle detection on the highway.

Extract the highway area, divide it into far and near zones, and place it in a convolutional network for traffic detection.

3. Multiple monitoring and analysis methods for highways. The ORB algorithm is used to extract and match the content of the search object, determine the search path, and calculate traffic direction and traffic flow. This research will be explained in the following sections. The Vehicle Information section describes the vehicle information used in this document. The "System Architecture" section presents the entire flow of the planning process. The "Method" section explains our strategy in detail. Experiments and related analyzes are presented in the "Results and discussion" section.

The "Select" section shows all the articles.

Vehicle dataset

Road surveillance cameras are installed all over the world, but traffic footage is rarely made public due to legal, privacy, and safety concerns. In terms of image acquisition, image data can be divided into three groups: images captured by the onboard camera, images captured by the surveillance camera, and images captured by non-surveillance cameras [30]. The KITTI benchmark dataset [31] contains images of highway landscapes and public roads that can solve problems such as 3D object detection and tracking. The Tsinghua-Tencent bus dataset [32] contains 100,000 images from the dash camera covering various lighting and weather conditions, although no vehicles are registered. The Stanford car dataset [33] is a database of cars with car lights captured by non-surveillance cameras.The database contains 19,618 vehicles, including make, model and year of manufacture. The Synthetic Cars dataset [34] is similar to the Stanford Cars

dataset but includes more images. 27,618 photos include the fastest car, number of doors, number of seats, modifications, and models. 136,727 images containing all the views of the car. Information is captured by security cameras; an example is the BIT-Car dataset with 9,850 images [35]. This data divides vehicles into 6 types: SUV, sedan, minivan, car, bus, and minibus; but the camera angle is good and the vehicle objects are too small for any image, which is difficult for CNN training. generalize

Traffic and accident data (TRANCOS) [36] includes images of vehicles on highways captured by security cameras, with a total of 1,244 images. Most images have some closure. The dataset contains thumbnails and doesn't give vehicle models, so it's not good to use. Therefore, several documents contain important explanations and a few photos of traffic situations.

This section introduces the vehicle dataset from the perspective of the highway surveillance videos we created. The docs has been published at: http://drive.google.com/open?id=1li858elZvUgss8rC_yDsb5bDfiRyhdrX.
Images were taken from highway surveillance videos in Hangzhou, China (Figure 1). The highway surveillance camera is installed on the side of the road and installed at 12 meters; It has an adjustable view and has no preset position. Images taken from this perspective span the length of the highway and feature vehicles that vary widely. The images were taken by 23 security cameras for different situations, different times, and different lighting conditions.
The data divides vehicles into three groups: cars, buses, and trucks (Figure 2). The registration information is stored in a text file containing numerical values for product groups and corresponding values for product containers. As shown in Table 1, the data contains a total of 11,129 images and 57,290 text boxes.
The image has RGB format and 1920*1080 resolution. Note that we collect small objects in adjacent road areas, so the data includes vehicle items of different sizes. Annotations closer to the camera have more features, while annotations far from the camera have fewer features. The description of various variables is useful for improving the accuracy detection of small car parts. The data is divided into two parts: the training set and the test set. In our data, the share of automobiles is 42.17%, the share of buses is 7.74% and the share of trucks is 50.09%. There are 5 of them.
There is an average of 15 samples per picture. Figure 3 compares the number of annotations in our dataset with the PASCAL VOC, ImageNet, and COCO datasets. Our dataset is a general-purpose, multi-purpose dataset that can be used in many regions such as Europe. Compared to current traffic data,

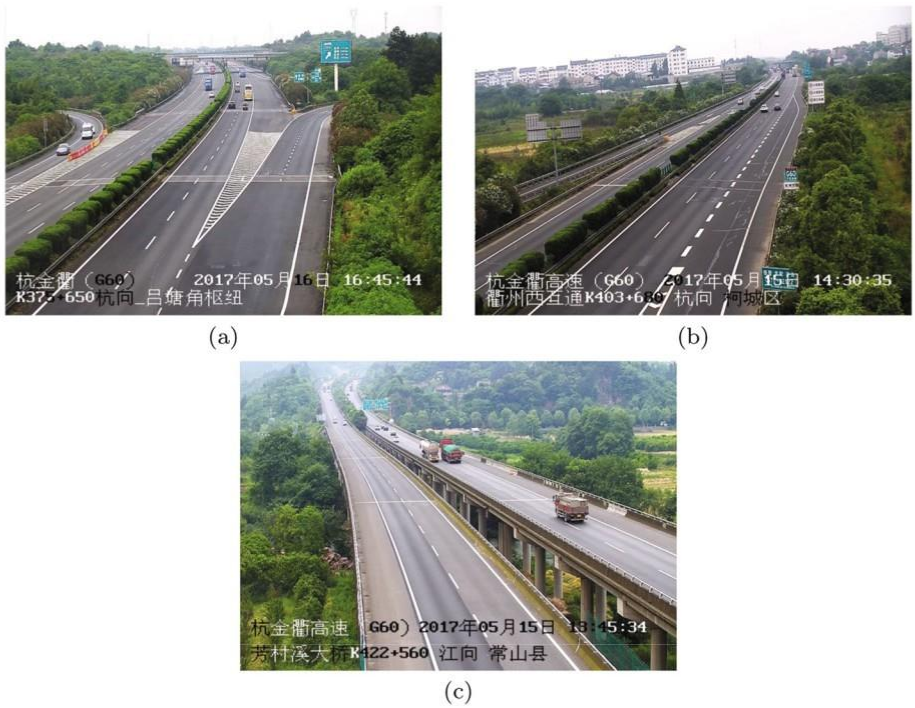

Fig. 1 Scenes taken by multiple highway surveillance cameras. **a** Scene 1; **b** Scene 2; **c** Scene 3

Fig. 2 Vehicle labelling category of the dataset

our dataset has higher resolution images, adequate lighting and complete instructions.

The system structure

This chapter introduces the main structure of the vehicle inspection and calculation system. First, the video data of a traffic zone is entered. After that, extract and distribute the sections of the path. YOLOv3's depth sensing method is used to detect vehicle objects in traffic. Finally, ORB feature extraction is performed in the vehicle detection framework to achieve multi-target monitoring and traffic information acquisition.

According to Figure 4, the road boundary of the highway was deduced by the road segmentation method. The path area is divided into two parts, far field and near field, according to the location of the camera mount. Then, vehicles in both directions are detected using the YOLOv3 object detection algorithm. This algorithm can improve the detection of small objects and solve the problem that objects are difficult to detect due to changes in measured objects.Then the ORB algorithm is used to track multiple objects. The ORB algorithm extracts the features of the detection box and performs matching to obtain the correlation between the same object and different images. Finally, traffic analysis is calculated. Create the object-to-target trajectory, determine the vehicle's driving direction, and collect traffic information such as the number of vehicles of all types. The system improves the accuracy of target detection from the perspective of highway surveillance video and generates a full view search and tracking and traffic information across the entire camera.

Methods

Road surface segmentation

This chapter presents methods for improving pavement extraction and segmentation. We perform subtraction and segmentation using image processing techniques such as Gaussian blend models, which provide better vehicle detection when deep search methods are used. There are many ways to watch the video. As the main focus in this research is the vehicle, the area of interest in the picture is the highway pavement. At the same time, the area of the road is concentrated in a certain part of the picture according to the shooting angle of the camera.

Using this video we can extract the large pavement area in the video. The surface removal process is shown in Figure 5.

To eliminate the effect of traffic in the road segmentation area, as shown in Figure 5, we use a Gaussian blend model to subtract the background from the first 500 frames of the video movie. The values of the pixels in the image are Gaussian distributed over some critical points in time, and statistics are generated for each pixel in each frame of the image. If a pixel is off-center, the pixel belongs to the foreground. If the pixel point's value deviates by a different variance from the position value, the pixel point is considered to belong to the background.

The combination of Gaussian models is particularly useful in images with a lot of background pixels, such as the highway surveillance images used in this study.

After subtraction, the background image is smoothed with a 3*3 core Gaussian filter. The MeanShift algorithm is used to soften the color of the input image,

Table 1 Vehicle dataset information published in this study

Image format	Size	Total number of images	Total number of instances	$\frac{Totalnumberofinstances}{Totalnumberofimages}$
RGB	1920 x 1080	11,129	57,290	5.15

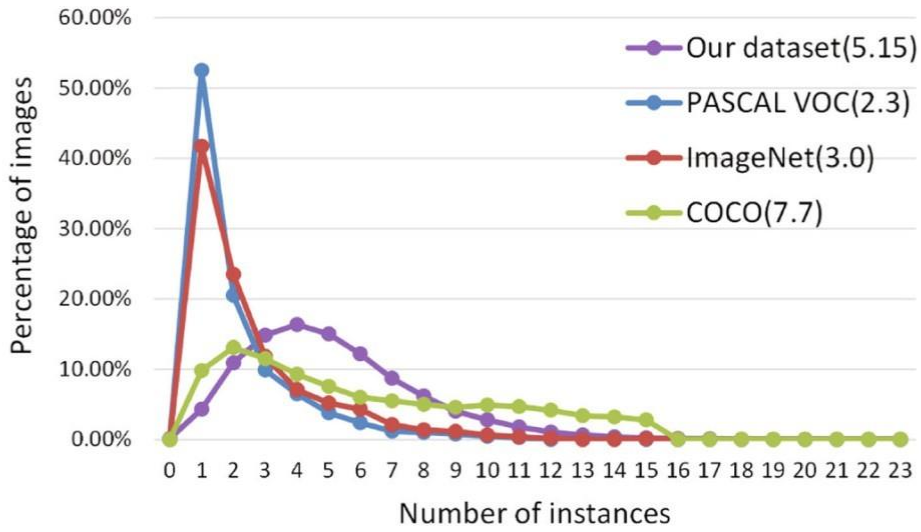

Fig. 3 Annotated instances per image (average numbers of annotated instances are shown in parentheses)

average the color of the same color, and make color with small area. On this basis, flood fill algorithm is used to separate the pavement area. The flood fill algorithm selects a point in the road region as the seed point and makes the continuous area of the road adjacent to the pixel value of the seed point. The pixel value of the continuous field is close to the pixel value of the core content.

Finally, hole filling and morphological widening is done to further remove the surface. We subtract the pavement of different highways (Figure 6) and the results are shown in Figure 7.

We separate sections of the road to provide the right ideas for controlling the next car. For the subtracted overlay image, a minimal contoured rectangle is created for the unrotated image. Divide the finished image into five equal parts, the 1/5 area near the starting point of the intersection is defined as the near-far area of the road, and the remaining 4/5 area is defined as the near-near. road surface area. The proximity of the area and the proximity of the area 100 pixels overlapping area (as shown in the red image in Figure 8) to solve the problem, the car in the image will be divided into two parts by the above operation.

Check the pixel values of the near field and near end line by line. If the pixel value in a column is all zero, the entire image in the column is black, not unidirectional; removed. After removing the uncoated areas, the remaining areas are referred to as the far and near coverage areas.

Vehicle detection using YOLOv3

This section presents the research methodology used in this study. The main way to find cars is to use the YOLOv3 network. The YOLOv3 algorithm continues the main idea of the previous two generations of YOLO algorithms. Convolutional Neural Networks are used to extract features

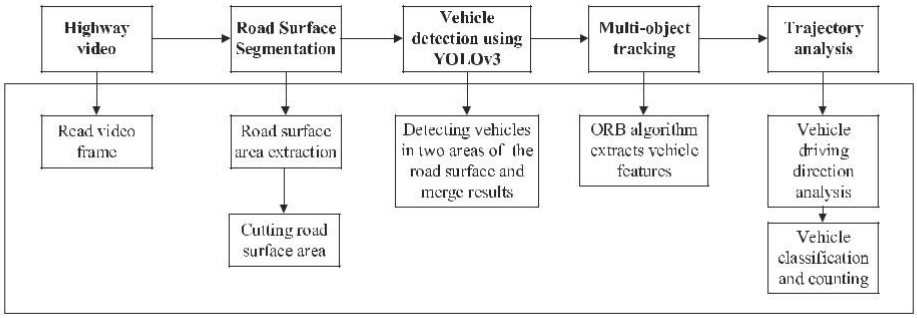

Fig. 4 Overall flow of the method

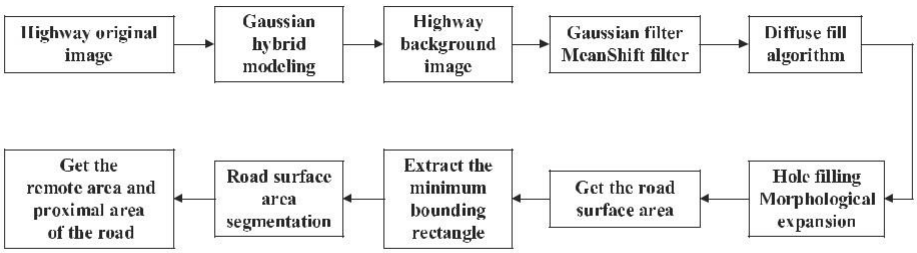

Fig. 5 Overall flow of road surface extraction

from input images. According to the size of its price - it is called Darknet-53.

The model uses the full frame map such as 13*13 to split the input image into convolution method and replaces the 13*13 grid of the previous version. The center of the text box is directly connected to the grid unit in the convolutional neural network, and the grid unit is now responsible for predicting the structure. The branch is used directly on the crop. The network model adopted by YOLOv3 is based on a deep understanding of the network.

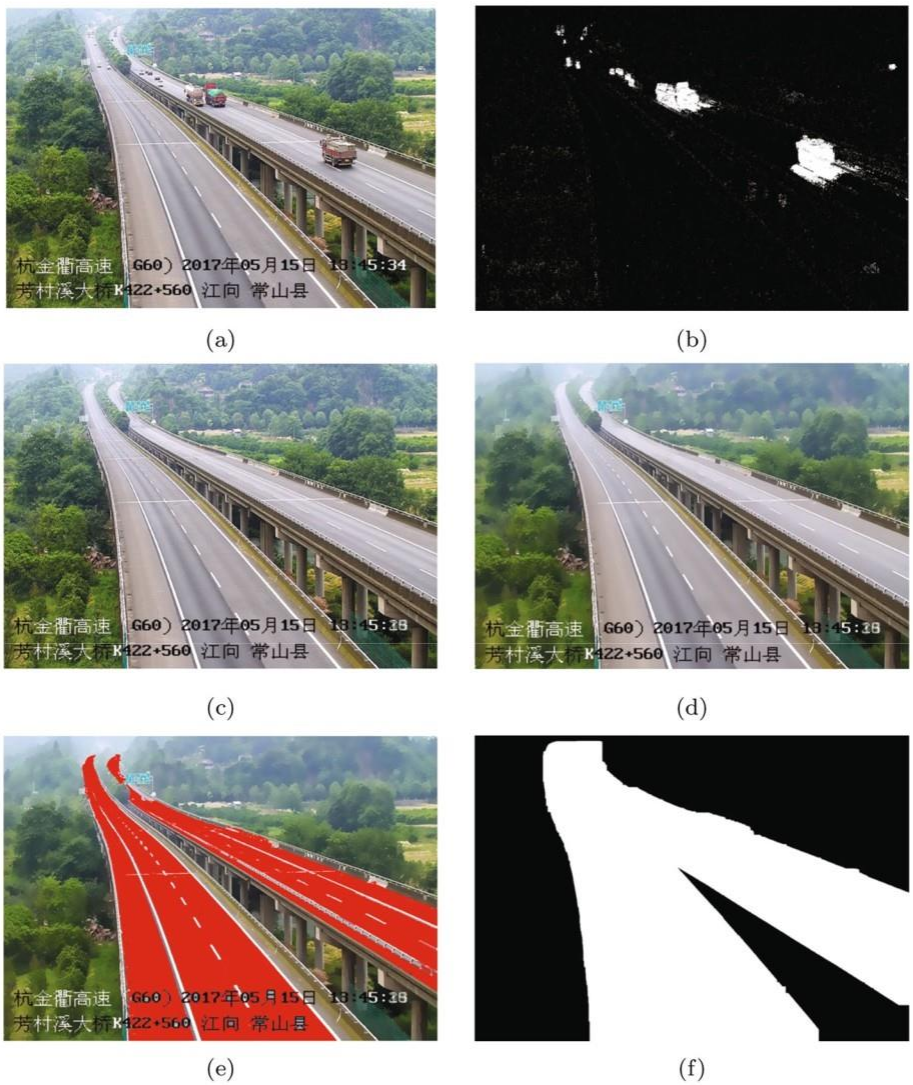

Fig. 6 Process of road surface area extraction. **a** Original image; **b** image foreground; **c** image background; **d** Gaussian filter and MeanShift filter; **e** diffuse filling; **f** road surface area mask

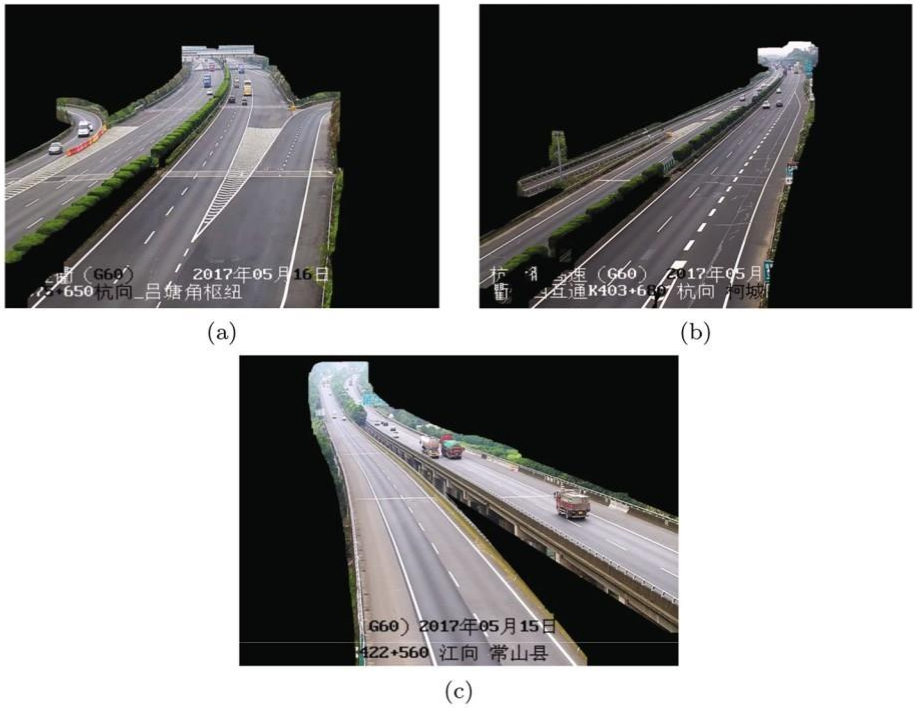

Fig. 7 Road surface extraction results for different highway scenarios. **a** Scene 1; **b** Scene 2; **c** Scene 3

Learning the rest directly ensures the integrity of the image quality data, simplifies the training process and improves the overall detection accuracy of the network. In YOLOv3, each cell will have three different boxes at different scales per object. The candidate box that most overlaps with the text box will be the final guess. Also, the YOLOv3 network has three sets of outputs and three branches are bundled together. Shallow features are used to capture small objects and deep features are used to capture large objects; so the network can detect objects with different values.

The search speed is fast and the search accuracy is high. Traffic events captured by highway video surveillance are highly adaptable to the YOLOv3 network. The network finally releases the shared product, trust, and sequencing.

When using YOLO detection, the image is converted to the same size as 416 * 416 when sent to the network. Because the image is segmented, the distant road surface will be distorted and large in size. Therefore, many points of a small car's target can be achieved, and some of the target features can be avoided because the target car is too small. Put the data presented in "Vehicle dataset" for training into the YOLOv3 network to achieve target vehicle detection. The vehicle detection model can detect three types of vehicles: cars, buses, and trucks (Figure 9).

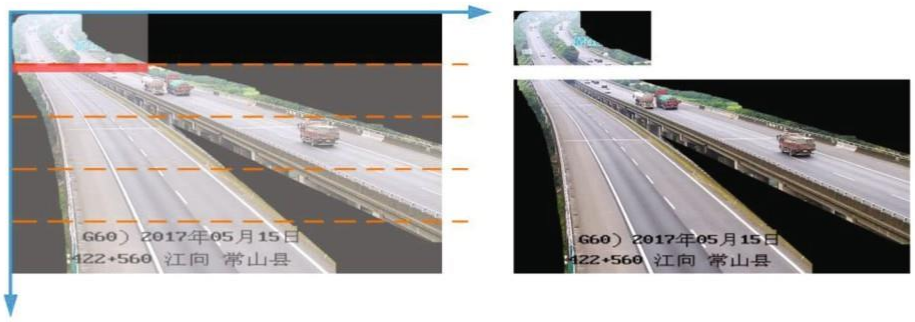

Fig. 8 Road surface segmentation

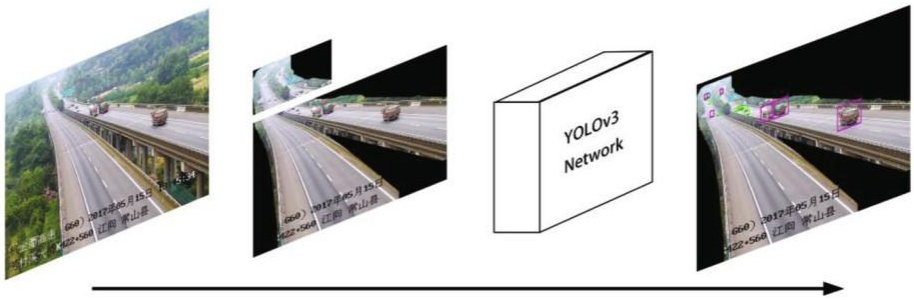

Fig. 9 Segmented image sent to the detection network and detected results merging. (Green, blue, and fuchsia boxes are labelled to indicate the “car”, “bus”, and “truck” regions, respectively.)

They are not in our search as there are very few motorcycles on the highway. Areas far and near the road surface are sent to the network for details. Return the position of the two areas of the car to the original image and get the object in the original image. Find out the category and location of the vehicle using the vehicle's object finder to obtain the necessary information to track the object. The above information is sufficient for the car calculation, so the car search method does not show the car's characteristics or the car's condition.

Multi-object tracking

This section explains how to track multiple items based on the package detected in the "Cars Using YOLOv3" section. In this study, the ORB algorithm was used to extract the characteristics of the detected vehicles and good results were obtained. The ORB algorithm performs best in terms of computational efficiency and cost consistency. This algorithm is a good alternative to SIFT and SURF image identification algorithms. The ORB algorithm uses Features from Fast Segment Analysis (FAST) to identify specific points and then uses the Harris operator for corner detection.

Use the BRIEF algorithm to calculate the descriptor after obtaining certain points. A coordinate system is created where the prime point is the center and the center of the area is along the x-axis of the coordinate system.So, when the image is rotated, the coordinate system rotates with the rotation of the

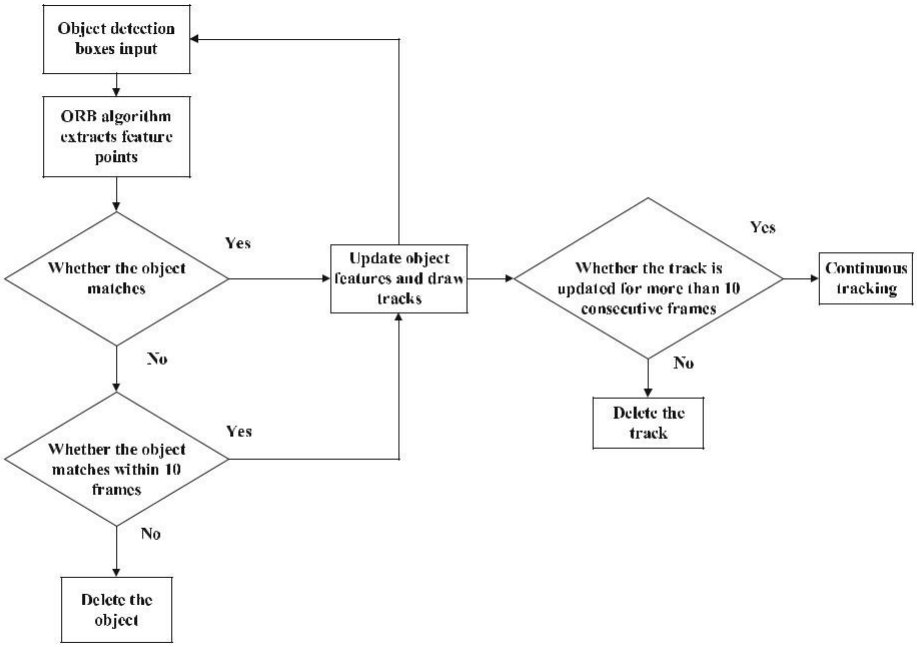

Fig. 10 Process of multi-object tracking

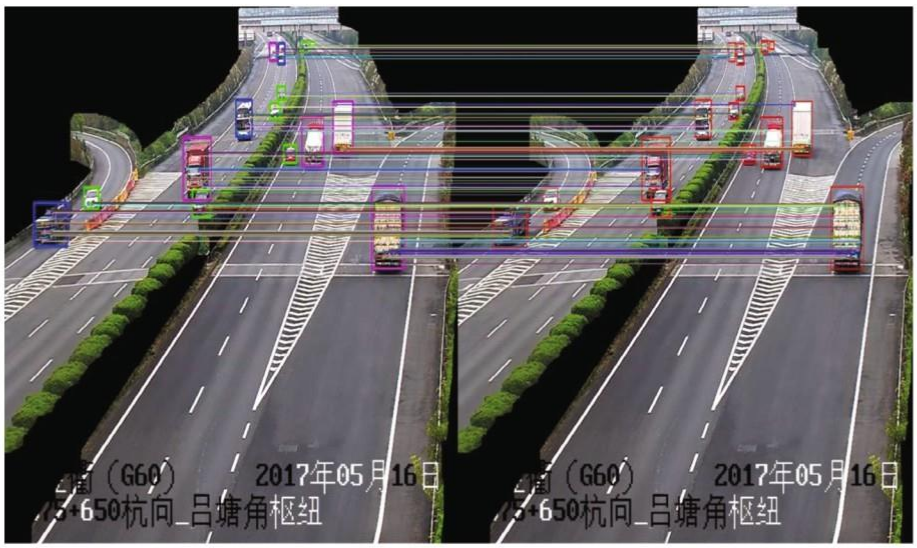

Fig. 11 Features of the detection object extracted by the ORB algorithm

image and the feature point identifier is thus rotated. A similar point can be made when changing perspective. After the binary feature point identifier is obtained, the feature points are matched with the XOR operation, which improves the matching function.

The follow-up process is shown in Figure 10. When the number of matching points is greater than the threshold, the point is determined to be similar and the match of the product is drawn. The basis of the estimation is as follows: use the RANSAC algorithm to clean up certain points that can remove the false noise matching the error and predict the homography matrix. Based on the predicted homography matrix to find the frame and the position of the original object, a perspective transformation is performed to meet the predicted frame.

We use the ORB algorithm to extract the unique content in the objects detected by the vehicle detection algorithm. There is no object feature extraction from the entire path area, greatly reducing

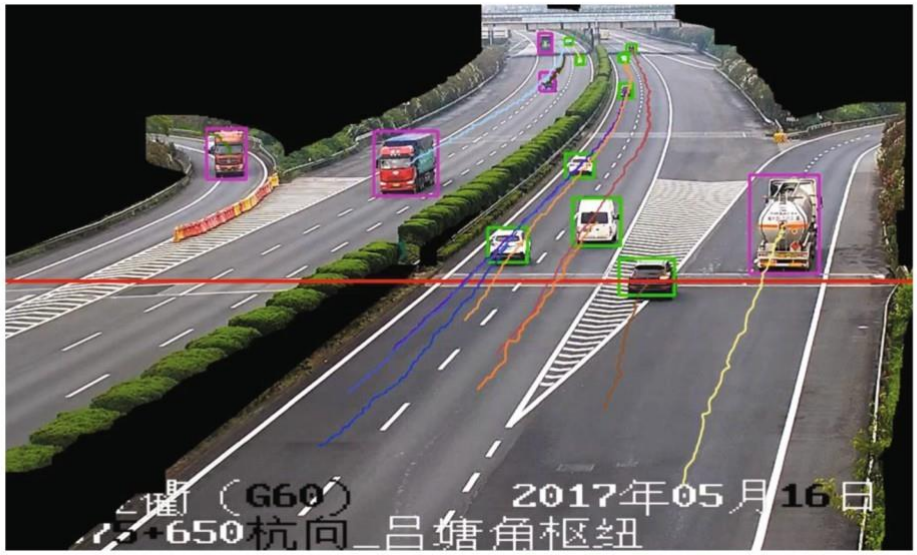

Fig. 12 Trajectory of the vehicle and detection line

the computational cost. In target tracking, since the cascade of the car's target on the video is subtle, the prediction of the target in the frame bit is drawn first, as the ORB attributes are extracted from the target. If frame estimation and finding the frame of the next frame meet the shortest distance requirement from the center, the same object is successfully obtained in both frames (Figure 11). We mean the threshold T, which expresses the maximum in pixels of the center of the vehicle detection box moving between two adjacent images. The travel position of the same car in two adjacent places is less than the T threshold. Therefore, when the main point of the car frame moves more than T in the two frames next to each other, the car in the two frames is the same and the information association does not work. Given that the parameter changes when the car moves, the value of the threshold value T is related to the size of the target box of the car. Different parts of the car have different parts.

This transparency can meet the needs of on-the-go vehicles and different video input sizes. T is calculated from Equation. 1. Box height is the height of the car box.

$$T = \frac{box\ height}{0.25} \quad (1)$$

In accordance with the scene of the camera acquiring a wide-angle view on the highway, we remove trajectories that are not updated ten times in a row. In this type of situation, the surface of the camera is too far away. In ten consecutive videos, the car will move further. Therefore, the trajectory is deleted when the trajectory is not updated for ten frames. At the same time, the

Table 2 Number of objects under different detection methods

vehicle trajectory and the detection line intersect only once, and the initial setting does not affect the final calculation.

If the prediction frame is not consistent across consecutive frames, the absence of objects in the video scene is determined and the prediction image is removed. Through the above process, the target of the earth can be detected and viewed from a full highway surveillance video perspective.

Trajectory analysis

This section provides an analysis of the moving parts and statistical data of various vehicle components. Most highways run in both directions with barricades separating the roads. According to the direction of the vehicle tracking path, we distinguish between the camera- registered direction of the vehicle in the global coordinate system (direction A) and the distance of the camera (direction B). Place a straight line on the traffic diagram as a search line for traffic classification and statistics. The detection line is located at a height above 1/2 of the vehicle image (Figure 12).Count two traffic flows at the same time. When the motion of the object affects the detection line, the data of the object is recorded. Finally, the number of products in different directions and groups can be obtained in one go.

Results and discussion

In this section, we describe the performance measures of the methods presented in the Policy section. We performed a test on the vehicle data set made in the "Vehicle Data Set" section. Our experiment uses HD video highways in three different locations, as shown in Figure 1.

Scenes	Video frames	Vehicle category	Total number of vehicle objects					
			Our method		Full-image detection method		Actual number of vehicles	
			Remote area	Proximal area	Remote area	Proximal area	Remote area	Proximal area
Scene 1	3,000	Car	6,128	8,430	493	6,616	6,849	8,550
		Bus	535	459	92	379	582	483
		Truck	5,311	5,320	840	4,703	5,792	5,471
Scene 2	3,000	Car	1,843	3,615	192	3,356	1,914	3,654
		Bus	194	364	82	295	207	382
		Truck	3,947	4,709	922	3,738	4,169	4,731
Scene 3	3,000	Car	1,774	2,336	224	2,188	1,834	2,352
		Bus	415	516	56	495	483	529
		Truck	3,678	3,490	731	2,662	3,726	3,507

Network training and vehicle detection

We use the YOLOv3 network for vehicle detection and the data we generate for network training. In network training, there is no optimal solution for dataset partitioning. Our dataset partitioning method follows the usual method. We divided the data into 80% training and 20% testing. Our dataset contains 11,129 images and the training set and test images were selected from the dataset.

Due to the large number of images in the data set, the speed of the test set and the training set is sufficient to obtain the model. The cost of the training process must be high to get the right model. The training set consists of 8,904 images and many vehicle models can be trained to get accurate models to detect cars, buses and trucks. The test set contains 2225 images of the target car, which are completely different from the training set, sufficient to test the accuracy of the trainer. We use a batch size of 32, set the distortion weight to 0.0005, momentum value to 0.9, and maximum training iteration count to 50,200. We use a learning rate of 0.01 for the first 20,000 iterations and then change that to 0.

001 after 20,000 iterations. This method reduces gradient descent and loss rate. We changed the default link box to match the text box using k-language++. The training process of our dataset calculated the default junction box size of 832 * 832 network resolution and yielded nine sets of values: [13.2597, 21.

4638], [24.1990, 40.4070], [39.4995, 63.8636], [61.

4175, 96.3153], [86.6880, 137.2218], [99.3636, 189.

3695, 342.4765], with an average IOU of 71.20%. In order to improve the detection results of small objects, we did not provide samples smaller than 1 pixel during training, we put them in the network for training. What we've published is a combination of Darknet-53's previous layer of routing before the last yolo layer, and a custom map of Darknet-53's 11th layer.

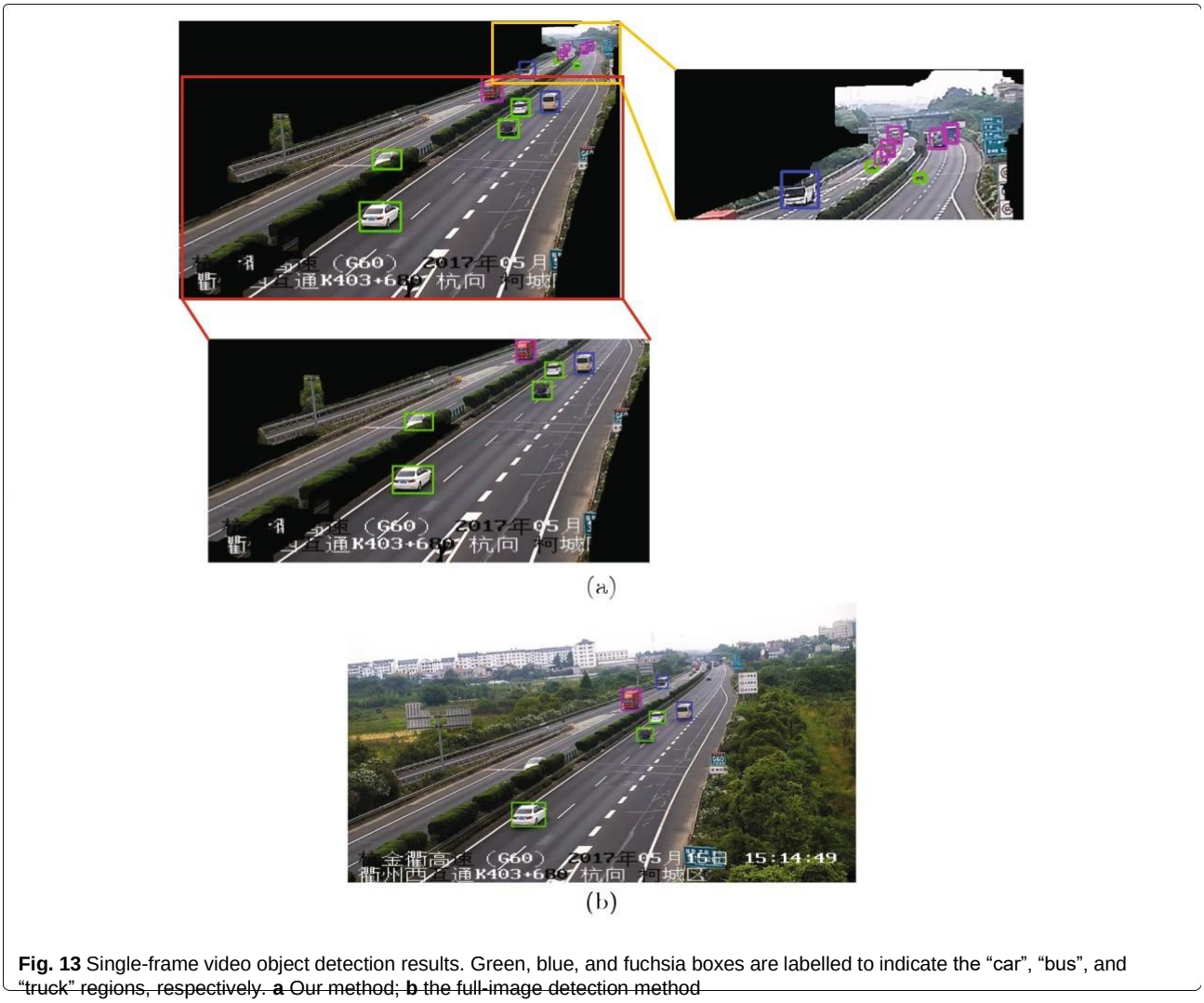

Fig. 13 Single-frame video object detection results. Green, blue, and fuchsia boxes are labelled to indicate the “car”, “bus”, and “truck” regions, respectively. **a** Our method; **b** the full-image detection method

Table 3 Comparison of actual vehicle numbers by using different methods

Vehicle category	Remote area		Proximal area		Average correct rate	
	Our method	Full-image detection method	Our method	Full-image detection method	Our method	Full-image detection method
Car	91.96%	8.58%	98.79%	83.54%	95.375%	46.06%
Actual number of vehicles Bus	89.94%	18.08%	96.05%	83.86%	92.995%	50.97%
Truck	94.51%	18.21%	98.61%	80.99%	96.56%	49.6%
Overall correct rate	92.14%	14.96%	97.82%	82.80%	94.976%	48.86%

In the upsampling layer before the last yolo layer, we set the step to 4. When we set the image to be displayed on the network, the network resolution becomes 832*832 instead of the default 416*416 resolution. After increasing the resolution, the mesh can have a larger resolution when released from the yolo layer, thus improving the accuracy of target detection.

Using our training model, 3000 consecutive frames of images are used for vehicle detection in many large scenarios. We extract and classify road areas and feed them into the network for vehicle detection. We will compare our method with testing images with a resolution of 1920 * 1080 on the network (without distribution of sites); the results are shown in Table 2 and Figure 13. We compare the detected objects with real vehicles in various ways as shown in Table 3.

Compared with the actual number of vehicles, our method is closer to the actual number of vehicles when the object is in an area close to the highway. When the product is placed in the far part of the road, the difference in detection is still less than 10%. The full-screen image detection system does not detect small objects in the far part of the road. Our system improves the detection of small objects in remote areas of the road. Meanwhile, our method is the most effective method for detecting images in the region close to the road.

But injustice is not real. CNNs can detect false objects or detect non-objects, resulting in false traffic. Therefore, we calculate the average accuracy for the data in Table 4. We use the testing method to evaluate the center of truth (map) of the model, based on 80% of the training process and 20% of the testing process; the graph represents the true mean (ap) mean.For each class, ap defines an average of 11 points for each class performance precision/recall curve. [0, 0.1, 0.2, ..., one]. For returns greater than any threshold (the threshold in the experiment is 0.25), there will be an equal maximum pmax(return). The 11 precisions above are calculated and ap is the average of 11pmax (return).

We use this value to describe the quality of our models.

$$ap = \frac{1}{11} \sum_{recall=0}^{1} p_{max}(recall), \quad recall \in [0, 0.1, ..., 1],$$

$$map = \frac{ap}{class\ number} \tag{2}$$

The calculation of *precision* and *recall* is as follows:

$$Precision = \frac{TP}{TP+FP},$$

$$Recall = \frac{TP}{TP+FN} \tag{3}$$

where TP, FN, and FP are positive, negative, and negative numbers, respectively. We got the final report value of 87.88% showing that the road is a good way to localize and deploy different tools. From the above analysis, we can see that the overall accuracy of our target detection is 83.46%, which indicates that the location and distribution of the car's target is different, which is great and gives better results for finding multiple targets.

Tracking and counting

After receiving the target, we follow the vehicle according to the ORB feature point matching method and perform trajectory analysis. In the experiment, when each object has more than ten matches, the corresponding ORB prediction function is generated. We use the search line to

Table 4 Accuracy of the network model

Parameters	ap			Precision	Recall	Average IoU	mAP
	Car	Bus	Truck				
Results	86.46%	88.57%	88.61%	0.88	0.89	71.32%	87.88%

detect the movement of the vehicle according to the direction created by tracking the trajectory and to make statistics. We do the tests in our other videos as part of "network training and car detection" with the same situation but with different numbers. We measure the speed of the system proposed in this paper using the real-time value, which is defined as the ratio of the time the system takes to process the video to the time it takes to film the original.

in balance. 4. Uptime is the time it takes the system to play the video, and movie time is the time it takes to play the original video. The smaller the value of real time, the faster the system will perform the calculations. When the value of the real-time value is less than or equal to 1, the video input can be processed in real time.

$$real\ time\ rate = \frac{system\ running\ time}{video\ running\ time} \quad (4)$$

The results are shown in Table 5. The results show that the average accuracy of vehicle routing and vehicle counting is 92.3% and 93.2%, respectively. In video surveillance on the highway, small objects such as small cars are easily blocked by large vehicles.

At the same time, many cars will appear from one side, which will affect the accuracy of the calculation. Our raw video runs at 30 frames per second. It can be seen from the calculation of the speed that the vehicle tracking algorithm based on the ORB features is quite fast. The operating speed of the system is related to the traffic in the situation. The more cars there are, the more features need to be extracted and the longer the process takes.

In general, the car counting system proposed in this paper is very close to real-time processing.

Conclusions

In this study, a high speed car target was determined for security cameras and a suitable target detection and tracking system for highways was proposed. By removing the large walking area, the ROI area becomes better. The YOLOv3 target detection algorithm obtains an end-to-end traffic detection model based on the specified traffic target data. For the problem of detecting small objects and many changes of objects, the field of the path is defined as the far and near field. The two path areas of each frame are detected as a link and good vehicle detection is obtained in the inspection area.

Depending on the target detection results, the position of the target in the image is estimated by the ORB feature extraction algorithm. Vehicle trajectories can then be obtained by monitoring the ORB properties of various objects. Finally, the driver, type of vehicle, number of vehicles etc. The train is analyzed to collect current traffic information such as Experimental results prove that the proposed method for vehicle detection and tracking in highway surveillance video scenes is effective and efficient. Compared to traditional vehicle equipment monitoring methods, the method proposed in this document is low cost, highly secure and does not require major construction or installation work on existing maintenance equipment.

Based on the research reported in this document, security cameras can be further measured to obtain internal and external measurements of the camera. Thus, the position information of the vehicle's trajectory is transferred from the image coordinate system.

Table 5 Track counting results

Scenes		Scene 1			Scene 2			Scene 3			Direction correct rate
Video frames		11000			22500			41000			
	Vehicle category	Car	Bus	Truck	Car	Bus	Truck	Car	Bus	Truck	
	Our method	29	21	3	110	40	21	287	141	22	
	Actual number of vehicles	32	21	3	117	43	22	297	150	24	
Direction A	Extra Number	3	0	0	8	3	2	15	13	3	0.92
	Missing number	0	0	0	1	0	1	5	4	1	
	Correct rate	0.906	1	1	0.923	0.930	0.864	0.933	0.887	0.833	
	Our method	41	37	4	117	69	13	300	168	15	
	Actual number of vehicles	43	38	4	125	77	13	311	172	17	
Direction B	Extra Number	2	2	0	11	10	0	15	8	2	0.931
	Missing number	0	1	0	3	2	0	4	4	0	
	Correct rate	0.953	0.947	1	0.888	0.844	1	0.939	0.930	0.882	
	Real time rate		1.27			1.35			1.48		
	Average correct rate		0.967	0.911		0.917					0.932

world management system The speed of the car can be calculated according to the speed of the camera. With the vehicle detection and tracking system, bad parking and traffic conditions can be detected and more traffic information can be obtained. In summary, cars in Europe, such as Germany, France, the UK, and the Netherlands, have the same characteristics as those in our car dataset, and the angles and heights of security cameras installed in these countries can be seen clearly for a long time. - distance. Therefore, the methods and results of vehicle inspection and calculation methods provided by this analysis will be an important reference for traffic investigation in Europe.

Confirmation

Not valid.

Authors' contributions

H-SS: Research and data analysis, article writing. H-XL: Content preparation and analysis, drafts. H-YL: Content Planning, Labels. ZD: Data analysis, text editing. XY: Data Collection, Labels.

The final article was read and approved by all authors.

Funding

This work was supported by the National Natural Science Foundation of China (No.61572083), the joint project of the Ministry of Education (No.614102022610), the special fund (No.300102248402) for the economic research copy foundation of the central school education group.) and the research and development key of Shaanxi Province (No.2018ZDXM- GY-047).

Availability of data and materials

The vehicle data created in Chapter 2 can be found in Google Drive, http:// drive.google.com/open?id=1li858elZvUgss8rC_yDsb5bDfiRyhdrX. Additional information was analyzed while the current study was not made public due to confidentiality.

The file contains personal information that may not be publicly available. Ensure that data produced for a research project is used only in the context of that research.

Competing interests

The author declares on behalf of all authors that

has no competing choice of text.

References

1. Al-Smadi, M., Abdulrahim, K., Salam, R.A. (2016). Traffic surveillance: A review of vision based vehicle detection, recognition and tracking. *International Journal of Applied Engineering Research*, *11*(1), 713–726.
2. Radhakrishnan, M. (2013). Video object extraction by using background subtraction techniques for sports applications. *Digital Image Processing*, *5*(9), 91–97.
3. Qiu-Lin, L.I., & Jia-Feng, H.E. (2011). Vehicles detection based on three-frame-difference method and cross-entropy threshold method. *Computer Engineering*, *37*(4), 172–174.
4. Liu, Y., Yao, L., Shi, Q., Ding, J. (2014). Optical flow based urban road vehicle tracking, In *2013 Ninth International Conference on Computational Intelligence and Security*. https://doi.org/10.1109/cis.2013.89: *IEEE.*
5. Park, K., Lee, D., Park, Y. (2007). Video-based detection of street-parking violation, In *International Conference on Image Processing*. https://www.tib.eu/en/search/id/BLCP%3ACN066390870/Video-based-detection-ofstreet-parking-violation, vol. 1 (pp. 152–156). Las Vegas: IEEE.
6. Ferryman, J.M., Worrall, A.D., Sullivan, G.D., Baker, K.D. (1995). A generic deformable model for vehicle recognition, In *Procedings of the British Machine Vision Conference 1995*. https://doi.org/10.5244/c.9.13: *British* Machine Vision Association.
7. Han, D., Leotta, M.J., Cooper, D.B., Mundy, J.L. (2006). Vehicle class recognition from video-based on 3d curve probes, In *2005 IEEE International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance*. https://doi.org/10.1109/vspets.2005.1570927: IEEE.
8. Zhao, Z.Q., Zheng, P., Xu, S.T., Wu, X. (2018). Object detection with deep learning: A review. arXiv e-prints, arXiv:1807.05511.
9. Girshick, R., Donahue, J., Darrell, T., Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation, In *2014 IEEE Conference on Computer Vision and Pattern Recognition*. https://doi.org/10. 1109/cvpr.2014.81: IEEE.
10. Uijlings, J.R.R., van de Sande, K.E.A., Gevers, T., Smeulders, A.W.M. (2013). Selective search for object recognition. *International Journal of Computer Vision*, *104*(2), 154–171.
11. Kaiming, H., Xiangyu, Z., Shaoqing, R., Jian, S. (2014). Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, *37*(9), 1904–16.
12. Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C.Y., Berg, A.C. (2016). Ssd: Single shot multibox detector, In *2016 European conference on computer vision*. https://doi.org/10.1007/978-3-319-46448-0_2 (pp. 21–37): Springer International Publishing.
13. Redmon, J., Divvala, S., Girshick, R., Farhadi, A. (2016). You only look once: Unified, real-time object detection, In *2016 IEEE conference on computer vision and pattern recognition*. https://doi.org/10.1109/cvpr.2016.91 (pp. 779–788): IEEE.
14. Erhan, D., Szegedy, C., Toshev, A., Anguelov, D. (2014). Scalable object detection using deep neural networks, In *2014 IEEE conference on computer vision and pattern recognition*. https://doi.org/10.1109/cvpr. 2014.276 (pp. 2147–2154): IEEE.
15. Redmon, J., & Farhadi, A. (2017). *Yolo9000: Better, faster, stronger*: IEEE. https://doi.org/10.1109/cvpr.2017.690.
16. Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.
17. Cai, Z., Fan, Q., Feris, R.S., Vasconcelos, N. (2016). A unified multi-scale deep convolutional neural network for fast object detection, In *2016 European conference on computer vision*. https://doi.org/10.1007/978-3-319-46493-0_22 (pp. 354–370): Springer International Publishing.
18. Hu, X., Xu, X., Xiao, Y., Hao, C., He, S., Jing, Q., Heng, P.A. (2018). Sinet: A scale-insensitive convolutional neural network for fast vehicle detection. *IEEE Transactions on Intelligent Transportation Systems*, *PP*(99), 1–10.
19. Palubinskas, G., Kurz, F., Reinartz, P. (2010). Model based traffic congestion detection in optical remote sensing imagery. *European Transport Research Review*, *2*(2), 85–92.
20. Nielsen, A.A. (2007). The regularized iteratively reweighted mad method for change detection in multi-and hyperspectral data. *IEEE Transactions on Image processing*, *16*(2), 463–478.
21. Rosenbaum, D., Kurz, F., Thomas, U., Suri, S., Reinartz, P. (2009). Towards automatic near real-time traffic monitoring with an airborne wide angle camera system. *European Transport Research Review*, *1*(1), 11–21.
22. Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, *PAMI-8*(6), 679–698.
23. Asaidi, H., Aarab, A., Bellouki, M. (2014). Shadow elimination and vehicles classification approaches in traffic video surveillance context. *Journal of Visual Languages & Computing*, *25*(4), 333–345.
24. Negri, P., Clady, X., Hanif, S.M., Prevost, L. (2008). A cascade of boosted generative and discriminative classifiers for vehicle detection. *EURASIP Journal on Advances in Signal Processing*, *2008*, 136.
25. Fan, Q., Brown, L., Smith, J. (2016). A closer look at faster r-cnn for vehicle detection, In *2016 IEEE intelligent vehicles symposium (IV)*. https://doi.org/ 10.1109/ivs.2016.7535375: IEEE.
26. Luo, W., Xing, J., Milan, A., Zhang, X., Liu, W., Zhao, X., Kim, T.K. (2014). Multiple object tracking: A literature review. arXiv preprint arXiv:1409.7618.
27. Xing, J., Ai, H., Lao, S. (2009). Multi-object tracking through occlusions by local tracklets filtering and global tracklets association with detection responses, In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. https://doi.org/10.1109/cvprw.2009.5206745