
Article

Real-Time Traffic Flow Statistics Based on Dual-Granularity Classification

Yanchao Bi, Yuyan Yin, Xinfeng Liu*, Xiushan Nie, Chenxi Zou, and Junbiao Du

School of Computer Science and Technology, Shandong Jianzhu University, Jinan 250101, China

* Correspondence: liuxinfeng18@sdjzu.edu.cn

Received: 17 June 2023

Accepted: 13 September 2023

Published: 26 September 2023

Abstract: Traffic detection devices can cause accuracy degradation over time. Considering problems such as time-consuming and laborious manual statistics, high misdetection probabilities, and model tracking failures, there is an urgent need to develop a deep learning model (which can stably achieve detection accuracy over 90%) to evaluate whether the device accuracy still satisfies the requirements or not. In this study, based on dual-granularity classification, a real-time traffic flow statistics method is proposed to address the above problems. The method is divided into two stages. The first stage uses YOLOv5 to acquire all the motorized and non-motorized vehicles appearing in the scene. The second stage uses EfficientNet to acquire the motorized vehicles obtained in the previous stage and classify such vehicles into six categories. Through this dual-granularity classification, the considered problem is simplified and the probability of false detection is reduced significantly. To correlate the front and back frames of the video, vehicle tracking is implemented using DeepSORT, and vehicle re-identification is implemented in conjunction with the ResNet50 model to improve the tracking accuracy. The experimental results show that the method used in this study solves the problems of misdetection and tracking effectively. Moreover, the proposed method achieves 98.7% real-time statistical accuracy by combining the two-lane counting method.

Keywords: dual-granularity classification; vehicle re-identification; traffic; multi-objective tracking algorithm; dual-lane counting

1. Introduction

With the rapid development of the economy, the traffic flow on arterial routes increases sharply. Real-time and accurate traffic flow statistics can assist the traffic police and corresponding departments in making reasonable decisions. Specifically, such statistics can provide road traffic information for the planning and construction of national highways, as well as the operation and management of road networks and traffic emergencies.

According to a national highway traffic situation survey, many regions currently use a large amount of traffic survey equipment to monitor the traffic on arterial routes. Owing to various reasons, the detection accuracy of the equipment decreases as time goes by, failing to provide accurate detection of arterial traffic situations. The original inspection method is to 1) perform manual frame-by-frame statistical counts; 2) compare the count results and equipment results; and 3) determine whether the current equipment detection accuracy satisfies the standard or not. If the detection accuracy is below the standard, the equipment needs to be repaired. This method is inefficient as it not only wastes considerable manpower and material resources but also fails to ensure detection accuracy. Therefore, there is a need to develop arterial traffic detection systems with a detection accuracy of over 90%. It is effective to consider a standard to evaluate whether the accuracy of the purchased equipment satisfies the requirements or not.

In recent years, the amount of traffic investigation equipment has increased largely. The workload has also increased for determining whether the detection equipment satisfies standards or not. Relying only on the original manual inspection methods consumes more time and wastes more human and material resources. With the continuous development of deep-learning algorithms, many deep-learning algorithms have been applied to arterial traffic inspection, which gradually becomes an important research area. Combining deep learning with traffic survey equipment detection as well as tracking systems can significantly improve the accuracy of arterial traffic flow detection. Thus, these combination methods can be used as measures to test the usage status of traffic survey equipment, and

assist traffic management departments in allocating traffic resources, improving road traffic efficiency and managing urban traffic congestions.

In this study, we evaluate the previous research on vehicle detection and tracking based on deep learning. A real-time traffic flow statistics method is proposed based on dual-granularity classification to reduce the probability of false detection and tracking recognition and realize real-time high-precision statistics of the traffic flow of different vehicle models. This evaluation reveals that the use of only one classification method to classify vehicle models results in the identification of non-motorized vehicles as motorized vehicles, which is an open scenario detection problem that is difficult to solve. Motivated by this point, we try to develop a dual-granularity classification method to obtain all motor vehicles and non-motor vehicles, and classify models only for motor vehicles, thereby lowering the probability of false detection.

We find that YOLOv5 + DeepSORT is a mature video tracking classification and counting model framework that obtain good results which is used as the basic framework for improvement. Google proposed the EfficientNet model in 2019 by evaluating the preferable combination of balancing the network depth, width, and image size. The YOLOv5 model is used to realize coarse-grained classification in the first stage. Furthermore, the EfficientNet model is used to realize fine-grained classification in the second stage, which constitutes the dual-grained classification model framework of this method. It is necessary to identify targets repeated in the front and back frames in DeepSORT, and assign an identical number to these targets to indicate if they are the same target or not. The above operation is called target re-recognition. Nevertheless, the original re-recognition in DeepSORT only can be applied to pedestrians. Therefore, we select the ResNet50 model with higher accuracy to replace the original pedestrian re-recognition model realize vehicle re-recognition, which improves the accuracy of vehicle re-recognition and increases the probability of successful tracking. Existing counting methods are used to determine the vehicle category when a vehicle arrives in the middle area. Research and analysis reveal that vehicles in the middle position would have problems of identity alteration, counting failures, and repeated counting because of two-lane vehicle occlusions. This severely affects the statistical accuracy. Therefore, we propose a dual-lane counting method that prevents counting in the middle area to significantly reduce the occurrence probability of these problems.

Several tests reveal that the method can significantly reduce the probability of false detection and tracking failures, and achieve real-time statistical accuracy of 98.7%. The contributions of this study are summarized below.

1) A dual-granularity classification method is proposed to obtain motorized and non-motorized vehicles for coarse-grained classification, where only motorized vehicles are classified and counted at a fine-grained level by the model. This simplifies the problem and reduces the probability of incorrectly inspecting a non-motorized vehicle as a motorized vehicle.

2) The YOLOv5 and EfficientNet lightweight models are used to achieve real-time detection. DeepSORT is combined with the ResNet50 model to realize vehicle re-identification., which reduces the number of identity transformation and decreases the probability of tracking failures.

3) The dual-lane counting method is used to realize the split vehicle-counting function, which further reduces the problems of counting failures and repeated counting caused by vehicle blocking in the middle.

2. Related Works

Deep learning-based vehicle detection and tracking have the advantages of high speed and accuracy in both complex environmental backgrounds and edge devices with low arithmetic power. Vehicle detection, tracking, and classification algorithms based on deep learning in the image field perform better than the conventional manual statistical approach. Such a statistic is used to determine whether the device's accuracy satisfies the requirements or not. Many classical algorithms have been proposed to achieve better accuracy, yielding new solutions to the problem of accurate inspection of traffic detection equipment.

2.1. Deep Learning Network Models

2.1.1. Object Detection Model

In the first stage of the proposed method, a classifier with high speed and high precision is required to realize the separation of motor vehicles and non-motor vehicles from the scene. YOLO + DeepSORT, as a mature vehicle detection framework, achieves good detection performance, and is used to achieve the coarse-grained classification task at the first stage.

The YOLO series is one of the most widely used single-stage models in the industry. It displays high detection speeds and accuracy, and this satisfies the industrial requirements of real-time detection. Although the advanced versions are YOLOv7 and YOLOv8, the training cost of the two versions is higher than YOLOv5. Therefore, the YOLOv5 model (shown in [Figure 1](#)) is selected to undertake the above task. YOLOv5 has significant advantages in

terms of flexibility, speed, and model deployment speed. Its minimum weight file is 14 M, which can be detected rapidly at 140 FPS on the TeslaP100 graphics card. Therefore, we select YOLOv5 as the detector of DeepSORT to realize the real-time detection of vehicles, and the detection results will be directly used for subsequent tracking and classification.

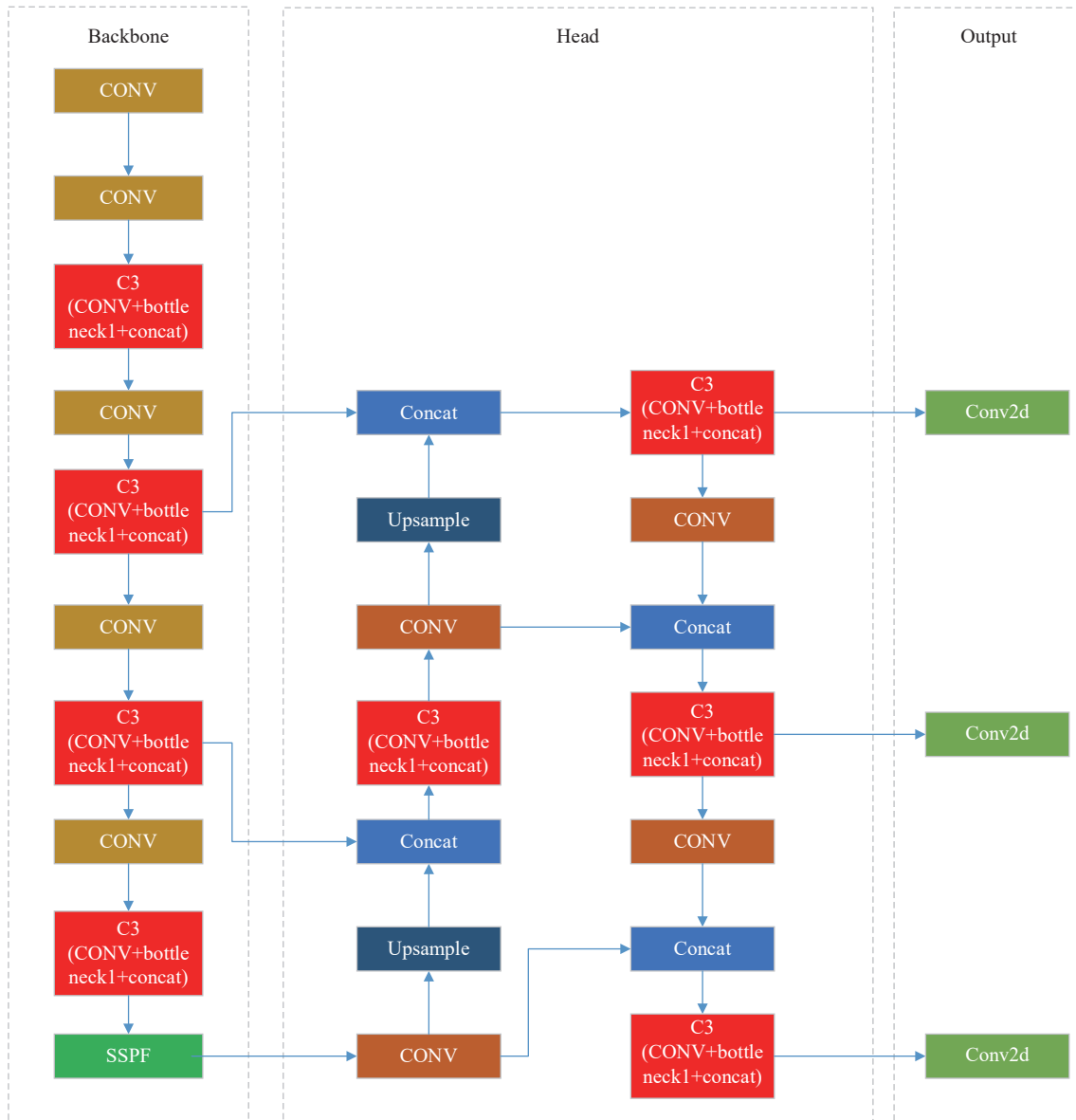


Figure 1. YOLOv5 structure diagram.

The fine-grained classification of the second stage can still be implemented using YOLOv5. Nevertheless, it is found that the EfficientNet model proposed by Google has higher accuracy than the YOLO series for the same number of parameters. Note that in the second stage of fine-grained classification, the large number of classification categories, complex tasks and parameters will lead to low detection accuracy. Therefore, in order to achieve the real-time detection, we use the EfficientNet model to achieve fine-grained classification tasks.

It is well known that increasing the width, depth, and image size of the network can improve the generalization capability of the model. Nevertheless, if only one of the parameters is increased, the accuracy of the model would reach a saturation state rapidly. Considering that the depth, width, and image size of the network interact with each other, Google evaluated the preferable combination and proposed the EfficientNet model [1] in 2019, see Figure 2. The performance of the EfficientNet network is shown in Figure 3.

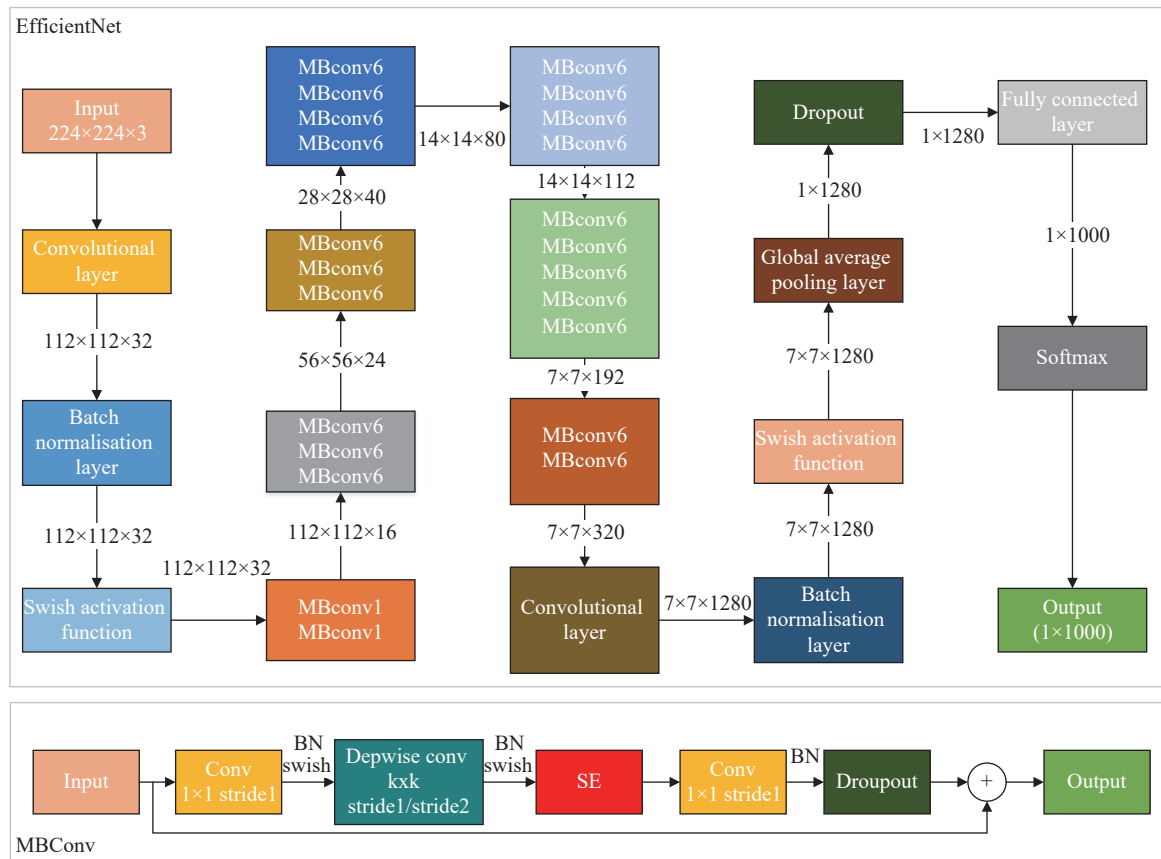


Figure 2. EfficientNet-b1 structure diagram.

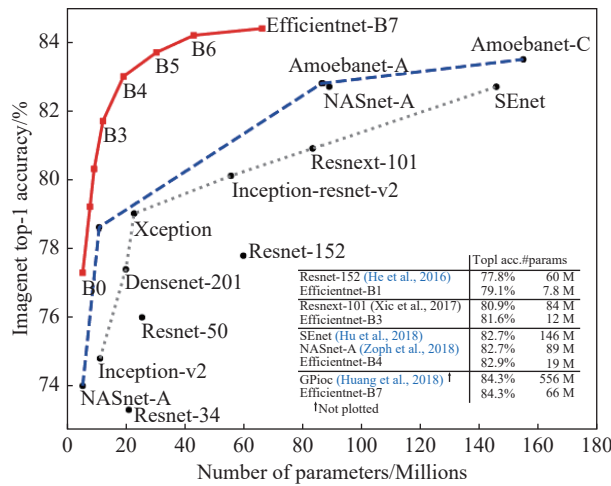


Figure 3. Comparison of EfficientNet network with other networks.

The red curve in Figure 3 indicates the performance of the efficient network model, the horizontal axis indicates the model size, and the vertical axis indicates the model accuracy. It can be observed that the method proposed by Google extends the model by adjusting the width and depth of the network. Moreover, the resolution of the input image can substantially improve the accuracy of the model under the premise of ensuring computational performance. Therefore, we select this model as the classification model for fine-grained classification to realize real-time detection and high-precision classification.

2.1.2. Object Tracking Model

Since the ultimate goal of this method is to count vehicle flow, it is extremely important to identify the same vehicle in different frames of the video. In order to prevent the same vehicle from repeated counting and improve the counting accuracy, a good tracking method should accurately identify whether the vehicles are the same vehicle or not, thereby reducing the problem number of the identity change during counting. DeepSORT [2] is a popular and

effective target tracking method that reduces ID switching by integrating awareness information. Figure 4 shows the flowchart. As shown in the figure, a cascade matching strategy is added to the DeepSORT algorithm to assign one tracker to each detector. To calculate the motion and appearance information of the target, marsupial and cosine distances are added, and the confirmation state is introduced. When a target is occluded, if it is in the verification state, the system considers that it has a high probability of reappearing until the maximum number of mismatches is exceeded. This improves the tracking accuracy. Target spacing and feature similarity are considered simultaneously. Moreover, a validation mechanism is adopted for the newly generated tracking trajectories to exclude erroneous prediction results.

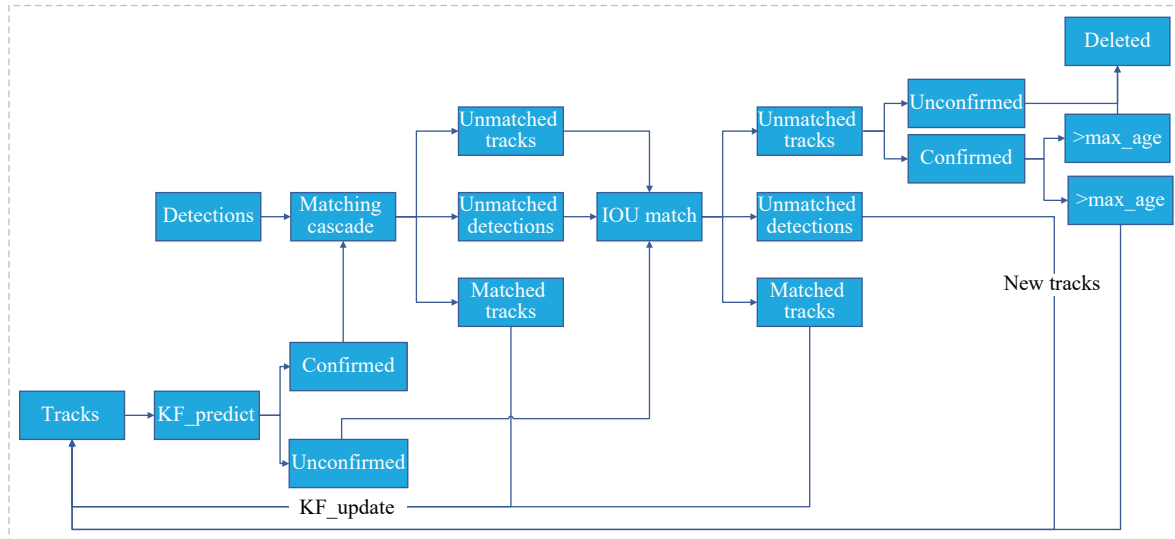


Figure 4. Schematic diagram of DeepSORT algorithm sequence.

The re-recognition algorithm in DeepSORT can be used only for pedestrian re-recognition. Therefore, the Resnet50 model was used in [3] (rather than the original re-recognition algorithm) to realize vehicle re-recognition. The ResNet50 model structure is shown in Figure 5. This model effectively solves the degradation problem of deep networks caused by gradient explosions via introducing shortcut connections. Therefore, this method extracts the apparent information of the vehicle to realize vehicle re-recognition and reduce the frequency of the identity transformation during tracking.

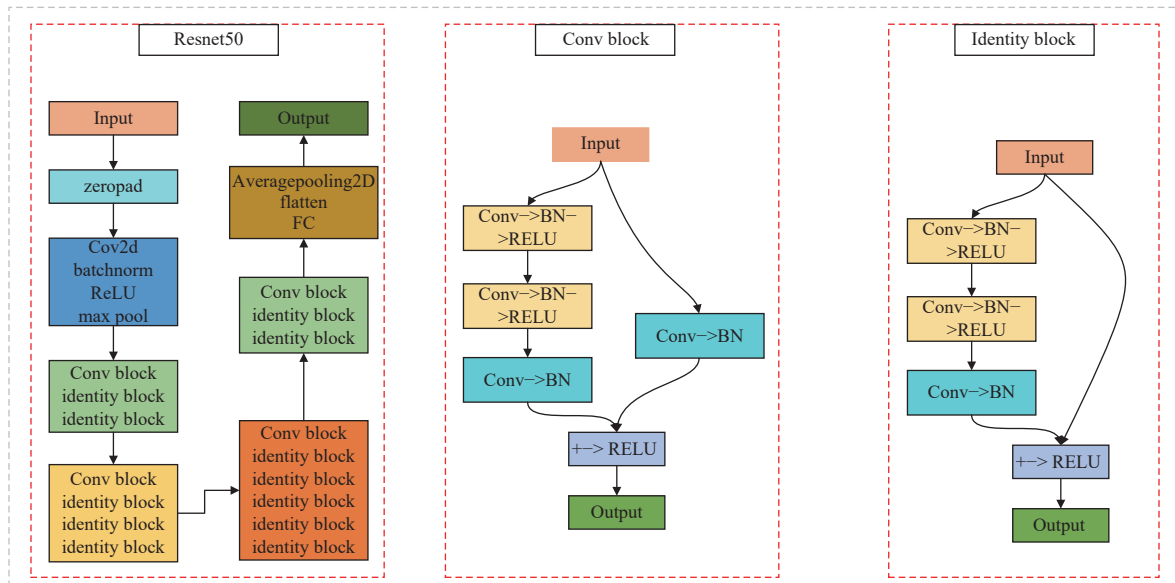


Figure 5. Structural diagram of ResNet50, conv block module and identity block module.

2.2. Deep Learning Based Vehicle Counting Approach

Currently, vehicle detection has become a hot research direction, and there exists a rich body of research on vehicle detection based on deep learning.

Combined with DeepSORT tracking, Li et al. [4] proposed a method to improve YOLOv5 for traffic flow statistics in surveillance videos with low accuracy. This improves localization accuracy and speeds up the rate of bounding box regression. Replacing NMS with DIOU-NMS reduces the problem of missing detection after target occlusion. Nevertheless, the problem of vehicle occlusion is not considered, and the brightness of light in the detection process will also affect the accuracy of recognition.

Jin et al. [5] proposed a frontal multi-target vehicle tracking algorithm with optimized DeepSORT in order to improve the recognition of the surrounding environment when the car is driven automatically. Combining the central loss function and cross-entropy loss function, the target has better functions such as inter-class aggregation and inter-class recognition. Nevertheless, since the light in the scene changes, especially when it changes suddenly, the robustness of the target detection will be reduced, resulting in the ID-switch phenomenon in the tracking process.

To eliminate shortcomings of low detection accuracy and poor robustness of traditional target detection and tracking algorithms, Jia et al. [6] proposed a real-time traffic flow detection method at intersections to realize end-to-end real-time traffic flow detection for intersection monitoring based on improved YOLOv5 and DeepSORT algorithmic models. The experimental results show that the method detects faster and better for vehicles in complex environments with an average accuracy of 96.6%.

Kumar et al. [7] proposed a robust vehicle detection technique based on improved RVD-YOLOv5 to improve the accuracy of vehicle detection. In the first stage of the method, data clustering is performed on the dataset using K-mean algorithms to generate classes of objects. In the second stage, YOLOv5 is applied to create bounding boxes, and the non-maximum suppression technique is used to eliminate the overlapping of vehicle bounding boxes. Finally, in the third stage, the loss function CIOU is used to obtain the exact regression bounding box of vehicles. The experimental results show that the proposed method achieves better results compared to the state-of-the-art methods.

In recent years, various types of target tracking algorithms have been increasingly used in the field of computer vision. Besides, to improve the accuracy of target tracking algorithms, more and more research has been conducted. Zhang et al. [8] reconstructed the feature extraction network of YOLOv5 (using the improved EfficientNetv2) to reduce the model complexity and improve the real-time detection speed of the model. This solves the problem that the existing multi-target tracking algorithm has many parameters and large computation, but makes it difficult to meet the real-time requirements of mobile devices.

Gai et al. [9] addressed the problems of low tracking accuracy and large tracking errors in pedestrian target tracking. The high-performance YOLOv5 algorithm is introduced into the DeepSORT algorithm to detect the tracking video frame by frame, predict the target position using Kalman filtering, and match the same target using the Hungarian algorithm. The MOTA of the algorithm improves by 9.9%, the number of false and missed detection significantly reduces, and the number of target ID switching reduces by 23 times.

Zhou et al. [10] used the YOLOv5 detector to detect vehicles in the range of interest in consecutive frames and adopted the DeepSORT tracker to achieve stable and high-speed vehicle tracking. This successfully reduces the sudden change of target ID and improves the stability of tracking. Nevertheless, it also affects the detection and tracking accuracy due to environmental and climatic conditions.

For the task of analyzing pedestrian-vehicle interactions at non-signalized intersections, Zhang et al. [11] proposed a processing framework for analyzing pedestrian-vehicle interactions based on YOLOv3-DeepSort. The study of exit interactions can provide a theoretical basis for proposing protective measures to ensure the safety of pedestrians and vehicles at non-signalized intersections. A variety of comparative experiments show that the proposed framework has excellent performance in trajectory extraction.

In order to improve the tracking failure caused by small target pedestrians and partially obstructed pedestrians in dense crowds in complex environments, Chen et al. [12] proposed a deep learning-based intelligent vehicular pedestrian target detection and tracking method. Based on YOLO, a channel attention module and a spatial attention module are introduced to achieve the weight amplification of important feature information in the channel and spatial dimensions, while improving the model's ability to represent important feature information. The experimental results show that the improved model can improve the detection accuracy of small-target pedestrians, deal with the target occlusion problem, and avoid the tracking failure due to occlusions.

In the vehicle detection, there are small vehicles with small targets and frequent occlusions which are difficult to detect. To solve this problem, a number of scholars have proposed a large number of improvement methods. In order to cope with the small targets captured by UAVs, Zhan et al. [13] applied four improvement methods to the original YOLOv5s model including redesigning the anchor size, adding an attention module to the CSPDarknet53 backbone network, replacing the GIoU loss function with the CIOU loss function, and adding the P2 functional level. To a certain extent, this method prompts the development of the object detection algorithm for UAV platforms. Nevertheless, the improvement of this algorithm only works for small objects, and it is not suitable for vehicle detection with large object volume differentiation.

Combining detection and feature matching techniques, Ye et al. [14] proposed a multi-target tracking algorithm based on the phenomenon that the target association accuracy of the multi-target tracking algorithm is reduced under frequent occlusions. All the proposed methods can effectively reduce the occlusion of the target, thus achieving stable tracking. Nevertheless, this method divides detection and tracking into two stages of training, which affects the real-time performance.

Based on background-related and identity-related confidence scores, Wang et al. [15] designed pyramid background-aware attention to extract background-related features from the global image and construct dual confidence metrics to obtain robust clustering results in the clustering process. Further to mitigate the background interference, a background filtering module for local branching is proposed in local branching. The experimental results validate that the performance of the proposed method for balancing background information is superior to the most advanced method.

Aiming at solving problems of insulator defects and low detection accuracy of small targets in the current complex background, Wang et al. [16] proposed a neural network detection algorithm (named the EfficientNet-YOLOv5) based on the deep learning framework. In this algorithm, the images of various insulators are photographed in the transmission line by UAVs, and the YOLOv5s backbone network is replaced by the EfficientNet network. As a result, the leakage detection problem caused by insulator target overlapping is solved. The experimental results show that the average accuracy of the improved network reaches 98.5%.

Zhang et al. [17] proposed a vehicle detection method based on YOLOv5. The method is highly adaptable and performs well under various conditions such as heavy traffic, night environment, multi-vehicle overlap, and partial loss of vehicles. Experimental results show that the algorithm can utilize materials such as pictures, videos and real-time surveillance, and has a high recognition rate in real-time.

As vehicle inspection pursues real-time detection, it is necessary to significantly increase the detection speed while reducing the model size. There are some related studies in this regard. To reduce the disadvantages of traditional video surveillance algorithms such as low accuracy, poor robustness and inability (to achieve real-time multi-target tracking), Xu et al. [18] proposed a multi-target tracking algorithm DeepSORT to achieve real-time detection and tracking of multiple targets in end-to-end surveillance videos. The high accuracy of YOLO algorithm for object detection makes DeepSORT less dependent on detection, less interfered with occlusions and illumination, and more robust in tracking.

Li et al. [19] proposed a new network structure of MobileNetv2-YOLOv5s in order to reduce the size of the model and improve the detection speed. Comparing lightweight networks MobileNetv2 and YOLOv5s with other object detection algorithms, the improved YOLOv5s has better detection results.

For the problem of vehicle counting, choosing an appropriate method of counting can reduce the probability of duplicate vehicle counts and omissions, thereby improving the statistical accuracy. Currently, there are also a lot of excellent counting methods.

Wen et al. [20] proposed a mobile target tracking and multi-lane vehicle counting model DCN-Mobile-YOLO. The deformable convolutional networks and mobile convolutional network MobileNetv3 are used instead of the regular convolutional kernel and backbone network of YOLO v4. Such networks are combined with the DeepSORT algorithm to achieve tracking and counting of multi-targets. Then, adaptive lane detection rules are established and accurate counting of vehicles is achieved in lanes. The results show that the DCN-Mobile-YOLO model not only improves the detection accuracy of the target detection model, but also achieves the speed of real-time detection of mobiles.

Vehicle flow detection has an important position in the field of intelligent transportation. Dai et al. [21] proposed the background difference method and the hybrid Gaussian model to improve the detection quality of moving vehicles. Specifically, virtual coils are set up in the lanes, and the moving vehicles are detected by using two methods: the connectivity domain analysis method and the regional grey level mean value method. The experimental results show that the regional grey mean method can solve the problem of vehicle leakage detection. The accuracy of vehicle counting is significantly improved, and the accuracy can reach over 94%.

3. The Proposed Method

In this study, we propose a real-time traffic flow statistics method of two stages based on dual-granularity classification. The first stage is implemented using YOLOv5 to obtain all the motor and non-motor vehicles appearing in the scene. The second stage is implemented using EfficientNet to obtain the motor vehicles obtained in the previous stage, and such vehicles are classified into six categories: the car, bus, minivan, medium truck, large truck, and container truck. Through this dual-grained classification, the problem of misclassifying non-motorized vehicles as motorized vehicles is solved, and only the model categories of motorized vehicles are classified among the motorized vehi-

cles. This simplifies the problem and significantly reduces the probability of misclassification. To correlate the front and back frames of the video, vehicle tracking is implemented using DeepSORT, and vehicle re-identification is achieved by combining the ResNet50 model. This reduces the number of identity transformation and improves the tracking accuracy. Finally, the dual-lane counting method is used to alleviate the identity transformation, counting failure, and repeated counting caused by vehicles in the middle area owing to occlusions.

3.1. Dataset Processing

The first stage of this method is to identify all the motorized and non-motorized vehicles in a scene. To train the YOLOv5 model to achieve the above target, the following dataset is constructed. The frames and filtered real scene videos from professional testing organizations are transformed to obtain 1500 images. We use the labeling tool to draw frames to label all the vehicles appearing in the images. The images in the dataset are processed by adding noises. This enables the model to adapt to complex shooting environments, enhance generalization capability, and improve robustness. To address the misclassification of non-motorised vehicles as motorised vehicles, motorised vehicles were labelled as jdc category and non-motorised vehicles were labelled as fjdc category.

The second stage involves the acquisition of motor vehicles obtained in the previous stage which are classified into six categories: cars, buses, minivans, medium trucks, large trucks, and container trucks. This stage is difficult, but plays a major role in the final classification effect and is difficult. Therefore, a large amount of data is required to train the model. We construct the required dataset in the following manner. Given four hours of real-scene videos from professional testing organizations, four videos are first selected every second. After the initial manual screening, the YOLO model is used to remove the vehicles in each image to remove the interference of scene information. Subsequently, the effective images are screened manually, and the vehicle data is stored according to six categories: the small bus, large bus, small truck, large truck, container truck and medium truck. After data enhancement, a small number of vehicle types contain 32,034 images. The main data-enhancement methods used here include the adjustment of brightness, contrast and saturation, and the addition of Gaussian noises.

To reduce the number of identity transformation and improve the tracking accuracy, we train the ResNet50 model to implement vehicle re-identification based on the DeepSORT model. To train ResNet50 for vehicle re-recognition, we extend the dataset by rotating, scaling, and adding noises to videos (of real scenes from professional testing organizations after transforming, cropping, and filtering the frames) to increase the diversity of the samples. The dataset is constructed such that images (of various perspectives from each vehicle) are stored in the same folder. Moreover, a label file is written to record the vehicle category corresponding to the vehicles in each folder. The 854 files are divided into six categories: cars, buses, minivans, medium trucks, large trucks, and container trucks.

3.2. Model Architecture

The design of the algorithm model for the proposed method is shown in Figure 6.

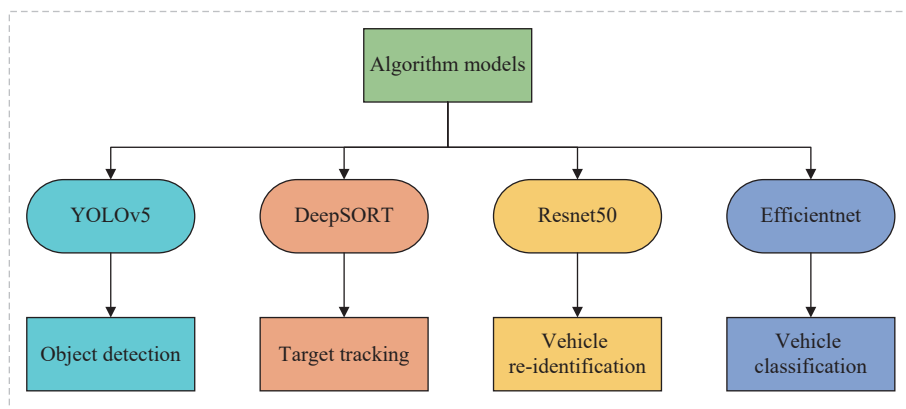


Figure 6. Algorithm model design diagram.

Based on dual-granularity classification, the proposed method for real-time traffic flow statistics is implemented conjointly using four model algorithms. To achieve dual-grained classification, in the first stage of the coarse-grained classification, we identify the motor and non-motor vehicles in the scene in real time using the YOLOv5m model. In the second stage of the fine-grained classification, we categorize the motor vehicles acquired in the first stage into six types of vehicles (cars, buses, minivans, medium trucks, large trucks, and container trucks) with the aid of the EfficientNet model. The DeepSORT model is introduced to achieve a connection between the front and back frames acquired during video detection. Nevertheless, the original method could achieve only pedestrian re-identification. Therefore, we combine it with the ResNet50 model to achieve vehicle re-identification.

Based on the above discussions, a counting method should be introduced for traffic counting. It is observed that the existing methods count in the middle region. Nevertheless, this approach is not suitable for lateral two-way lane counting scenarios, where the middle region can cause many identity transformation, counting failures, and duplicate-counting problems owing to the occlusion of vehicles in the two-way lanes. To solve these problems, this study proposes a lane-splitting counting method that realizes staggered counting of two lanes through two separate counting areas. Specifically, the blue and yellow lines are set at $1/3$ and $2/3$, respectively, from left to right. The counting point is the upper-right corner of the detection frame. When the vehicle travels from left to right, i.e., when the front end hits the blue line, the number is added to the blue and yellow array. When the vehicle is driving from right to left, i.e., when the rear of the vehicle hits the yellow line, the number is also added to the blue and yellow arrays. Two separate counting areas are selected to achieve staggered counting of vehicles traveling in the left and right lanes at approximately $1/3$ of the path. If one of the counted areas is not hit, the other area could still be remedied. This approach further reduces the problem of identity alterations when vehicles are tracked. Thereby, it solves the problem of incorrect or duplicate counts due to identity alterations.

After realizing the above functions, it is only necessary to connect the above models and methods in series and adjust the input and output to realize the method. The specific operation process is as follows. First, we initialize two areas (to be counted by crashing lines) whose values are assigned as 1 and 2. Blue and yellow are used in this study. Subsequently, the YOLOv5 model is loaded, and all the motorized and non-motorized vehicles in the scene are detected. The detected frames labeled with jdc are passed into DeepSORT for vehicle tracking, the EfficientNet model is loaded for vehicle classification, and the frame labels are updated. During the tracking process, the ResNet50 model is used to extract the appearance features of vehicles in the prediction frames before and after the frames. In addition, it is compared with the vehicles in the detection frames to achieve re-identification. When the vehicles reach the counting area, the vehicle models are assessed and counted. A flowchart of the above system is shown in Figure 7.

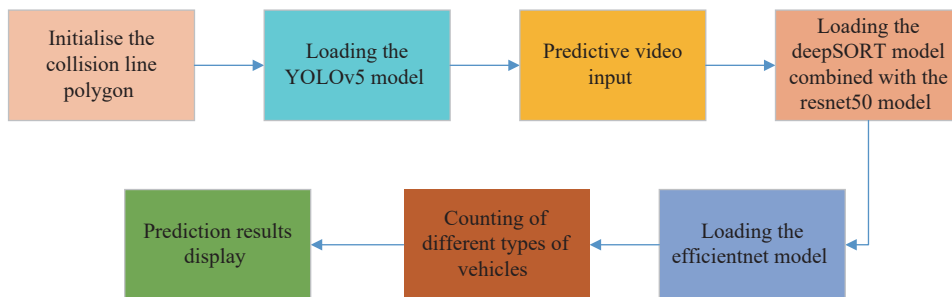


Figure 7. Traffic survey equipment detection and tracking system flow chart.

4. Experimental Results and Analysis

4.1. Evaluation Metrics

The samples can be divided into the following four categories.

- a. True Positive (TP) indicates the number of samples that are positive with positive classification results.
- b. False Negative (FN) indicates the number of samples that are positive with negative classification results.
- c. False Positive (FP) indicates the number of samples that are negative with positive classification results.
- d. True Negative (TN) indicates the number of samples that are negative with negative classification results.

There are also many types of evaluation criteria for the model. These include indicators such as accuracy, error rate, precision, recall, and statistical accuracy. The performance of the model is assessed through the comprehensive evaluation of multiple indicators.

1) Accuracy

It is calculated as the number of samples that are classified as pairs divided by all the samples. The higher the accuracy, the higher the number of correctly classified samples, and the better the classifier performance. The formula for the correctness rate is

$$acc = \frac{(TP + TN)}{(TP + FP + TN + FN)} \quad (1)$$

2) Error Rate

The error rate indicates the number of samples with incorrect prediction. The formula for the error rate of a sample is

$$error_rate = \frac{(FP + FN)}{(TP + FP + TN + FN)} \quad (2)$$

3) Precision

The positive-sample accuracy rate indicates the number of samples that are positive because the prediction is positive. The formula for the positive sample precision rate is

$$pre = \frac{TP}{(TP + FP)} \quad (3)$$

The precision of a negative sample can be calculated similarly.

4) Recall

The positive sample recall is an indication of the number of positive samples that also have positive prediction. The formula for positive sample recall is

$$rec = \frac{TP}{(TP + FN)} \quad (4)$$

Similarly, one can calculate the recall of negative samples.

5) Statistical Accuracy

Statistical accuracy is an indication of the number of samples that are detected and predicted correctly. If we assume that the number of samples (predicted correctly) is T, the number of samples (predicted incorrectly) is R, and the number of samples overlooked is L, then the formula for statistical accuracy is

$$\text{Statistical_accuracy} = \frac{T}{(T + R + L)} \quad (5)$$

4.2. Results of the First Stage of Classification Experiment

The model training parameters are shown in the second column of [Table 1](#). The initial learning rate is set to 0.01. The SGD optimizer is used, and the weight attenuation is set to prevent the model-overfitting problem (which can be observed after analyzing the training results). The current detection classification accuracy can reach 0.995. The vehicle detection accuracy is 0.9991, which can completely identify all the vehicles in the scene. The method in this study can largely alleviate the problem of misclassifying non-motorized vehicles as motorized vehicles. Its accuracy on the validation set is 0.999, which can completely identify non-motorized vehicles in the scene. This ensures that the second stage can classify motor vehicles only by the model. This reduces the difficulty of the problem. Accuracy and recall curves with the number of training sessions are shown in [Figure 8](#). YOLOv5 training has three loss values: box_loss, obj_loss, and cls loss. These values represent the localization loss (which indicates the error between the prediction box and calibration box), confidence loss (which indicates the confidence of the grid and whether there is a target in the grid for dichotomous classification), and classification loss (which determines whether the category labeled on the detection box is correct). The training loss curve for this training set is shown in [Figure 9](#).

Table 1 Model training parameters

Model	YOLOv5	ResNet50	EfficientNet
Initial learning rate	0.01	0.01	0.01
Optimizer	SGD	SGD	Adam
Weight decay ratio	5e-4	~	1e-4
Learning rate decay ratio	~	0.1	0.8
Input image size	640×640	224×224	224×224
Batch_size	24	24	40
Epoch	100	80	100

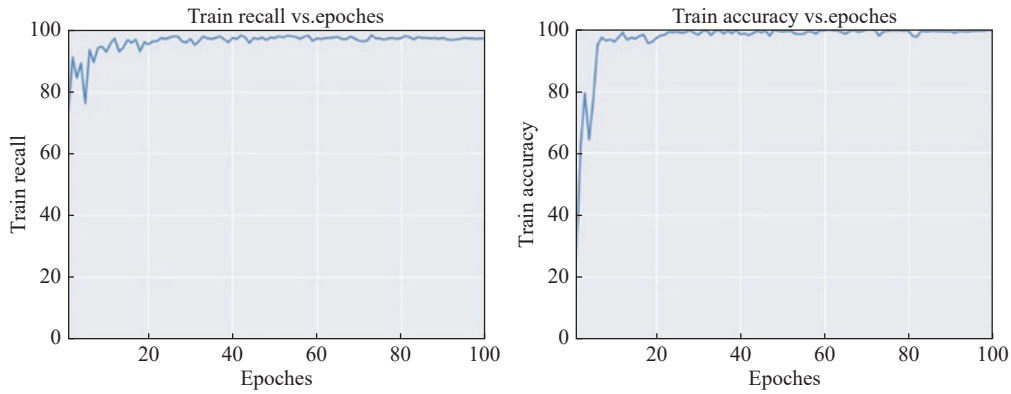


Figure 8. Variation curves of accuracy and recall with epoch.



Figure 9. Box, obj, and cls loss variation curve with epoch.

4.3. Results of the Second Stage of Classification Experiment

The model training parameters are shown in the fourth column of [Table 1](#). The initial learning rate is set to 0.01, the Adam optimizer is used, and the learning rate is set to accelerate the convergence of the model. The early stopping strategy and weight attenuation prevent the overfitting problem. Analysis of the training results reveals that the accuracy stabilizes at approximately 0.998. In addition, the initial accuracy is high. Their data similarity is large accounting for the data augmentation used in the production process of the EfficientNet dataset. Considering that the training and validation sets are divided automatically in the code, the accuracy is perpetually high. The statistical accuracy of the current system reaches 98.7%. Therefore, the classification accuracy of the model satisfies the requirement of the classification task of the test video in this scenario. The accuracy variation curve with the number of training times and the loss variation curve with the number of training times are shown in [Figure 10](#).

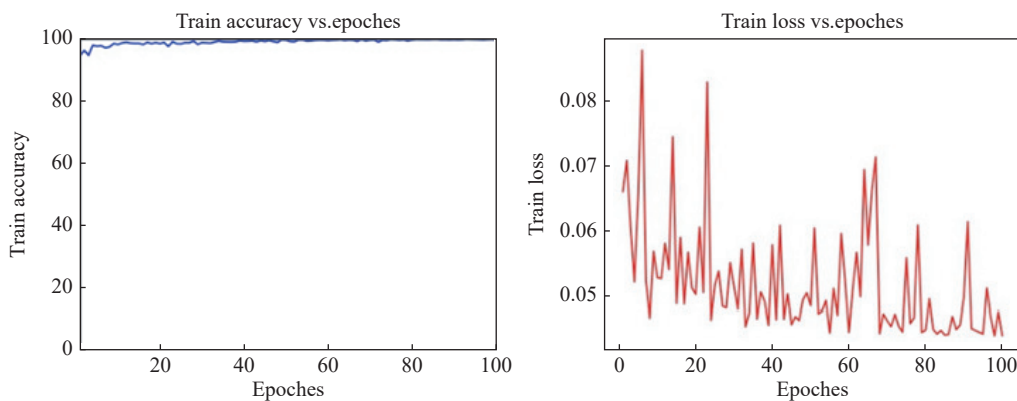


Figure 10. Variation curves of accuracy and loss with epoch.

4.4. Experimental Results of Vehicle Re-Identification

The model training parameters are listed in the third column of [Table 1](#). Here, the initial learning rate is set to

0.01, the SGD optimizer is used, and the learning rate decay is set to accelerate the model convergence. Analysis of the training results shows that the current classification accuracy and recognition accuracy are approximately 90% on average. Given that the model only re-identifies vehicles during the DeepSORT tracking process to improve the tracking accuracy, the impact on the overall statistical accuracy of the system is relatively marginal. A test reveals that the use of this vehicle re-identification model significantly reduces the probability of tracking failures and identity transformation. Combined with the crash line counting technique, the problem of repeated counting (owing to identity transformation in the tracking process) can be solved. The accuracy variation curve (with respect to the number of training times) and the loss curve (with respect to the number of training iterations) are shown in Figure 11.

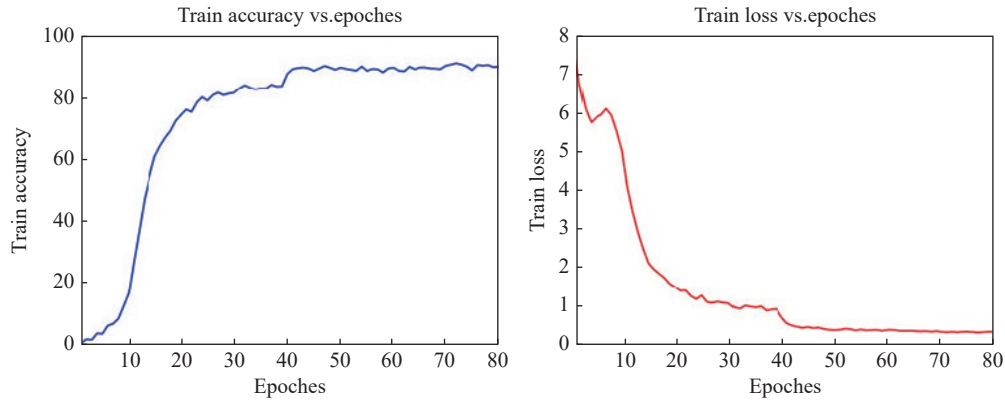


Figure 11. Accuracy and loss curves with epoch.

4.5. Overall Performance

The overall performance of this method is presented here. The model performance is presented in Table 2. The methods listed in the table display statistical accuracy only when a video is detected and tracked. The table lists the results of the YOLOv5 and EfficientNet models for real-time high-accuracy detection. Subsequently, single classification using only YOLOv5 is provided. Although real-time detection can be achieved, the statistical accuracy is low. The reasons are related to the identification of non-motorized vehicles as motorized vehicles and the identity transformation that occurs during tracking. Next, real-time detection could be achieved and the statistical accuracy is improved to a certain extent, although the detection speed is reduced with the incorporation of the ResNet50 and EfficientNet models into YOLOv5 + DeepSORT. It is observed that by using all the improvements and combining the dual-lane counting method, the real-time detection effect of 45 FPS is achieved with 98.7% statistical accuracy.

Table 2 Overall model performance

Models / Systems	Size	FPS	AP	Statistical accuracy
YOLOv5	640	98	95%	~
EfficientNet	640	55	92%	~
YOLOv5 + DeepSORT	640	75	~	88.6%
YOLOv5 + DeepSORT (ResNet50)	640	68	~	92.4%
YOLOv5 + DeepSORT + EfficientNet	640	50	~	95.1%
Our Approach	640	45	~	98.7%

4.6. Interface Visualization

We implement a visual interface using Tkinter for convenience of use. The forecast video can be selected, and the camera is used to realize a real-time forecast and display the forecast results and their statistics. When the GUI interface is used for prediction, the real-time prediction results are displayed in the middle part. The total number of cars and the number of various models are displayed in real time at the top. This is shown in Figure 12. Clicking on the camera opens it for prediction. If the EfficientNet weights need to be modified, only the weight file in the CLS.py file needs to be altered. To modify the weight of YOLOv5, we need to alter the weight file in Detector.py. To modify the weights of the deep learning model that implements vehicle weight recognition during tracking, the weights need to be altered in the deep_sort.yaml file.

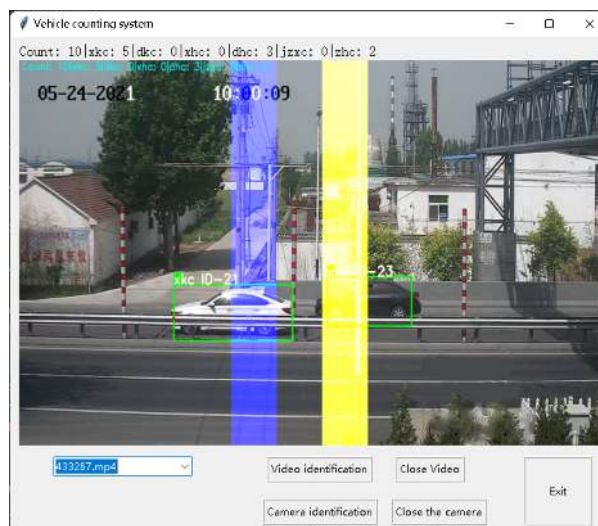


Figure 12. GUI prediction video interface display effect.

5. Conclusion

In this study, a real-time traffic flow statistical method based on dual-granularity classification has been proposed to solve the accuracy discrimination problem of traffic detection equipment. Through this dual-granularity classification, the problem of misdetection of non-motorized vehicles as motorized vehicles has been solved, and the model category of motorized vehicles has been classified only among motorized vehicles. This significantly reduces the probability of misdetection. Vehicle re-identification has been achieved using DeepSORT combined with the ResNet50 model. This reduces the number of identity transformation and improves the tracking accuracy. The experimental results have shown that, in combination with the dual-lane counting method, the proposed method solves the problem of false detection and vehicle tracking failures with a real-time statistical accuracy of 98.7%. This substantially exceeds the industry requirement for the detection accuracy ($> 90\%$) of the trunk line interchange equipment. Nevertheless, the detection results still show problems such as missing vehicles and duplicate counting owing to identity alterations. This leaves room for continuous improvement as the proposed method can be used only for side-shot views, which limits its application scope.

Author Contributions: **Yanchao Bi:** method implementation, method testing, dataset production, writing-original draft preparation; **Yuyan Yin:** method testing, data processing and labeling, writing-original draft preparation; **Xinfeng Liu:** data acquisition, idea sorting and guidance, project management, thesis review; **Xiushan Nie:** idea sorting and guidance, answering of difficult questions, guidance on thesis writing, thesis review; **Chenxi Zou:** thesis format review and revision, data processing and labeling; **Junbiao Du:** visualization, thesis format review and revision, data processing and labeling. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Shandong Provincial Key Research and Development Program (Major Science and Technology Innovation Project) under Grant No. 2021CXGC011204.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Acknowledgments: We would like to thank the traffic control department for providing real video surveillance data that supported our experiment with a lot of real data.

References

1. Tan, M.X.; Le, Q.V. Efficientnet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning, Long Beach, USA, 9–15 June 2019*; ICML: Honolulu, 2019; pp. 6105–6114.
2. Wojke, N.; Bewley, A.; Paulus, D. Simple online and realtime tracking with a deep association metric. In *2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017*; IEEE: New York, 2017; pp. 3645–3649. doi:10.1109/ICIP.2017.8296962
3. He, K.M.; Zhang, X.Y.; Ren, S.Q.; et al. Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 27–30 June 2016*; IEEE: New York, 2016; pp. 770–778.

- doi:10.1109/CVPR.2016.90
4. Li, Y.S.; Ma, R.G.; Zhang, M.Y. Traffic monitoring video vehicle volume statistics method based on improved YOLOv5s+DeepSORT. *Comput. Eng. Appl.*, **2022**, *58*: 271–279.
 5. Jin, L.S.; Hua, Q.; Guo, B.C.; et al. Multi-target tracking of vehicles based on optimized DeepSort. *J. Zhejiang Univ. (Eng. Ed.)* **2021**, *55*, 1056–1064. doi:10.3785/j.issn.1008.973X.2021.06.005
 6. Jia, Z.; Li, M.J.; Li, W.T. Real-time vehicle detection at intersections based on improved YOLOv5+DeepSort algorithm model. *Comput. Eng. Sci.*, **2023**, *45*: 674–682.
 7. Kumar, S.; Jailia, M.; Varshney, S.; et al. Robust vehicle detection based on improved you look only once. *Comput., Mater. Continua*, **2022**, *74*: 3561–3577.
 8. Zhang, W.L.; Nan, X.Y. Road vehicle tracking algorithm based on improved YOLOv5. *J. Guangxi Norm. Univ. (Nat. Sci. Ed.)* **2022**, *40*, 49–57. doi:10.16088/j.issn.1001-6600.2021081303
 9. Gai, Y.Q.; He, W.Y.; Zhou, Z.L. Pedestrian target tracking based on DeepSORT with YOLOv5. In *2021 2nd International Conference on Computer Engineering and Intelligent Control (ICCEIC), Chongqing, China, 12–14 November 2021*; IEEE: New York, 2021; pp. 1–5. doi:10.1109/ICCEIC54227.2021.00008
 10. Zhou, X. A deep learning-based method for automatic multi-objective vehicle trajectory acquisition. *Trans. Sci. Technol.* **2021**, *135*–140, 144.
 11. Zhang, Q. Multi-object trajectory extraction based on YOLOv3-DeepSort for pedestrian-vehicle interaction behavior analysis at non-signalized intersections. *Multimedia Tools Appl.*, **2023**, *82*: 15223–15245.
 12. Chen, X.W.; Jia, Y.P.; Tong, X.Q.; et al. Research on pedestrian detection and DeepSort tracking in front of intelligent vehicle based on deep learning. *Sustainability*, **2022**, *14*: 9281.
 13. Zhan, W.; Sun, C.F.; Wang, M.C.; et al. An improved yolov5 real-time detection method for small objects captured by UAV. *Soft Comput.*, **2022**, *26*: 361–373.
 14. Ye, L.L.; Li, W.D.; Zheng, L.X.; et al. Multiple object tracking algorithm based on detection and feature matching. *J. Huaqiao Univ. (Nat. Sci.)* **2021**, *42*, 661–669. doi:10.11830/ISSN.1000-5013.202105018
 15. Wang, S.; Wang, Q.; Min, W.D.; et al. Trade-off background joint learning for unsupervised vehicle re-identification. *Visual Comput.*, **2023**, *39*: 3823–3835.
 16. Wang, N.T.; Wang, S.Q.; Tang, L.; et al. Insulator defect detection based on EfficientNet-YOLOv5s network. *J. Hubei Univ. Technol.*, **2023**, *38*: 21–26.
 17. Zhang, K.J.; Wang, C.; Yu, X.Y.; et al. Research on mine vehicle tracking and detection technology based on YOLOv5. *Syst. Sci. Control Eng.*, **2022**, *10*: 347–366.
 18. Zhang, X.; Hao, X.Y.; Liu, S.L.; et al. Multi-target tracking of surveillance video with differential YOLO and DeepSort. In *Proceedings of the SPIE 11179, Eleventh International Conference on Digital Image Processing, Guangzhou, China, 14 August 2019*; SPIE: San Francisco, 2019; pp. 701–710. doi:10.1117/12.2540269
 19. Li, M.A.; Zhu, H.J.; Chen, H.; et al. Research on object detection algorithm based on deep learning. In *Proceedings of 2021 3rd International Conference on Computer Modeling, Simulation and Algorithm, Shanghai, China, 4–5 July 2021*; IOP, 2021; p. 012046. doi:10.1088/1742-6596/1995/1/012046
 20. Wen, N.; Guo, R.Z.; He, B. Multi-lane vehicle counting based on DCN-Mobile-YOLO model. *J. Shenzhen Univ. (Sci. Eng.)* **2021**, *38*, 628–635. doi:10.3724/SP.J.1249.2021.06628
 21. Dai, J.H.; Guo, J.S. Video-based vehicle flow detection and counting for multi-lane roads. *Foreign Electron. Meas. Technol.*, **2016**, *35*: 30–33.

Citation: Chao, B.; Yan, Y.; Feng, X.; et al. Real-Time Traffic Flow Statistics Based on Dual-Granularity Classification. *International Journal of Network Dynamics and Intelligence*. 2023, 2(3), 100013. doi: 10.53941/ijndi.2023.100013

Publisher’s Note: Scilight stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2023 by the authors. This is an open access article under the terms and conditions of the Creative Commons Attribution (CC BY) license <https://creativecommons.org/licenses/by/4.0/>.