LATEX

1023040911 Qihang Cai

June 12, 2024

1 introduction

The identification of intelligent transportation systems can be traced back to the development of intelligent transportation systems (ITS) and the advancement of computer vision and artificial intelligence technology. In the early 1980s, with the acceleration of urbanization and the increasing popularity of automobiles, traffic management faced increasing challenges, such as traffic congestion and frequent traffic accidents, which urgently needed to be solved. In order to solve these problems, people began to explore how to use advanced information technology to improve traffic management and service levels, and thus intelligent transportation systems emerged. Traffic signs and markings, as important components of traffic management, are crucial for ensuring road traffic safety. Therefore, using computer vision and artificial intelligence technology to recognize traffic signs and markings has become an important research direction. The development of this technology benefits from the advancement of computer vision technology, enabling computers to "see" and "understand" the content in images like humans. Meanwhile, the application of artificial intelligence technology endows computers with the ability to learn and reason, enabling them to process and analyze image data more intelligently. The application of traffic sign recognition technology is of great significance in intelligent transportation systems. Firstly, it can help traffic management departments achieve automatic monitoring and recognition of traffic signs, improving the management efficiency of traffic signs. Traditional traffic sign inspections usually require manual patrols, which consume manpower and resources, and there are problems of missed and false detections. Through the identification technology of intelligent transportation systems, it is possible to monitor traffic signs 24/7, greatly reducing labor costs and improving the accuracy and efficiency of monitoring. Secondly, traffic sign recognition technology can also provide drivers with real-time traffic sign information, improving driving safety. When drivers encounter complex traffic signs while driving, they often need to be distracted to find and understand the meaning of the signs, which can easily lead to distraction and driving errors. Through the identification recognition technology of intelligent transportation systems, the identified traffic sign information can be displayed in real-time on the driver's onboard display screen, making it more convenient and efficient for the driver to obtain traffic sign information, improving driving safety and comfort. In addition, traffic sign recognition technology can also provide rich traffic data support for traffic management departments and provide scientific basis for traffic management decision-making. By identifying and analyzing traffic signs, real-time monitoring and evaluation of road traffic status can be achieved, helping traffic management departments to timely discover and solve traffic safety hazards, optimize traffic signal control strategies, improve road traffic efficiency, and thus achieve intelligent and refined traffic management. Therefore, the development of intelligent transportation system identification technology is of great significance. It can not only improve the management efficiency and driving safety of traffic signs, but also provide rich traffic data support for traffic management departments, promoting the intelligent and refined development of traffic management. With the continuous progress of computer vision and artificial intelligence technology, it is believed that intelligent transportation system identification technology will play an increasingly important role in the future, making greater contributions to the improvement of traffic management and service levels. The focus of this article is to establish a traffic sign detection application to detect popular traffic signs in Vietnam. This application receives traffic videos as input. Then, it locates the area of traffic signs in the video and recognizes these signs. In order to train a traffic sign detection model, we established a large dataset containing 16770 images of 54 types of traffic signs. The performance of the proposed application has been tested and evaluated in various metrics. Based on the experimental results, the detection errors in the application were analyzed.

2 related works

Sign recognition in intelligent transportation systems has been a research hotspot in the fields of computer vision and artificial intelligence in recent years. Through image recognition and analysis of traffic signs on the road, intelligent transportation systems can achieve automatic detection, recognition, and understanding of traffic signs, thereby improving traffic management efficiency and driving safety levels. The primary task of sign recognition is to detect and locate traffic signs in road images. Traditional methods are mainly based on feature engineering, such as Haar features and HOG features. These methods can to some extent detect the position of markers, but their adaptability to complex scenes and occlusion situations is limited. In recent years, the rise of deep learning technology has made detection methods based on convolutional neural networks (CNN) mainstream. For example, object detection models such as Yolov3 and Faster R-CNN have achieved good results in traffic sign detection [1]. Once traffic signs are detected and located, the next task is to classify and recognize them. Traditional classifiers such as SVM and KNN often rely on manually designed features, which to some extent limits their performance. In contrast, deep learning based methods can learn higher-level feature representations, thereby achieving better classification performance. For example, classic CNN models such as AlexNet, VGG, and ResNet perform well in image classification tasks [?, ?]

. For the problem of joint recognition and understanding of multiple signs, traditional methods mainly rely on rules and logical reasoning, which often rely on manually designed rules and constraints, making it difficult to cope with complex and changing road conditions. In recent years, deep learning models such as Graph Neural Networks (GNN) have been introduced into multi label joint recognition, achieving more accurate and robust recognition results by modeling the semantic relationships between labels [?]. Real time and robustness are another challenge faced by sign recognition in intelligent transportation systems. In order to improve real-time performance, researchers usually use methods such as model lightweight and hardware optimization. For example, lightweight CNN models such as MobileNet and EfficientNet significantly reduce computational complexity while maintaining high accuracy, making them suitable for deployment on embedded devices [?]. In order to improve robustness, researchers usually use methods such as data augmentation and domain adaptation to ensure that the model can maintain good generalization performance in different environments. The continuous updating and iterative optimization of sign recognition technology in intelligent transportation systems is crucial for their practical applications. With the continuous updating and evolution of traffic signs, models need to be able to adapt to new sign types and layouts in a timely manner. At the same time, by continuously collecting and providing feedback on actual data, the model is optimized and improved to make its performance more robust and reliable in practical scenarios. Therefore, sign recognition in intelligent transportation systems involves research on multiple aspects such as object detection, image classification, multi sign joint recognition, real-time performance, robustness, and automatic updating. With the continuous development and popularization of deep learning technology, it is believed that sign recognition technology in intelligent transportation systems will continue to make new breakthroughs and progress, providing stronger support for the development and application of intelligent transportation systems.

3 system architecture

The system inputs traffic video frames and uses a transfer learning model based on YOLOv4 to detect traffic signs in each video frame, obtaining the labels of these signs. Then, the content of these tags is displayed to the user through a web-based interface of the system. The transfer learning model based on Yolov4: YOLOv4 [?] has many special enhancements that can improve the accuracy and speed of its brother YOLOv3 [?] on the same COCO dataset and V100 GPU. The structure of v4 is divided into four parts: backbone, neck, dense prediction, and sparse prediction. The backbone network for target recognition is usually pre trained through ImageNet classification problems. Pre training means that the weights of the network have been adjusted to recognize relevant features in the image, although they will be fine tuned in new object detection tasks. The author is considering using the backbone: CSPResNext50, CSP-Darknet53, EfficientNet-B3. The neck is responsible for mixing and matching the feature maps learned through feature extraction (backbone) and recognition process (YOLOv4 called dense prediction). YOLOv4 allows customization of neck structures, such as FPN, PAN, NASFPN, BiFPN, ASFF, SFM, SSP.

4 experiment

4.1 datasets

This article adopts two methods to collect traffic sign datasets: collecting images and videos from Google search pages. Most of the data is collected through video recording because it is closer to reality, the diversity of backgrounds, and the minimal noise caused by images provided on Google. The video recording is divided into two directions: one is for actual combat (shooting traffic signs on the street), and the other is based on satellite projected images onto Google Maps and then transmitted back to the screen. More images will be collected for the first direction than for the second direction. However, the second direction is used to supplement data on symbols that are difficult to encounter in real life, as it is impossible to correctly locate the remaining symbols. If this second direction still does not match the quantity, the sewn sample logo will be integrated into the actual context to create a realistic image and ensure the quantity of the logo. The collected signs are common signs that may be encountered in daily life based on the label names in the traffic manual. There are a total of about 54 labels, of which 53 are single signs, a type of image that is considered complex or non-existent in the selected number of signs. This label was added for future issues. After annotating the images and videos, there were a total of 16770 images, including 13439 in the training set and 3331 in the test set. Figure 2 shows a statistical chart of the number of assigned labels.

4.2 data preprocessing

Each image has many different features. Therefore, to use image data in the model, several preprocessing steps must be taken. The following are the preliminary preprocessing steps for the image dataset: - Read the image, then convert the color channels of all images to RGB format to create consistency in the number of color channels for all images to match the model input. Adjust the photo size to the appropriate size - height: 416 pixels, width: 416 pixels. So all the images are converted to 416 * 416 *The size of 3. After preprocessing, we use yolov4 to train the labeled images in the dataset, with the following parameters: batch = 64, subdivision = 16, max batch = 108000, steps = 86400,97200, filters = 177, classes = 54, width = 416, height = 416.

4.3 method evaluation

The performance indicators of object detection problems include - IoU (Intersection over union), which measures the ratio of the degree of intersection between two contours (usually the predicted contour and the actual contour) to determine whether two frames overlap. This ratio is calculated based on the intersection area of two contour lines and the total area of their intersection and non intersection. Accuracy measures the accuracy of a model's predictions, which is the percentage of correct predictions made by the model. Recall the degree to which the measurement model finds all positive patterns. From the definition of precision and recall above, we can also evaluate the model by changing a threshold and observing the values of precision and recall. The concept of area under a curve (AUC) also has a similar definition. For the precision recall curve, AUC has another name, Average Precision (AP). Assuming there are N thresholds for precision and recall, each corresponding to a pair of precision values, with a recall of Rn, n=1,2,..., N. The method of drawing the Precision Recall curve is to draw each point on the coordinate axis using coordinates (Pn) and connect them together. In multi class object detection, mAP is the average AP calculated for all classes.

4.4 result

During the long training period (specifically, about 4000 rounds/day of training, completed in approximately 27 days), there were many models trained in 10000 rounds, 20000 rounds,... 10000; And the model saved from the mAP calculation of each small circle. And compare the obtained models. Finally, it is concluded that mAP@0.5 =94.81%, mAP@0.75 =68.53% of the models are optimal. From Figures 3a and 3b, it can be seen that the overall evaluation of the dataset by the model is very good, mAP@0.5 Up to 94.81%, mAP@0.75 =68.53%, only a few cases are not high, such as class ID=7 which is a sign prohibiting motorcycles and tricycles, with very low accuracy, rated as mAP@0.5 When AP=22.85%, rated as mAP@0.75 When AP=0.

5 conclusion

In the future, we plan to improve and expand the dataset by recording more videos of different routes, making it closer to real life. We have also developed other methods to improve the quality of recognizing QR codes by combining more frames. In this article, we used the Yolov4 method. The recognition result of this method for mAP @ 0.5 is as high as 94.18%, but it is effective for mAP@75 Due to the presence of signs prohibiting motorcycles and three wheeled vehicles that cannot be recognized, the recognition result is very low, only 68.53%. The accuracy of identifying these above symbols is extremely low: mAP@0.5 AP=22.85%, mAP@0.75 AP=0. This is because the number of different symbols in the dataset is uneven. In the future, we plan to improve and expand the dataset for logo types with a small number of images. In addition to identifying these signals, we will also develop problems based on the images collected from the dash cam and provide explanations or warnings. We hope to contribute this dataset to various sectors of society to stimulate research on traffic sign recognition and improve recognition efficiency with better methods.