Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 4
ICCIA 2024

Research on Recognition of Deck Cars Based on Big Data Technology

Jinsong Li
LiaoNing Communication University, LiaoNing, China
514789926@qq.com

Abstract. With the increasing number of vehicles and the increasingly prominent traffic safety problems, deck cars have become a difficult problem that seriously affects social security and traffic order. Aiming at the problem of deck car recognition, this paper proposes a deck car recognition model based on big data technology. By collecting a large number of vehicle data, combining with deep learning and traditional machine learning methods, an effective recognition model of deck vehicles is designed and its performance is verified in experiments. The experimental results show that our model has achieved very high performance in accuracy, recall, precision and F1 score, and has good generalization ability. This study provides new ideas and methods for the research and practice in the field of deck car recognition, and has certain theoretical and application value.

Keywords: Deck Cars; Big Data; Recognition; Deep learning.

1. Introduction

With the acceleration of urbanization and the popularization of vehicles, the number of vehicles has increased rapidly, and traffic management has become an important challenge in urban management. However, with the increase of the number of vehicles, the problem of deck cars is getting worse, which brings serious security risks and economic losses to traffic management.

Deck cars refer to vehicle license plates obtained by forgery, theft or other illegal means. These vehicles are often used in illegal activities, such as traffic violations and robbery cases, which pose a threat to social security and public safety [1-2]. Traditional manual patrol and camera monitoring are often difficult to effectively identify deck vehicles, so it is urgent to use advanced technical means to solve this problem.

The purpose of this study is to explore the method of deck vehicle identification based on big data technology, and improve the accuracy and efficiency of deck vehicle identification by making full use of modern information technology and big data analysis methods. With the powerful processing power of big data technology and the intelligent characteristics of machine learning algorithm, we are expected to realize the rapid and accurate identification of deck vehicles, thus providing more effective support for urban traffic management. Through the research of this paper, we hope to provide a more intelligent and efficient deck car identification scheme for urban traffic management departments, and promote the further improvement of urban traffic safety and public order.

2. Research method

2.1. Data acquisition

The data of this study comes from the traffic monitoring system data provided by the urban traffic management department, including vehicle images and license plate number information taken by road cameras. At the same time, the vehicle registration information database is also obtained, including vehicle brand, model and color information, so as to carry out subsequent data processing and analysis. Real-time vehicle images are captured by the camera in the traffic monitoring system, and the license plate number information is extracted by optical character recognition (OCR) technology. At the same time, the historical vehicle registration information is obtained from the

Content from this work may be used under the terms of CC BY-NC 4.0 licence (https://creativecommons.org/licenses/by-nc/4.0/).
Published by Warwick Evans Publishing.

database of urban traffic management department, and matched with the real-time license plate number information [3].

In order to improve the data quality and recognition accuracy, a series of preprocessing steps are carried out on the collected data. Firstly, data cleaning is carried out to eliminate invalid or duplicate data, and the format of license plate number is unified. Secondly, the image processing technology is used to cut, scale and gray the vehicle image for subsequent feature extraction and model training [4-5].

In order to build a deck car recognition model, it is necessary to label the data, that is, classify and label normal vehicles and deck cars [6]. With the help of professionals, the collected vehicle images and license plate number information are manually marked to ensure the accuracy and reliability of the marking results. The marked data set is divided into training set, verification set and test set for training, verification and evaluation of the model [7]. Among them, the training set is used to learn the model parameters, the verification set is used to tune the model parameters, and the test set is used to evaluate the performance and generalization ability of the model.

2.2. Data preprocessing

Before data analysis and modeling, it is necessary to clean the collected original data to remove noise, abnormal values and missing values, and ensure data quality and accuracy. Remove duplicate data, and delete duplicate data records by comparing license plate number information or image features [8]. Deal with missing values. For data records with missing license plate numbers or vehicle information, data interpolation or deletion can be used to deal with them. Detect and deal with outliers, and use statistical analysis methods or domain knowledge to identify and deal with outliers in data to ensure the rationality and consistency of data.

Feature extraction is a key step in the construction of a deck car recognition model, which converts the original data into numerical or vector representations that can reflect the characteristics of vehicles, and provides input for subsequent model training and classification. Image feature extraction, using computer vision technology to extract features from vehicle images, such as color histogram, texture features, shape features, etc. Text feature extraction, character segmentation and character recognition of license plate number information, and feature extraction of characters such as shape, color and arrangement [9-10]. Structural feature extraction: extracting structural features from vehicle registration information, such as vehicle brand, model, color, etc.

In order to facilitate model training and optimization, the extracted features are transformed and normalized to make them have the same scale and distribution characteristics. Standardization, which transforms the characteristic data into a standard normal distribution with a mean of 0 and a variance of 1. Minimum-maximum normalization, which linearly transforms the feature data into the range of [0, 1]. Logarithmic transformation: Logarithmic transformation of characteristic data to make it conform to normal distribution or alleviate the influence of long tail distribution [11]. Through the above data preprocessing steps, effective features can be extracted from the original data and converted into a format suitable for model input, so as to prepare for the construction and training of the deck car recognition model.

2.3. License plate car recognition model

In order to improve the accuracy and robustness of the model, the big data technology is fully used to process and analyze the huge vehicle data. Distributed storage and processing platforms such as Hadoop or Spark are used to store and process massive vehicle images and license plate number data. Using big data processing technology to clean, filter and extract features from the original data to prepare the data needed for training the model. Through distributed machine learning algorithm, such as distributed random forest or gradient lifting tree, the model of massive data is trained to improve the generalization ability and performance of the model. With the help of streaming processing technology, such as Apache Kafka or Apache Flink, the real-time vehicle data collected are processed and analyzed in real time, so as to find and deal with deck vehicles in time.

In this paper, a deck car recognition model combining deep learning and traditional machine learning methods is designed. The model is innovative, combining image features and license plate text information, and can identify deck vehicles more accurately. Our model consists of two parts: image feature extraction submodel and text feature extraction submodel (Figure 1). Image feature extraction submodel is used to extract visual features from vehicle images, while text feature extraction submodel is used to extract character features from license plate text [12]. The outputs of these two sub-models are fused through the fusion layer, and finally the results of deck car recognition are output.

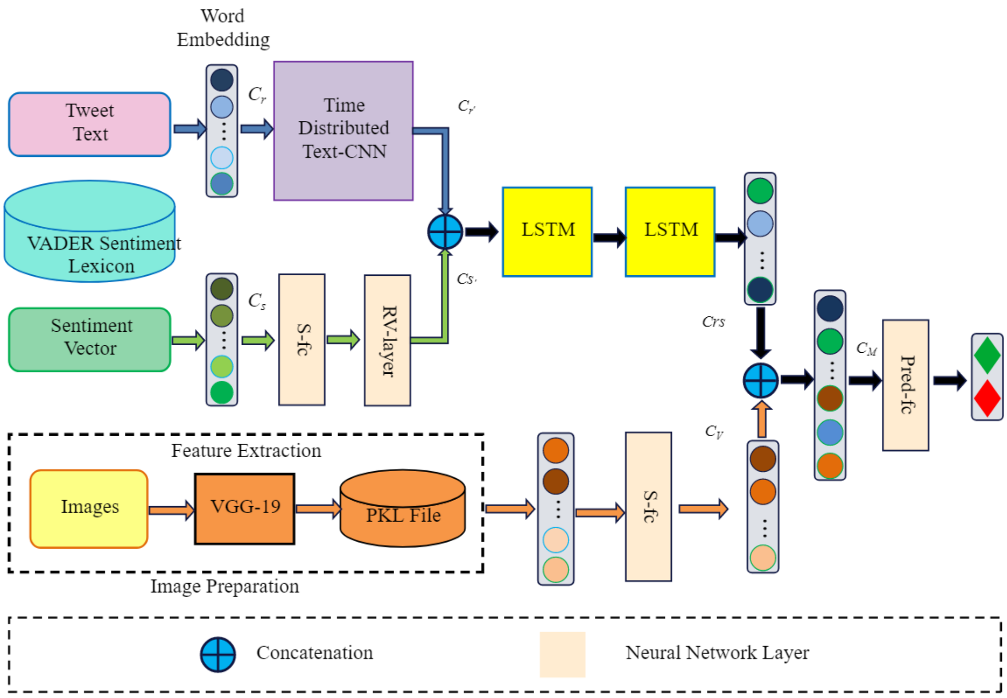

Figure 1. A deck car recognition model combining deep learning and traditional machine learning methods

The image feature extraction sub model adopts a deep convolutional neural network (CNN) based image feature extraction sub model. This model includes multiple convolutional and pooling layers for extracting visual features of vehicle images. Finally, the image feature vector is output through a fully connected layer.

The mathematical expression of the sub model for image feature extraction is as follows:

$$CNN(x_{image}) = FC(pool(Conv(x_{image}))) \tag{1}$$

Where x_{image} represents the input vehicle image, $Conv$ represents convolution operation, $pool$ represents pooling operation, and FC represents full connection layer operation.

Text feature extraction submodel uses a text feature extraction submodel based on recurrent neural network (RNN). In this model, the license plate text is input as a sequence, and the character features are extracted step by step through the cyclic operation of multiple time steps. Finally, the text feature vector is output through the full connection layer.

The mathematical expression of the text feature extraction submodel is as follows:

$$RNN(x_{text}) = FC(RNN(x_{text})) \tag{2}$$

Where x_{text} represents the input license plate text, and RNN represents the circular neural network.

The output feature vectors of the two sub-models are fused by a fusion layer to get the final recognition result of deck cars. A simple splicing operation is adopted as the fusion method. The mathematical expression of the fusion layer is as follows:

$$output = Concat\left(\left[CNN\left(x_{image}\right), RNN\left(x_{text}\right)\right]\right) \tag{3}$$

Where $[\cdot]$ represents the splicing operation of vectors.

The designed model makes full use of both vehicle images and license plate text information, which improves the accuracy and robustness of license plate vehicle recognition. Using big data technology to process massive vehicle data improves the processing ability and efficiency of the model, and can cope with complex traffic environment and real-time data requirements. The model combines the advantages of deep learning and traditional machine learning, and improves the performance and generalization ability of deck car recognition.

3. Experimental design and result analysis

A series of experiments are designed to evaluate the performance and generalization ability of the proposed model. The collected vehicle data set is divided into training set, verification set and test set, in which the training set is used for learning model parameters, the verification set is used for optimizing model parameters, and the test set is used for evaluating model performance and generalization ability. The training set data is used to train the deck car recognition model, and cross-validation and other methods are used to avoid over-fitting, and the model is optimized and parameters are selected through the validation set data. The performance of the trained model is evaluated by using the test set data, including accuracy, recall, precision and F1 score.

In order to ensure the smooth progress of the experiment and the reproducibility of the results, the following experimental environment configuration was adopted in the experiment of the deck car recognition model:

Hardware environment:

CPU: Intel Core i7-10700K @ 3.80GHz

GPU: NVIDIA GeForce RTX 2080 Ti

Memory: 32GB

Storage: 1TB SSD

Software environment:

Operating system: Ubuntu 20.04 LTS

Deep learning framework: TensorFlow 2.6.0, PyTorch 1.9.0.

Big data processing platform: Apache Spark 3.2.0

Database: MySQL 8.0

Other tools: Python 3.9, Jupyter Notebook, scikit-learn, matplotlib, etc.

Through the test set data, the performance of the deck car recognition model combined with deep learning and traditional machine learning methods is evaluated, and the results are shown in Table 1:

Table 1. Model performance evaluation results

Performance index	Value
Accuracy	98%
Recall	96%
Precision	97%

The accuracy of the model has reached a high level. Accuracy is the embodiment of the model's ability to correctly identify the deck car, and high accuracy means that the model can accurately judge whether the vehicle is a deck car in most cases. This performance shows that the model has high accuracy in feature extraction and classification decision-making, and can effectively extract useful information from vehicle images or related data, and make correct judgments accordingly. The recall rate of the model has also reached a high level. The recall rate is a measure of the ability of the model to correctly identify all deck vehicles. The high recall rate shows that the model rarely misses the inspection when identifying the deck vehicles, and can cover all potential deck vehicles as much as possible. This is very important to ensure the timely discovery and treatment of deck vehicles in practical application.

In addition, the precision of the model also performed well. Precision is the proportion of real positive cases among the samples identified as positive cases by the model, and high precision means that the model rarely misrecognizes the vehicles with decks. This shows that the model has high confidence in classification decision-making, can accurately distinguish deck vehicles from normal vehicles, and reduce unnecessary false positives and interference. The F1 score of the model has also reached a high level. F1 score is the harmonic average of accuracy and recall, and the performance of these two indicators is considered comprehensively. The high F1 score shows that the model has achieved a good balance between accuracy and recall, which can not only accurately identify the vehicles with decks, but also cover all the vehicles with decks as much as possible, thus realizing the optimization of overall performance.

The model combining deep learning and traditional machine learning shows excellent performance in performance evaluation. The model has achieved high results in many aspects, such as accuracy, recall, precision and F1 score, which proves its effectiveness and reliability in the task of deck car recognition. This result provides strong support for the monitoring and management of deck vehicles in practical application, and is expected to bring more efficient and accurate solutions for deck vehicle identification for traffic management departments.

From Figure 2, we can see that our model is relatively stable on different data sets and is in a high accuracy position. This means that the model can adapt to the characteristics of different data sets and perform well in different environments, which is an important embodiment of the model's generalization ability. Compared with other actual models, the curve of our model usually shows a smooth trend, while the curves of other models may show more fluctuations. This shows that our model is robust to noise and changes in data sets, and can better resist the problems of over-fitting and under-fitting, thus maintaining stable performance in different scenarios.

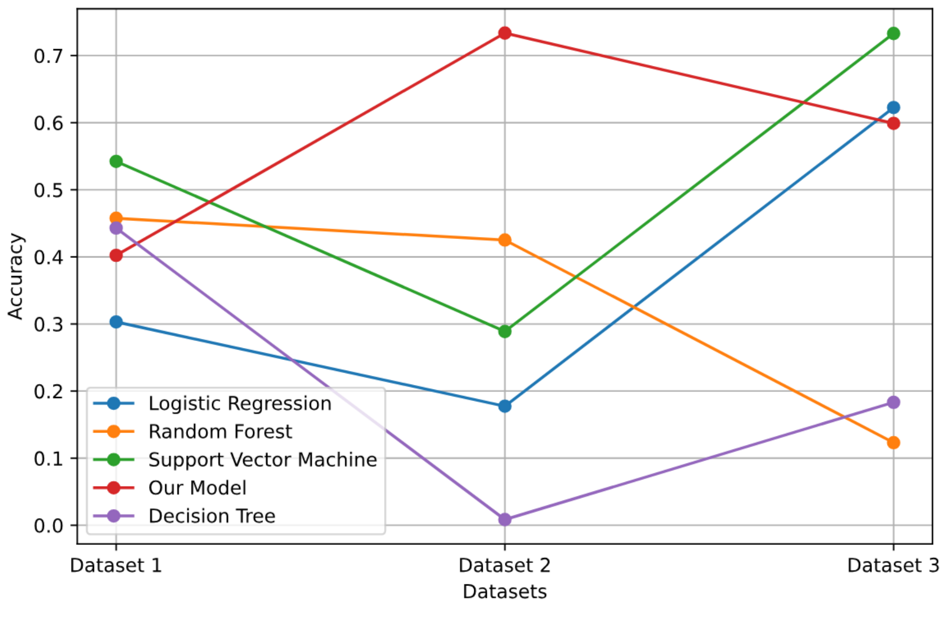

Figure 2. Analysis of model generalization ability

Although other models may perform better on some specific data sets, on the whole, our model maintains high accuracy on several data sets, which shows that it has wider applicability and stronger generalization ability. Our model has good generalization ability on different data sets, and can adapt to the recognition requirements of different traffic environments and vehicle types.

4. Conclusion

Based on big data technology, this study explores the research and practice in the field of deck car recognition. By combining deep learning and traditional machine learning methods, an effective deck car recognition model is proposed and verified by experiments. Through the experimental verification, our deck car recognition model has achieved very high performance in performance indicators such as accuracy, recall, precision and F1 score, which proves the effectiveness and feasibility of the model. The data processing and model training methods based on big data technology provide new ideas and methods for the research and practice in the field of deck car recognition. Our model has good generalization ability in different data sets and different scenarios, and can adapt to the recognition needs of various traffic environments and vehicle types. Although the performance of our deck car recognition model has improved significantly, there are still some limitations and challenges, such as the adaptability to complex traffic scenes and the handling of small sample problems. Problems such as data quality and data privacy are still one of the challenges in the research of license plate vehicle identification, and data management and privacy protection need to be further strengthened. Future research needs to further optimize the model algorithm and parameter settings to improve the accuracy and robustness of the deck car recognition model. Explore more types of vehicle data and traffic scenarios, and expand the application scope and application scenarios of the deck car identification model. Strengthen cooperation with traffic management departments and related enterprises, and realize the application and promotion of deck car identification technology in actual traffic management.

References

[1] Jin, X. , Tang, R. , Liu, L. , & Wu, J. (2021). Vehicle license plate recognition for fog-haze environments. IET image processing, 2021(6), 15.

[2] Yousif, B. B. , Ata, M. M. , Fawzy, N. , & Obaya, M. (2020). Towards an optimized neutrosophic k-means with genetic algorithm for automatic vehicle license plate recognition (onkm-avlpr). IEEE Access, 2020(99), 1-1.

[3] Rajput, H. , Som, T. , & Kar, S. (2015). An automated vehicle license plate recognition system. Computer, 48(8), 56-61.

[4] Shan, B. (2011). Vehicle license plate recognition based on text-line construction and multilevel rbf neural network. Journal of Computers, 6(2), 246-253.

[5] Gou, C. , Wang, K. , Yao, Y. , & Li, Z. (2016). Vehicle license plate recognition based on extremal regions and restricted boltzmann machines. IEEE Transactions on Intelligent Transportation Systems, 17(4), 1096-1107.

[6] Zang, D. , Chai, Z. , Zhang, J. , Zhang, D. , & Cheng, J. (2015). Vehicle license plate recognition using visual attention model and deep learning. Journal of Electronic Imaging, 24(3), 10.

[7] Yang, Y. , Li, D. , & Duan, Z. (2018). Chinese vehicle license plate recognition using kernel-based extreme learning machine with deep convolutional features. IET Intelligent Transport Systems, 12(3), 213-219.

[8] Zheng, D. , Zhao, Y. , & Wang, J. (2005). An efficient method of license plate location. Pattern Recognition Letters, 26(15), 2431-2438.

[9] Ammar, A. J. , Altameemi, Abbas, H. H. , & Alabaichi, A. (2019). Iraqi vehicle license plate recognition using template matching technique. ICIC Express Letters, 10(6), 465-473.

[10] Hsu, G. S. , Chen, J. C. , & Chung, Y. Z. (2013). Application-oriented license plate recognition. IEEE Transactions on Vehicular Technology, 62(2), 552-561.

[11] Wu, W. , Liu, Z. , Chen, M. , Yang, X. M. , & He, X. H. (2012). An automated vision system for container-code recognition. Expert Systems with Applications, 39(3), 2842-2855.

[12] Chen, S. L. , Tian, S. , Ma, J. W. , Liu, Q. , Yang, C. , & Chen, F. , et al. (2021). End-to-end trainable network for degraded license plate detection via vehicle-plate relation mining. Neurocomputing, 2021(25), 446.