Signal Processing 147 (2018) 146–153

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Vehicle color recognition using Multiple-Layer Feature Representations of lightweight convolutional neural network

Qiang Zhang [a,c], Li Zhuo [a,b,c,*], Jiafeng Li [a,c], Jing Zhang [a,c], Hui Zhang [a,c], Xiaoguang Li [a,c]

[a] Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing University of Technology, Beijing, China
[b] Collaborative Innovation Center of Electric Vehicles in Beijing, Beijing, China
[c] College of Microelectronics, Faculty of Information Technology, Beijing University of Technology, Beijing, China

ARTICLE INFO

Article history:
Received 30 July 2017
Revised 6 December 2017
Accepted 16 January 2018

Keywords:
Multiple-Layer Feature Representations
Lightweight convolutional neural network
Vehicle color recognition
Spatial Pyramid Matching

ABSTRACT

In this paper, a vehicle color recognition method using lightweight convolutional neural network (CNN) is proposed. Firstly, a lightweight CNN network architecture is specifically designed for the recognition task, which contains five layers, i.e. three convolutional layers, a global pooling layer and a fully connected layer. Different from the existing CNN based methods that only use the features output from the final layer for recognition, in this paper, the feature maps of intermediate convolutional layers are all applied for recognition based on the fact that these convolutional features can provide hierarchical representations of the images. Spatial Pyramid Matching (SPM) strategy is adopted to divide the feature map, and each SPM sub-region is encoded to generate a feature representation vector. These feature representation vectors of convolutional layers and the output feature vector of the global pooling layer are normalized and cascaded as a whole feature vector, which is finally utilized to train Support Vector Machine classifier to obtain the recognition model. The experimental results show that, compared with the state-of-art methods, the proposed method can obtain more than 0.7% higher recognition accuracy, up to 95.41%, while the dimensionality of the feature vector is only 18% and the memory footprint is only 0.5%.

1. Introduction

Vehicle recognition plays an important part in the applications of Intelligent Transportation System, criminal investigation and so on. Its task is to recognize the type, color and license plate of the targeted vehicle. Color is one of the basic attributes of vehicles, therefore, color recognition plays a significant role in vehicle recognition. However, due to complex impacts of illuminations, weather conditions, noises and image capture qualities, vehicle color recognition has become a challenging task. Firstly, illumination, noises and special weather conditions will lead to obvious changes in visual appearance of vehicle colors. For instance, the vehicle colors under different illuminations vary greatly, which makes the vehicle recognition very difficult. Secondly, considering the differences in camera positions and parameters, there are significant differences in angles and focal distances. In addition, the size and area of vehicles in the image change greatly, which poses another great challenge to vehicle color recognition. Finally, the interference of complex scenes and non-vehicle color components further escalate the difficulty of vehicle color recognition.

In recent years, many researchers have proposed various solutions for vehicle color recognition. In general, the research on vehicle color recognition can be divided into two stages:

The first stage is based on handcrafted features combined with classifiers [1,4–6], which focus on designing various features of vehicle colors manually, such as color histogram [1], color moment [2] and color correlogram [3], and then these features are used to train the classifiers, such as Support Vector Machine (SVM), K-Nearest Neighbor (KNN) and Artificial Neural Network (ANN). Chen et al. [1] extracted the color histogram features and introduced spatial information using Spatial Pyramid Matching (SPM) and Feature Context (FC) to establish Bag of Words (BoW) model, then combined linear SVM to solve the vehicle color recognition problem. Baek et al. [4] also extracted the color histogram of H and S components in HSV color space to construct the bi-dimensional feature vectors and further combined SVM classifiers for vehicle color recognition. Another approach was proposed by Dule et al. [5], which extracted different color features from various color spaces, and tested the vehicle color recognition performance using different classifiers, including KNN, ANN and SVM. The experimental result indicates that ANN classifier combined with 8-bin color

* Corresponding author.
E-mail addresses: zhuoli@bjut.edu.cn (L. Zhuo), zhj@bjut.edu.cn (J. Zhang), huizhang@bjut.edu.cn (H. Zhang), lxg@bjut.edu.cn (X. Li).

https://doi.org/10.1016/j.sigpro.2018.01.021
0165-1684/

histogram feature vector can obtain the highest recognition accuracy.

Apart from above vehicle color recognition method based on global feature extraction, Hu et al. [6] provided an alternate way of vehicle color recognition by constructing the color reflection model to remove the non-vehicle color areas in the image, such as background, vehicle wheels and windows, then directly extracting features from the major vehicle color regions to train the SVM classifier.

These methods mentioned above basically adopt handcrafted features. Therefore, they are tending to achieve higher execution speed but weaker generalization capability and lower recognition accuracy. In addition, the design of handcrafted features requires professional knowledge. An appropriate feature usually takes massive experience and time to validate its effectiveness for a certain task. Therefore, it is difficult to manually design proper features for new data and tasks.

The second stage is based on deep learning. The research of this stage mainly focuses on two aspects: the first is to make use of deep neural network to obtain the feature representation of the images then apply it for vehicle color recognition in combination of traditional classifier. Hu et al. [7] proposed a vehicle color recognition method, in which the deep features of AlexNet [8], and kernel SVM are used, combined with SPM learning strategy. The other is to use an end-to-end deep neural network structure, which merges the feature extraction and classifier into a unified framework through joint optimization. Rachmadi et al. [9] designed a parallel CNN network to achieve end-to-end vehicle color recognition, which learns the recognition model from big data by using two convolutional networks, and integrated the two parts by a fully connected layer. The research results indicate that for vehicle color recognition, compared to the traditional recognition methods of handcrafted features + classifier, the feature representations learnt from deep learning have strong generalization capability and the recognition performance can be improved obviously.

The representative deep network structures used in vehicle color recognition mainly include AlexNet [8], GoogleNet [13], VGG-Net [14], and so on, which are originally designed for complex classification tasks and characteristic of a massive amount of data. Therefore, the network structures usually possess large number of parameters and are easily subject to the occurrence of over-fitting phenomenon. In addition, they require large computational and storage resources. But, recently, with the advance of research, some researchers found that for some specific tasks or applications, a lightweight network structure can also achieve the most advanced result [10].

Deep neural network demonstrates strong learning ability and highly efficient feature extraction capability, which can extract information from low-level raw data to high-level abstract semantic concepts. The hierarchical feature representation can entitle it with prominent advantages when extracting global features and context semantic information of the images. However, the existing methods commonly make use of the output features of the last layer for recognition while neglecting the feature information of the previous layers. Actually, these features of lower layers contain considerable information of images, which may promote the recognition performance. However, if all the features from the intermediate layers can be employed, the extremely high dimension of feature vectors will result in training the classification model too difficult or even failure. To address this problem, a traded-off solution is proposed in this paper. Firstly, a lightweight CNN network architecture is designed for the vehicle color recognition task, which contains five layers, i.e. three convolutional layers, a global pooling layer and a fully connected layer. The feature maps of convolutional layers are used for feature representation of the image. Specifically, SPM strategy is adopted to divide the feature maps of the convolutional layers into four sub-regions. Then, each sub-region is encoded to obtain a feature representation vector of the layer. Next, these feature representation vectors and the output feature vector of the final global pooling layer are normalized and cascaded as a whole feature vector to represent the content of the image. Finally, the linear SVM is used as the classifier for vehicle color recognition. The experimental results demonstrate that, the features from the intermediate layers can help to improve the recognition accuracy. And compared with the state-of-art methods, the proposed method can get higher recognition accuracy by more than 0.7%, up to 95.41%, while with the lower dimensionality of the feature vector and smaller memory size of CNN model.

Table 1
Comparison results of the proposed network structure and AlexNet structure.

	AlexNet	Proposed network
Input	227 × 227 × 3	227 × 227 × 3
Layer 1	conv,96	conv, 48
Layer 2	conv,256	conv, 128
Layer 3	conv,384	conv, 192
Layer 4	conv,384	–
Layer 5	conv,256	–
Layer 6	fc,4096	GAP,192
Layer 7	fc,4096	–
Output	fc,8	fc,8
Memory	227.6M	1.1M

The rest of this paper is organized as follows. Section 2 describes the main ideas and details of the proposed method. Experimental results and analysis are presented in Section 3. Finally, the conclusions are drawn in Section 4.

2. The proposed method

Convolutional neural network has become the dominant machine learning approach for visual recognition tasks. To fulfill complex tasks, which usually need to identify hundreds or even thousands of categories, CNN model usually has a large number of parameters. When used for small and medium size datasets, overfitting often occurs. In vehicle color recognition, color categories and the size of the datasets are both limited. For example, in literature [6], vehicle color contains red, yellow, blue and green, totally four categories. And in [5], the vehicle colors contains white, black, gray, red, blue, green, and yellow, totally seven categories. In this paper, the public vehicle color dataset in [1,7,9] is adopted, which includes eight color categories to recognize. In order to achieve a good tradeoff between the performance and the computational complexity, a lightweight CNN network architecture is designed to extract the features for vehicle color recognition, which includes three convolutional layers, a global pooling layer and a fully connected layer, totally five layers.

The framework of the proposed method is depicted in Fig. 1. Considering that the feature maps of the convolutional layers contains rich information of the vehicle images, all feature representations of the convolutional layers are employed and combined with the output feature vector of the global pooling layer to form a whole vector to represent the content of the images. The linear SVM classifier is used to train the classification model. In the following, the details of the proposed method will be described.

2.1. Lightweight convolutional neural network structure design

The lightweight convolutional neural network architecture designed in this paper is shown with the black dotted line in Fig. 1. The number of neurons at the three convolutional layers is 48, 128,