FINE-GRAINED RECOGNITION OF ROTATING MACHINERY AXIS TRAJECTORY BASED ON DEEP LEARNING

Puchun Yu
Jiangsu Food & Pharmaceutical Science College, Jiangsu 223003, China
Email: yupuchunfood1212@126.com

Abstract - In order to study the detection and recognition of rotating machinery, the attributes of each task and their relations, a joint detection and recognition algorithm for coarse-grained attributes was proposed. The algorithm of vehicle brand recognition was studied especially for fine - grained attribute recognition. First, the colour and type of the rotating machine were fused into the detection algorithm. The multi-task learning framework was used to model the attribute recognition task and positioning task of rotating machinery. At the same time of detection, attribute recognition was completed. Then, the deep learning method (DL) and convolution neural networks (CNN) were introduced, which was widely used characteristics of deep learning approach. Combined with its structural characteristics and generalization capabilities, the CNN structure was analyzed. Finally, experimental tests were conducted. The universal validity and environmental adaptability of the proposed detection algorithm were verified. The results showed that based on the rotating mechanical dataset, the proposed rotating machinery recognition algorithm not only accurately identified the known class samples in the test set, but also identified the samples of unknown categories. Therefore, the proposed rotating machinery detection and identification framework solves the problems in the current solution. The effect of rotating machine detection and recognition is enhanced. The overhead of solution computing resources is reduced. This has practical application value.

Keywords: Deep Learning; Detection of Rotating Machinery; Fine-Grained Identification.

1. Introduction

With the rapid development of urbanization and the wide popularization of automobile products, the negative impact of urbanization has led to frequent traffic accidents. The issue of driving safety has become an inevitable global concern. First, the color and type of the vehicle are fused into the detection algorithm. The multi-task learning framework is used to model the attribute recognition task and location task of motor vehicle. Attribute recognition is completed at the same time of detection. The detection and identification of motor vehicles are divided into two steps: first, coarse-grained attribute identification and detection; second, fine-grained attribute recognition.

In order to propose the vehicle brand recognition model, starting from the convolution neural network structure, the excellent convolution neural network substructures of the Inception structure and the residual learning module are comprehensively utilized. For the recognition task, the recognition accuracy of the proposed method for recognition and type attributes reached 91.1% and 91.8%, respectively. The color and type of the vehicle are important visual features of the motor vehicle. The comprehensive use of the above clues can help improve the detection effect of the motor vehicle. At the same time, good attribute recognition performance can be obtained. In addition, a highly integrated framework is used to accomplish multiple tasks, which can improve the operational efficiency of the intelligent transportation system.

2. Literature Review

As a basic task in intelligent transportation systems, motor vehicle detection and identification has received extensive attention and research [1-2]. Existing solutions usually separate vehicle detection from attribute recognition tasks. Researchers have proposed a number of optimization methods for vehicle detection and attribute recognition.

The basic visual attributes of a motor vehicle include the type of vehicle and the color of the vehicle. Existing attribute recognition methods usually distinguish these two attributes separately. In fact, these two attributes are interrelated.

Different vehicle types have different color distributions. The traditional method has the characteristics of high efficiency and low algorithm complexity. However, different features need to be designed for a specific scenario, and the image data

processed in the data center comes from different intersections. The scene is complex and changeable.

The traditional approach does not meet the requirements for processing all data. After the deep learning is mature, the existing object detection algorithms mainly include region-based detection algorithms and object detection algorithms based on regression [3-4].

Based on the framework of these two general object detection algorithms, the researchers design the vehicle detection algorithm from the following two aspects: On the one hand, combined with the size characteristics of motor vehicles, the vehicle detection performance is optimized. As a rigid object, the profile and size information of a motor vehicle have certain rules. On the other hand, in addition to using a priori information of the size of the vehicle, the resolution of the input image is also increased to 5000*1510. Finally, the detection accuracy of 91.4%, 86.0%, and 70.0% was achieved under the three difficulties of easy, normal, and hard data sets [5].

Based on this idea, a detection model with only three convolution layers was proposed for vehicle detection. Its detection accuracy is low, which is only 62.9% in the proposed data set. The detection rate reached 30 frames per second. However, the lower convolution layer has less semantic information.

Therefore, after completing the vehicle detection, the classification algorithm needs to be further used to identify the attributes of the motor vehicle. These vehicle detection algorithms all use the vehicle as a general category or use only the type information of the vehicle in the vehicle detection algorithm. The differences between the internal categories of the motor vehicle and other objects in the actual environment are not considered, so that the model has insufficient discrimination to the foreground samples, resulting in missed detection and false detection.

The focus of vehicle property identification is mainly on the type, color and brand of the vehicle [6]. There are two main difficulties in vehicle color recognition: one is the reflection on the surface of the vehicle, which affects the effect of color recognition. Second, the color distribution of the vehicle is complicated, and it is difficult to distinguish the main color components of the vehicle body. In order to solve the above problems, combined with the color distribution characteristics of motor vehicles, Kono et al. [7] used traditional color features to identify the color of the vehicle.

The algorithm achieved 92.5% color recognition accuracy on the proposed data set. On this basis, in order to obtain more excellent color features, CNN extraction features are used to replace traditional features for vehicle color recognition [8]. At the same time, the vehicle color is identified by combining the feature context. Finally, the color recognition accuracy rate is 94.7%. The two-way CNN model is used to solve the problem of vehicle color recognition [9]. There are different colors on the body of the same motor vehicle.

Two CNN models are used to model this phenomenon, so that the model can be responsive to different color information. Finally, two CNN model features are stitched together as vehicle color features. In addition, the effect of different color spaces on vehicle color recognition was verified. The RGB color space was used on the proposed data set to achieve 94.5% color recognition accuracy [10].

The difficulty in vehicle type identification is mainly that the vehicle texture information is complex and difficult to characterize with the underlying features. The CNN model is used as a vehicle type identification model. A miniature CNN with only two layers of convolution layer and two layers of fully connected layers was designed. A recognition accuracy of 98% was achieved on the proposed 1500 vehicle image test set. After the reconstruction of the vehicle image is completed, the vehicle features are extracted in conjunction with the three-dimensional description information. Finally, the vehicle brand classification is carried out. The biggest problem with this approach is that the cost is too high, and 3D reconstruction requires a lot of computing resources, which reduces the feasibility of the solution. The above algorithms all use traditional features for vehicle brand recognition. CNN was used to extract the features of the vehicle instead of the artificially designed features [11].
Pre-training is used to solve the over-fitting problem of convolution neural networks on small data sets. The accuracy rate of recognition is 100%. Further, the vehicle image is used as an input to a convolution neural network, and the network structure is used as a classifier for vehicle brand recognition [12]. The above method achieves better recognition accuracy on self-built data sets. However, no sample of brand categories outside the data set was considered. This cannot be used in actual scenarios.

3. Rotating Machinery Joint Detection and Recognition Algorithm

3.1 Joint inspection and attribute identification of rotating machinery

The two coarse-grained properties of the rotating machine color and type are incorporated into the vehicle detection algorithm. Specifically, a region-based object detection algorithm is used as a basic framework of the algorithm. A multi-task learning algorithm is integrated into the framework of the algorithm.

At the same time, the detection of the rotating machine, the identification of the type, and the recognition of the color are completed. Further, in order to alleviate the problems caused by the long tail phenomenon to the model optimization, the

difficult mining algorithm is improved. A multi-task difficult case mining algorithm is proposed.

Firstly, the model optimization target is clarified. After the rotating machinery type and color attributes are combined in the rotating machine detection model, the three tasks of color recognition, type recognition and vehicle positioning are included. The optimization goal is as shown in equation (1):

$$\min_{W} \sum_{i=1}^{3} L_{taski}\left(W_{taski}, I, y^{taski}\right) + f(W) \quad (1)$$

In the equation, $w = \{w_{task1}, w_{task2}, w_{task3}\}$ is the optimization parameter set. $\phi(W)$ is the regularization term. $task_1$, $task_2$ and $task_3$ correspond to the type classification of the vehicle, the color classification of the vehicle, and the detection task of the vehicle. $L_{task_i}(\bullet)$ is the loss function corresponding to each task. $I = \{I_1, ..., I_j, ...\}$ is the training image set. y is the real label of the corresponding task sample. Specifically, the real value of the attribute classification task is a label of the corresponding attribute. For example, the type of vehicle uses a bus, a car, etc. as a label, and the color uses yellow, green, etc. as a label. The cross-entropy loss function is used as the loss function of the classification task, as shown in equation (2):

$$cls(p_k) = \sum_{i=1}^{n} -c_i^* \times \log(c_i) \quad (2)$$

In the equation, t represents the coordinates of the candidate region. t^* is the true coordinates of the vehicle. In order to meet the requirements of the regression algorithm, the coordinate values are normalized to obtain t and t^*. The positioning loss is not calculated when the candidate region p_k is the background. The positioning loss is calculated for this area only when p_k is the foreground.

After determining the optimization goal, the model is used to implement the proposed method. The region of interest (ROI) pooling layer can be divided into the following three modules: feature extraction module, candidate region generation module, and multi-task learning module. The ROI pooling layer is to regularize the features extracted from candidate areas of different sizes. Features of different lengths are mapped to the same dimension vector as the input of multi-task module.

The feature extraction module inputs a three-channel color image I. The network structure is the same as that of VGG-16. The function is to map the image from the pixel domain to the feature domain to obtain the feature map F. The candidate region generation module inputs the feature map F. The module consists of a full convolution network with three convolution layers. The first layer consists of a convolution kernel of size 3*3. The other two layers are in a side-by-side relationship, and they are each composed of a convolution kernel of size 1*1.

The coordinates of the candidate region and the category of the candidate region are respectively output.

Finally, the module outputs a set of candidate regions $p = \{p_1, p_2, ..., p_M\}$ that may be present in the vehicle. The multitasking learning module inputs the features extracted from the candidate regions. After the ROI pooling layer is normalized, the vector set $\{x_{p1}, ..., x_{pM} \mid x_{pt} \in R^{25088}\}$ is obtained.

That is, each candidate region gets a feature vector of 25088 dimensions. Through the multitasking framework, these features are mapped to the vehicle color categories, type categories, and vehicle positioning results for each region.

In the training stage, the input of the multi-task learning module includes the foreground area and the background area. Therefore, on the one hand, the color and type information of the rotating machine will enhance the response of the foreground area in the model. On the other hand, the background area is suppressed, so that the ability of the model to distinguish between the front and back background is improved. Further, since the three tasks share one feature extraction module, the back-propagation capability of the pooling layer of the region of interest is benefited. The model parameters of the module are simultaneously affected by three tasks, so that the model can extract better rotating mechanical features.

In the testing phase, the flow of the algorithm is as follows: For an image I_j, the feature extraction module is used to extract the image features to obtain a feature map F. The generation module of the candidate region obtains all candidate regions $p = \{p_1, p_2, ..., p_M\}$ of the input image using the feature map F. The ROI pooling layer is used to normalize the features, and the feature vector $\{x_{p1}, ..., x_{pM} \mid x_{pt} \in R^{25088}\}$ corresponding to each candidate region is obtained. Finally, the multitasking module is used to map these features to vehicle type, vehicle color, and vehicle detection results. It can be seen that in the testing phase, the proposed method only needs to extract features once for an image. Each module reduces the amount of model calculations through a large number of feature reuse.

In general, the joint detection and recognition algorithm for rotating machinery has the following advantages: In the training phase, the end-to-end method is used to train the model. Detection task, color recognition task and type recognition task share training data. There is no interdependence between detection and classification tasks. In the testing phase, the proposed method takes full advantage of image features.

The multi-task module is used to accomplish the task of attribute recognition and detection

simultaneously, which saves a lot of computing resources.

In different road scenes, there are differences in the number of different types of rotating machines. For example, in the urban area, buses and private cars are the mainstay, while the proportion of trucks on the highway is high. Similarly, the number of rotating machines of different colors is also quite different. This phenomenon can have a negative impact on model optimization when training models using the Stochastic MiniBatch Gradient Descent.

Therefore, on the basis of the proposed model, the online hard example mining (OHEM) algorithm is introduced in the model training stage to improve the negative impact of this phenomenon.

When the batch random gradient descent method is used to train the model, if a total of M regions of interest are generated in one forward operation, the candidate region set is obtained after non-maximum suppression, which is denoted as $P = \{p1, p2, \cdots, p_n \mid N \leqslant M\}$. In the forward process, the loss function for each batch is given by equations (3) and (4):

$$\text{loss}(W, X) = \sum_{i=1}^{n} l(pi) \tag{3}$$

$$l(pi) = \begin{cases} \text{cls}(pi) \\ \text{cls}(pi) + \text{loc}(pi) \end{cases} \tag{4}$$

The first p_i is the background and the second p_i is the foreground. $\text{loss}(\bullet)$ is the loss function for the batch. $\text{cls}(\bullet)$ and $\text{loc}(\bullet)$ correspond to vehicle attribute classification and vehicle positioning loss function, respectively. X is the sample collection for the batch, $X \in P$. n is the number of samples in this batch. The n samples in X are randomly selected from the set X. The candidate regions generated by region generation algorithm are mostly background samples. To prevent an imbalance in the number of foreground and background samples, the ratio of foreground to background sample size is usually controlled at 1:3. This training strategy has the following two problems: First, a background sample will be selected throughout the training process. Second, the samples used to update the model depend on the random sampling algorithm.

In the long tail data, the categories with more samples have a dominant role in model optimization. In order to solve the above problems, the online difficult example mining algorithm only updates model parameters with samples of difficult examples in each iteration, and does not rely on random sampling strategy to determine which samples are used to update parameters. The loss value function of each sample in the single-task model is shown in equation (3). Each sample in set P will get a loss value. The samples in P are sorted according to the value to obtain a sequence $\{l(p_i) \geqslant \cdots \geqslant l(p_j) \geqslant \cdots\}$.

The larger the sample loss value, the less accurate the prediction of the model for the sample. The parameters that need to be updated are adapted to these samples, which are called hard examples. When the batch size is n, only n difficult samples are selected to update the model parameters. It can be seen that the online difficult example mining algorithm reduces the sample distribution to some extent through the method of difficult example mining, such as the influence of the long tail phenomenon on model optimization. Furthermore, the multi-task algorithm is combined with the online mining algorithm. The classification loss function is transformed from single task $\text{cls}(p_i)$ to equation (5):

$$cls(p_i) = cls_{type}(p_i) + cls_{color}(p_i) \tag{5}$$

In the equation, $\text{cls}_*(p_i)$ is the same as formula (2). It can be seen that after the vehicle type and color attributes are integrated, the source of the sample loss value is changed from the original one to two, which can provide more basis for mining difficult samples. By using multi-task on-line hard case mining algorithm, samples used to update model parameters during each iteration are no longer dependent on random sampling strategy. This can not only reduce the influence of the long tail phenomenon on the model optimization, but also make full use of the advantages of multi-task learning to improve the mining effect of difficult example samples.

Ultimately, the vehicle type, vehicle color and vehicle positioning tasks are combined. Moreover, after the multi-task online difficult case mining algorithm is integrated in the training process, the batch random gradient descent method is used to optimize the model. The L2 norm is used as the $\phi(W)$ in the regularization function. The updating formula of model parameters is shown in formula (6) and formula (7):

$$W' = W - \eta * \partial(\text{loss}(W, X) + f(W) / \partial W \tag{6}$$

$$\text{loss}(W, X) = \sum_{i=1}^{n} l(p_i) \tag{7}$$

In the formula, W is the parameter of the model, η is the learning rate, and n is the size of the batch.

The samples in each batch are generated by the candidate region generation module and processed by the multitasking online difficulty instance mining algorithm.

3.2 Data collection for road rotating machinery

In order to verify the performance of the combined detection and recognition algorithm for rotating machinery, a large amount of road image data was collected.

Also, some of the data was selected to construct a road rotating mechanical dataset with 12,712 images and 19,378 rotating machinery. The images in the dataset cover a variety of road conditions and weather conditions, which truly reflect the various scenarios in the actual application. The images in the data set are captured by road surveillance cameras.

The resolution of all images is above 2000*2000.

Each image is labeled with the position of the vehicle, as well as the type and color of the vehicle at the corresponding position. The types of vehicles are classified into six categories according to the size and use of the vehicle. The color of the vehicle is divided into eight types according to the color of the vehicle that is common in daily life.

An example image of each color and model is given. The number of rotating machines of different types and colors in the data set is discussed. It can be seen from the table that the number of rotating machineries shows a long tail phenomenon from the type of the vehicle and the color of the vehicle.

Moreover, the quantity difference between different categories is large, as shown in Table 1.

Table 1 Statistical results of the number of rotating machines of different types and colors

Types	Number of rotating machineries	Color	Number of rotating machineries
Car	8222	Light color	6923
Big truck	4124	Black	4482
Van	2326	Red	2562
SUV	2259	Blue	2091
Bus	1606	Green	1296
Small truck	861	Gray	1230
		Brown	414
		Yellow	400
Total	19398	Total	19398

3.3 Experiment analysis

The experiments involved in this study were all done on the GPU server. The software and hardware environment of the experimental platform are shown in Table 2.

In the division of the data set, 7630 images were randomly selected from the data set as the training set, and the other 5082 images were selected as the test set. When training the joint detection and recognition model of the rotating machine, the weights of the three tasks of vehicle positioning, vehicle color and vehicle type identification are 1:1:1.

In the experiment, both the detection and classification models used the VGG-16 convolution neural network structure.

The complete model structure and parameter names of rotating machinery detection are shown in Table 3. Brand network identification is shown in Figure 1.

Table 2 Configuration of experimental platform

Category	Name	Types	Performance
Hardware	CPU	Intel(R)Xeon(R)CPU	2.1GHz*6
	RAM	NVIDIA K40	1600MHz 64GB
	Graphics processor	Ubuntu 14.04	Single precision computing capability 4.29TFLOPS
Software	System	Ubuntu 14.04	
	Deep learning	Caffe	

Table 3 Parameter configuration of detection model

Module	Layer name/layer type	Core size / step size	Number of parameters
Feature extraction module	Conv1_1/ convolution layer	3*3/1	64*3*3*3
	Conv1_2/ convolution layer	3*3/1	64*3*3*3
	pooll/ convolution layer	2*2/2	
	Conv2_1/ convolution layer	3*3/1	128*3*3*64
	Conv2_2/ convolution layer	3*3/1	128*3*3*64
	pool2/ pooling layer	2*2/2	
	Conv3_1/ convolution layer	3*3/1	256*3*3*128
	Conv3_2/ convolution layer	3*3/1	256*3*3*128
	Conv3_3/ convolution layer	3*3/1	256*3*3*128
	Pool3/ pooling layer	2*2/2	
	Conv4_1/ convolution layer	3*3/1	521*3*3*256
	Conv4_2/ convolution layer	3*3/1	521*3*3*256
	Conv4_3/ convolution layer	3*3/1	521*3*3*256
	Pool4/ pooling layer	2*2/2	
	Conv5_1/ convolution layer	3*3/1	521*3*3*512
	Conv5_2/ convolution layer	3*3/1	521*3*3*512
	Conv5_3/ convolution layer	3*3/1	521*3*3*512
Candidate region generation module	Rpn_conv/ convolution layer	3*3/1	521*3*3*512
	Rpn_cls_score/ convolution layer	1*1/1	36*1*1*512
	Rpn_bbox_pred/ convolution layer	1*1/1	18*1*1*512
Region of interest pool	ROI_Poo15/ pool of interest area	7*7	

Table 4 Test results

Tyes	Index	SSD	Faster-RCNN	Proposed method
Car	AP	88.0	80.7	89.5
	R	92.0	88.8	93.1
Big truck	AP	89.7	90.2	90.2
	R	95.0	95.2	96.4
Van	AP	76.4	85.1	87.3
	R	87.4	86.3	93.9
SUV	AP	72.5	77.7	84.0
	R	85.6	87.8	95.2
Bus	AP	78.9	87.7	88.7
	R	89.2	90.7	94.2
Small truck	AP	51.9	68.8	73.9
	R	60.4	76.2	86.6
	mAP	76.2	80.9	85.6
	mR	72.8	75.0	79.9

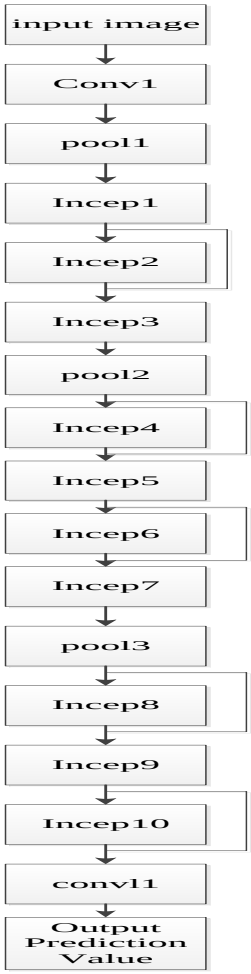

Figure 1 Network structure of brand recognition model

In order to make the experimental results more convincing, the SSD (single shot multi box detector) algorithm and the Faster-RCNN algorithm are improved. The default frame that conforms to the size characteristics of the rotary machine is selected as the parameter of the default frame in SSD and Faster-RCNN.

The optimized algorithm is used as the benchmark algorithm. When evaluating the test performance, the vehicle type is used as the test result.

The detection results of each type of vehicle are detected using the detection accuracy and the recall rate. mAP (mean AP) and mR (mean Recall) are used as evaluation indicators for data set detection results.

The calculation accuracy and recall rate are calculated as follows: Assuming that there are n_{class} vehicles in a certain category, the detection algorithm detects a total of r_{class} vehicles belonging to this category. In r_{class}, a total of t_{class} is correct, and the formula for calculating accuracy and recall rate is as shown in equations (8) and (9):

$$\text{Precision} = \frac{t_{class}}{r_{class}} \tag{8}$$

$$\text{Recall} = \frac{t_{class}}{n_{class}} \tag{9}$$

The final detection accuracy can be obtained by using the 11-point averaging method for the precision value. The experimental results of vehicle testing are shown in Table 4. The detection accuracy reflects the degree of false detection of the model.

The high detection accuracy indicates that the model is more accurate for different types of rotating machinery. The detection accuracy of the SSD and Faster-RCNN algorithms is compared. For the specific task of rotating machinery detection, the region-based detection algorithm is superior to the regression-based detection algorithm in detection accuracy. This indicates that the region-based detection algorithm has stronger modeling ability for rotating machinery and can distinguish different types of vehicles better.

Comparing SSD and Faster-RCNN, it can be seen that the proposed method has the highest detection accuracy. This shows that after adding the type and color information of vehicles, the model's ability to distinguish different types of rotating machinery is improved. For example, in the SSD and Faster-RCNN algorithms, the detection accuracy of minivans and off-road vehicles is low. This means that the two vehicle types are easily confused with other types of vehicles at the time of detection. The proposed method is significantly more accurate than the above methods for the detection accuracy of these two categories of rotating machinery. It shows that after combining the color and category information, the recognition accuracy of the model for the rotating machine is improved.

It can be seen from the calculation formula of the recall rate that the indicator reflects the missed detection of the model. Compared with SSD and Faster-RCNN, the recall rate of this method is significantly higher than other algorithms. This shows that the proposed model has a better

response to rotating machinery. By integrating the type and color information of rotating machinery, a better detection model of rotating machinery was established.

In order to verify the recognition effect of the joint detection algorithm on the vehicle attributes, it is compared with the classification algorithm based on the vehicle image.

The color and type classifier is used to train the vehicle image. From the road image dataset, a total of 12,206 rotating machines were obtained from the 7630 training set images, which were used to train the color and type of vehicle classifiers. The 7192 rotating machines in the test set were used to test the accuracy of the identification. The VGG-16 model is used as the color and type recognition model for vehicle images. First, the end-to-end training model is adopted to complete the end-to-end training. The full connection-layer response in the model is then used as a color or type characteristic of the rotating machine. The dimension of the feature vector is 4096 dimensions and the SVM classifier is trained.

The experimental design of multi-attribute classification is shown in Figure 2. By using the detection model of the rotating machine, the features of the rotating machine are extracted from the image. Similarly, the fully connected layer response is used as a rotating mechanical feature. Therefore, each rotating machine gets a length of 4096-dimensional vector xi. These features constitute a set of feature vectors $\left\{x1, x2, ..., x_k \mid x_{i \in R^{4096}}\right\}$. The SVM classifier is trained using the features extracted from the training set. The trained SVM classifier is then tested on the test set.

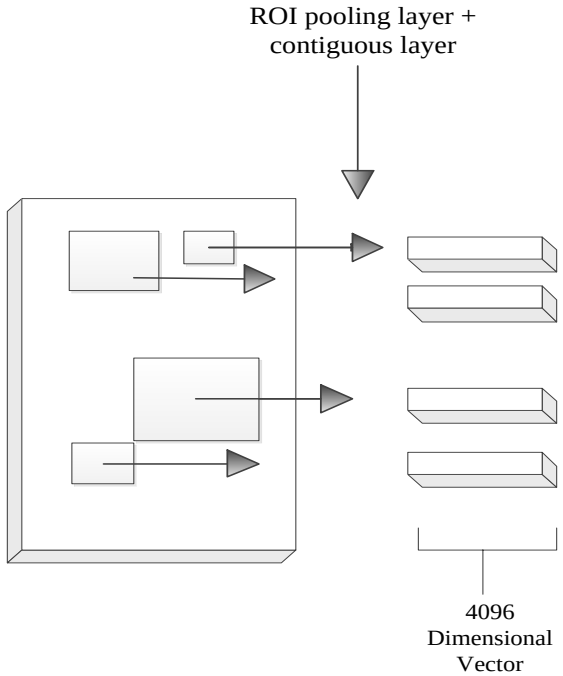

Figure 2: Experimental design of multi-attribute classification

Table 5 Multi-attribute recognition result

Tasks	Vehicle classification (%)	Proposed method (%)
Vehicle color	90.6	91.9
Vehicle type	92.3	91.8

The experimental results of attribute recognition are shown in Table 5. It can be seen from the experimental results that the color and type recognition rate of the joint detection algorithm is comparable to that of the classification model based on the vehicle image. However, since the joint detection and recognition algorithm performs classification simultaneously in the process of detection, it is not necessary to re-extract the features of the rotating machine. Therefore, the efficiency is much higher than the recognition method based on the vehicle image.

In the actual application environment, the algorithm not only needs to have excellent effects, but also must meet the actual needs in terms of computing performance; otherwise, it will not have application value. The detection and recognition algorithm of the rotating machine is input as a scaled image.

The main parameters of the algorithm are shown in Table 6.

Table 6 Parameter setting of the detection algorithm

Parameters	Valie
Number of candidate frames	100
Size of the input images	(500, *)

On the experimental platform, the detection and recognition algorithm processed a total of 50.82 million images in the test set, which required 696 days.

The amount of image data processed in one day can reach 635,000.

According to the amount of image data collected, an average of 3,000 images can be collected on a national road and a provincial road.

Based on this calculation, the algorithm can process image data collected by more than 200 sites per day.

3.4 Vehicle brand recognition model

In combination with the two characteristics of Inception structure, a similar structure is used as the substructure of the network, as shown in Figure 3 and Figure 4.

This substructure consists of three convolution layers.

The first layer consists of the convolution kernel of 1*1, which reduces the input dimension.

The second layer and the third layer are composed of 1*1 and 3*3 convolution kernels, respectively.

On the one hand, it is to make the neurons have different scales of receptive fields; on the other hand, the input is subjected to the ascending dimension operation to extract more information.

This first compressed and expanded structure is similar to the working principle of a sparse auto-encoder.

By limiting the number of hidden layer parameters, the model learns more robust features.

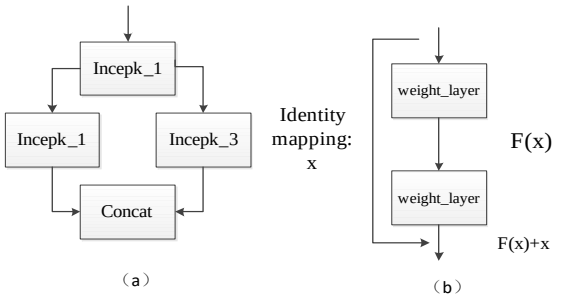

Figure 3: (a) Improved Inception structure; (b) Residual learning module

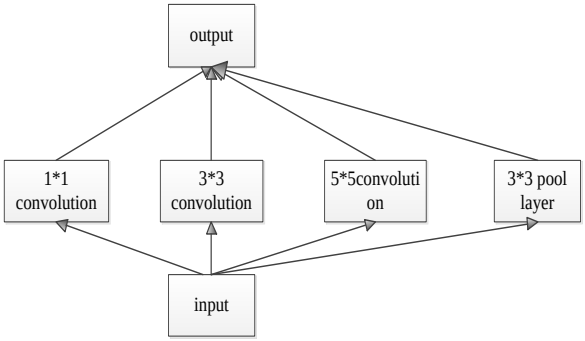

Figure 4: Example diagram of the Inception structure

The topological structure of traditional convolution neural network model decreases with the increase of depth due to problems such as gradient disappearance. In response to this problem, HeK et al. proposed a deep residual learning model in 2015. The performance of the convolution neural network model is optimized by residual learning.

The main principles of the mechanism are as follows: In a recognition task, a CNN model with a depth of L is optimized to obtain a classification accuracy of 85% on the data set. The depth of the model is continued to increase to the L+k layer. From an optimization point of view, the effect of the L+k layer model should not be worse than the model with the depth of the L layer. Because the newly added k-layer model only needs to implement H(x)=1 mapping, the effect of the model is guaranteed to be the same as the original model.

However, from the existing experimental results, simply increasing the depth of the CNN model will reduce the effect of the model. This shows that the existing optimization methods are difficult to simulate the identity mapping relationship when optimizing the CNN model. Therefore, in the CNN model, the identity map is introduced by means of a shortcut, so that the depth of the model is positively correlated with the effect of the model.

Further, the role of identity mapping for the CNN model is theoretically analyzed. The complete residual learning model can be used as in equation (9) and equation (10):

$$y_l = h(x_l) + F(x_{l,} w_l) \tag{10}$$

$$x_{l+1} = f(y_l) \tag{11}$$

In the formula, $x_{l,}$ is the input of the layer 1 residual module. W is the parameter of the residual module. $F(\square)$ is the parameter of the residual module. $f(\square)$ is a function that maps the output of the residual module. If $h(x_l) = x_l$, $x_{l+1} = y_l$, formula (12) can be obtained as follows:

$$x_{l+2} = x_{l+1} + F(x_{l+1}, w_{l+1}) = x_l + F(X_l, W_l) + F(x_{l+1}, w_{l+1}) \tag{12}$$

Further, formula (13) can be obtained:

$$x_L = x_l + \sum_{i=l}^{L-1} F(x_{i,} w_i) \tag{13}$$

In the formula, X_L is an input of any depth. It can be seen that after the introduction of the identity map, there are several excellent properties:

The residuals module at any depth inputs $x_{l=} x_0$. It can consist of a shallow input x1 and a residual $\sum_{i=1}^{L-1} F$ term. It shows that the whole model also has the characteristics of residual learning.

Further, if $x_{l=} x_0$, $x_L = x_0 + \sum_{i=l}^{L-1} F(x_{i,} w_i)$ can be obtained. This indicates that the final output of the deep learning model composed of residuals is obtained by summing. In VGG and Alex Net, the final output of the model is obtained by the product.

The specific form is $\prod_{i=0}^{L-1} w_i x_0$.

The depth model formed by the residual component module is optimized by the gradient descent algorithm, and the gradient is calculated as equation (14):

$$\frac{\partial \varepsilon}{\partial x_l} = \frac{\partial \varepsilon}{\partial x_L} g \frac{\partial x_L}{\partial x_l} = \frac{\partial \varepsilon}{\partial x_L} + \frac{\partial \varepsilon}{\partial x_L} g \frac{\partial}{\partial x_l} \sum_{i=l}^{L-1} F(X_i, W_i) \tag{14}$$

The residual module learning is shown in Figure 5.

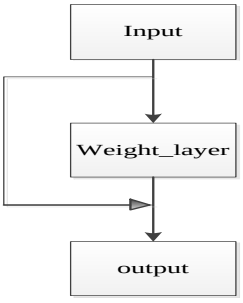

Figure 5: Residual learning module

General object recognition problems, such as type recognition, scene recognition, and test set sample categories are all included in the training set sample category. The model only needs to ensure the prediction accuracy rate for the known category.

As for the vehicle brand identification problem, as mentioned earlier, the category of the sample in the test set is likely to be out of the training set.

Therefore, the requirements for features must be distinguishable in addition to being separable. The gaps within the class are reduced as much as possible, and the distance between classes is increased. It can also be identified when a category sample is present in the test set that is not in the training set.

Compared with separable features, discernible features not only require a greater distance between classes, but also make the gap within classes smaller. Softmax regression is a multi-classification generalization of logistic regression. It is the most commonly used loss function in the deep learning model. Its expression is as shown in equation (15):

$$L_S = -\sum_{i=1}^{m} \log\left(\frac{e^{w_{yi}^T x_i + b_{yi}}}{\sum_{j=1}^{n} e^{w_j^T x_i + b_j}}\right) \quad (15)$$

In the formula, $x_i \in R^d$ is a d-dimensional feature vector. It belongs to the y_i category. $w_j \in R^d$ is the parameter of the model. $b \in R^n$ is the offset term. It can be seen that Softmax regression is the same as logistic regression, and they all belong to the linear classification algorithm. The Softmax loss function is used as a supervisory signal for the model, which can only produce features that are separable, and the requirements for the identification characteristics of the vehicle brand cannot be met. The central loss function not only reflects the separability of the features, but also measures the distance between the features. Its form is as in equation (16):

$$L_c = \frac{1}{2}\sum_{i=1}^{n} \left\| x_i - c_{yi} \right\|_2^2 \quad (16)$$

In the formula, $c_{yi} \in R^d$ is the center of the y_i sample. $\|\bullet\|_2^2$ is the Euclidean distance. n is the number of samples. The loss function calculates the distance of each sample from the center of the category.

The distance within each category feature is reduced.

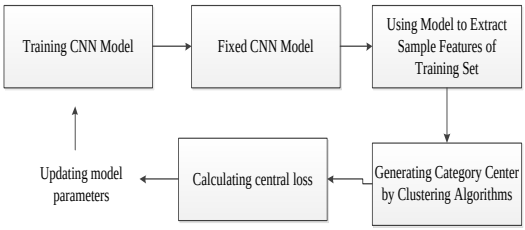

Figure 6: The flow of ideal center loss value calculation

Ideally, the center c_{yi} of each category should be generated from the entire training set sample feature in accordance with the flow shown in Figure 6.

However, this method of calculating center loss is very expensive. Every time the center point is calculated and model parameters are updated, all training set samples need to be completely traversed. Therefore, the algorithm needs to be improved in the following two aspects:

Combined with the batch stochastic gradient descent algorithm, the central loss value in each batch was calculated. The process of the batch gradient descent algorithm is to split the training set into multiple batches (mini-batch).

During each iteration, only the samples in the batch are used to calculate the loss values and the model parameters are updated. Therefore, the form of the central loss function becomes as in equation (17):

$$L_C = \frac{1}{2}\sum_{i=1}^{m} \left\| x_i - c_{yi} \right\|_2^2 \quad (17)$$

In the formula, m is the size of the batch. Center c_{yi} is calculated for all samples in the batch before calculating the center loss. The calculation method is as shown in formula (18) and formula (19):

$$c_j^{t+1} = c_j^t - \alpha \bullet \Delta c_j^t \quad (18)$$

$$\Delta c_j^t = \frac{\sum_{i=1}^{m} \delta(y_{i=j}) \bullet (c_j^t - x_i)}{1 + \sum_{i=1}^{m} \delta(y_{i=j})} \quad (19)$$

In the formula, $\delta(\bullet)$ is an indicative function. If the condition is true, the value is 1; otherwise, the value is 0. As can be seen from the update formula of category center c, although the category center is updated in each batch, the category center value is affected by the entire training set.

The impact of the wrong sample on the model is reduced. In the identification of vehicle brands, it is difficult to guarantee that the labels of training set samples are completely correct because some models are very similar. In order to prevent these samples from affecting each category center, parameter a is used as a smoothing factor when updating the category center to reduce the impact of wrong samples.

After the above improvements, the central loss function can be incorporated into the standard batch gradient descent algorithm for optimizing the convolution neural network model. However, it is not possible to use only the central loss function. It can be seen from the formula that if the loss function is not matched, the loss value of the center loss function will be reduced to zero after multiple iterations. Its physical significance is to gather the sample characteristics of the same category into a single point, resulting in the prediction variance of different samples being 0.

That is, the model after training does not have generalization performance.

Therefore, while using the central loss function, the Softmax regression loss function also needs to be used. Finally, the loss function for vehicle brand recognition is given by equation (20):

$$L = L_c + L_s \tag{20}$$

3.5 Data collection for vehicle brands

In order to verify the proposed vehicle brand classification model, 8559 vehicle images were collected, including cars, trucks, buses and other models. The common thirty-three rotating machinery brands were investigated.

The following phenomena can be observed from the data set:

All vehicles in the dataset are head-to-camera.

However, because the cameras at different intersections face different angles of the road, different vehicles will present different angles.

In addition to the difference in appearance, different brands of vehicles have different coverage. For example, Buick, Volkswagen, Audi, and BMW are mainly private cars in China, and there are basically no buses or trucks. Domestic brands such as Jianghuai and Jiangling, including private cars, trucks, minivans, and buses, cover a wide range of models and have large differences.

The quantity varies greatly among different vehicle brands. In the data set, Volkswagen has the most cars, reaching 1,388. Geely had the fewest cars, which is only 38.

3.6 Experiment analysis

In order to verify the performance of the proposed vehicle brand recognition model, the model is evaluated from two aspects of model recognition accuracy and calculation performance.

The recognition accuracy of the vehicle brand model includes the following two contents:

The first is the recognition accuracy of known brand categories in the training set. In order to verify the effect of the proposed vehicle brand recognition model, Alex Net and Google Net are used as the benchmark network structure. Further, in order to verify the influence of the residual learning and the central loss function on the brand recognition effect, these influencing factors are studied using the control variable method.

The second is the recognition accuracy for brand categories that are not included in the training set.

The purpose of this part of the experiment was to verify the distinguish ability of the brand recognition model for different brand category characteristics. A good classification model can not only accurately predict the known categories, but also accurately judge the sample categories that do not belong to the training set.

The performance evaluation of the model is mainly to evaluate the time consuming of the model prediction. In order to verify the proposed vehicle brand recognition algorithm recognition effect, the Alex Net and Google Net models were used as the benchmark comparison model. Further, in order to study the role of the residual learning algorithm and the central loss function on the vehicle brand, the vehicle brand recognition basic model and its corresponding improvement model are set. They are sequentially recorded as Basenet- Softmax (BS), Basenet- Res- Softmax (BRS), and Basenet- Res- Softmax-Center (BRSC).

It can be seen from the results of this set of experiments that Google Net has the best recognition result, which can reach 97.7%. Compared to Alex Net, it has increased by one percentage point.

This shows that the Inception structure helps to enhance the recognition of the vehicle brand. Compared with Alex Net and the proposed vehicle brand recognition model, the proposed vehicle brand recognition model has the same accuracy. However, Alex Net has 62.4 million parameters.

The proposed vehicle brand recognition model has only 1.5 million parameters, which saves a lot of computing and storage space.

A series of vehicle brand recognition models BS, BRS, BRSC are compared. It can be seen that the residual learning and the central loss function are helpful for improving the recognition effect of vehicle brand. In order to further understand the reasons for the misclassification of some vehicle samples by the model, the misclassified vehicle samples of the three models were collected. There are several reasons for misclassification:

The first is that the sample itself is misclassified. Due to the small differences among some vehicle brands, human error occurred in the data set production. For example, in the third line of rotating machinery, there is no such brand category in the training set, which will inevitably lead to errors.

The second is the existence of occlusion or unclear logo. If the sample is greatly affected by the environment, this will lead to classification failure.

The third is that the difference between different car models under the same vehicle brand is too large. If the samples are all rare models under this brand category, the model is unfamiliar to these samples, resulting in the identification error.

After verifying the effect of the brand recognition model, the computational performance of the proposed vehicle brand recognition model is further evaluated. On the experimental platform, a total of 2152 images were tested in the test set using all models.

The results show that the computational performance of the proposed model is better than that of Google Net and Alex Net.

The processing of a complete vehicle image takes less than 4 milliseconds. It can process 20 million images of the whole vehicle in 24 hours, which fully meets the requirement of processing massive data.

4. Conclusion

Various problems in vehicle identification and detection are studied, including vehicle type identification, vehicle colour identification, vehicle brand identification and vehicle detection.

Unlike current task-oriented solutions, no separate algorithms are designed to address each of the vehicle detection and identification issues. According to the nature and relationship of different tasks, the characteristics of deep learning algorithm are applied.

A new vehicle detection and identification framework is proposed, which firstly identifies and detects the coarse-grained attributes of the vehicle, and then performs fine-grained attribute identification. In the aspect of algorithm, the vehicle coarse-grained recognition and detection algorithm and the vehicle brand recognition algorithm with active learning ability are proposed respectively. In terms of data sets, a road image data set with 12,712 images and a vehicle brand data set with 8559 images were constructed for vehicle detection problems.

In order to solve the problem of motor vehicle detection, the expression ability of the model to the motor vehicle can be proposed by simultaneously detecting the motor vehicle between the vehicle type and colour. Further, the simultaneous integration of the vehicle type and colour in the detection algorithm helps to improve the detection model effect.

The idea is realized by using the framework of multi-task learning and region-based detection algorithm. The detection algorithm has achieved an accuracy of 85.6% in the proposed road image data set, which is better than SSD and Faster-RCNN algorithms. At the same time, the recognition accuracy of the detection algorithm for vehicle colour and vehicle type reached 91.1% and 91.8%, respectively. Aiming at the problem of vehicle brand recognition, the proposed algorithm achieves an accuracy of 96.5% in the data set of known brand categories.

The identification accuracy of the unknown category samples was 97.6%. Moreover, the parameters of the vehicle brand model only contain 1.5 million parameters, and the forward operation time is within 4 milliseconds. In order to put forward the vehicle brand recognition model, based on the structure of convolution neural network, the excellent convolution neural network substructures such as Inception structure and residual learning module are comprehensively utilized. Further, the central loss function is used to enhance the discriminating power of the model features.

Experiments show that the central loss function is very important for the active learning algorithm. Good features are helpful for the model to extract samples of unknown categories.

In summary, the proposed vehicle detection and identification solution solves the problems of vehicle detection, vehicle type identification, vehicle colour recognition, and vehicle brand identification using only two convolution neural network models.

At the same time, all the indicators have reached the performance of the mainstream algorithm, which has practical application value.

It hopes to provide some ideas for other researchers.

Acknowledgement

Research startup subject of Yangtze Normal University: 2017KYQD16

References

[1] Lin, H. Y., Zhao, C. Y., & Zhang, M. J. (2016). Frequency analysis of the non-principal-axis rotation of uniaxial space debris in circular orbit subjected to gravity-gradient torque. Advances in Space Research, 57(5), 1189-1196.

[2] Bi, Q., Huang, N., Chao, S., Wang, Y., Zhu, L., & Han, D. (2016). Identification and compensation of geometric errors of rotary axes on five-axis machine by on-machine measurement. International Journal of Advanced Manufacturing Technology, 84(1-4), 505-512.

[3] Yang, J., Han, D., Zhao, H., & Yan, S. (2016). A generalized online estimation algorithm of multi-axis contouring errors for cnc machine tools with rotary axes. International Journal of Advanced Manufacturing Technology, 84(5-8), 1239-1251.

[4] Lee, K. I., & Yang, S. H. (2016). Compensation of position-independent and position-dependent geometric errors in the rotary axes of five-axis machine tools with a tilting rotary table. International Journal of Advanced Manufacturing Technology, 85(5-8), 1677-1685.

[5] Inamori, T., Otsuki, K., Sugawara, Y., Saisutjarit, P., & Nakasuka, S. (2016). Three-axis attitude control by two-step rotations using only magnetic torquers in a low earth orbit near the magnetic equator. Acta Astronautica, 128, 696-706.

[6] Ibaraki, S., Tsujimoto, S., Yu, N., Sakai, Y., Morimoto, S., & Miyazaki, Y. (2018). A pyramid-shaped machining test to identify rotary axis error motions on five-axis machine tools: software development and a case study. International Journal of Advanced Manufacturing Technology, 94(1-4), 227-237.

[7] Kono, D., Moriya, Y., & Matsubara, A. (2017). Influence of rotary axis on tool-workpiece loop compliance for five-axis machine tools. Precision Engineering, 49, 278-286.

[8] Zhang, J., Li, J., Xie, Z., Chao, D., Liang, G., & Zhao, W. (2017). Rapid dynamics prediction of tool point for bi-rotary head five-axis machine tool. Precision Engineering, 48, 203-215.

[9] Arnold, B., Brinkschmidt, T., Casser, H. R., Diezemann, A., Gralow, I., & Irnich, D., et al. (2017). Multimodale schmerztherapie für die behandlung chronischer schmerzsyndrome. Der Schmerz, 36(05), 361-368.

[10] Wang, Z., Jia, Y., Lei, J., & Duan, J. (2016). Thrust vector control of upper stage with a gimbaled thruster during orbit transfer. Acta Astronautica, 127, 359-366.

[11] Ibaraki, S., & Yu, N. (2017). Formulation of the influence of rotary axis geometric errors on five-axis on-machine optical scanning measurement—application to geometric error calibration by "chase-the-ball" test. International Journal of Advanced Manufacturing Technology (2), 1-11.

[12] Xu, R., Xiang, C., Zheng, G., & Chen, Z. (2017). A tool orientation smoothing method based on machine rotary axes for five-axis machining with ball end cutters. International Journal of Advanced Manufacturing Technology, 92(9-12), 3615-3625.