

基于深度卷积神经网络的车型识别

石 磊¹ 王亚敏¹ 曹仰杰² 卫 琳²

(郑州大学信息工程学院 郑州 450001)¹ (郑州大学软件与应用科技学院 郑州 450002)²

摘 要 传统的基于卷积神经网络的车型识别算法存在识别相似车型的准确率不高,以及在网络训练时只能使用图像的灰度图从而丢失了图像的颜色信息等缺陷。对此,提出一种基于深度卷积神经网络(Deep Convolution Neural Network,DCNN)的提取图像特征的方法,运用深度卷积神经网络对背景较复杂的车型进行网络训练,以达到识别车型的目的。文中采用先进的深度学习框架 Caffe,基于 AlexNet 结构提出了深度卷积神经网络的模型,分别对车型的图像进行训练,并与传统 CNN 算法进行比较。实验结果显示,DCNN 网络模型的准确率达到 96.9%,比其他算法的准确率更高。

关键词 卷积神经网络,车型识别,深度学习

中图分类号 TP183 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2018.05.048

Vehicle Type Recognition Based on Deep Convolution Neural Network

SHI Lei¹ WANG Ya-min¹ CAO Yang-jie² WEI Lin²

(School of Information Engineering,Zhengzhou University,Zhengzhou 450001,China)¹

(School of Software and Applied Science and Technology,Zhengzhou University,Zhengzhou 450002,China)²

Abstract The accuracy of traditional convolution neural network recognizing the vehicle model is not high when recognizing similar models,and the gray scale of image can only be used in the network training with the loss of color information of the image. Based on this,a method of extracting image features based on deep convolution neural network (DCNN) was proposed. The deep convolution neural network is used to carry out network training for the vehicle model with complex background,so as to achieve the purpose of recognizing models. In this paper,by using advanced deep learning framework Caffe,a deep convolution neural network based on AlexNet structure was proposed with the training of the image of vehicle model and the comparison with traditional convolution neural network. The experimental results show that the accuracy rate of DCNN network model can reach 96.9% with a higher accuracy.

Keywords Convolution neural network,Vehicle identification,Deep learning

1 引言

21 世纪以来,随着生活水平的不断提高,车辆已经成为了普遍交通工具,其数量飞速增长。不断增加的车辆虽然给人们的生活提供了便利,但是也给交通管理带来了巨大的挑战,由交通事故导致的死亡人数比例一直居世界首位。

传统的依靠人力或者交通设施的管理方法已经不能满足目前的发展需求。基于此,人们提出了一种能够发挥实时性、高效性、准确性的交通管理系统——智能交通系统,它提高了车辆的运输效率,缓解了交通的拥挤压力,保障了交通的安全,减少了对环境的污染,降低了对能源的消耗。而车型识别

是智能交通管理系统的重要组成部分,目前比较成熟的车型识别方法有超声波检测^[1]、激光红外线检测^[2]等,这几种方法在设备安装的过程中容易被破坏,使交通无法正常运行,尽管识别效率比较高,但是需要很高的维护成本。

目前,深度学习算法已经在图像分类^[3]以及识别目标物体^[4]等方面取得了很好的效果。与传统的依赖先验知识进行特征提取的算法相比,深度神经网络具有更好的适用性和敏捷性,可以使用训练数据来适应不同的网络构建特征。而深度学习中一种重要的结构是卷积神经网络,该网络具有在深度学习中自动学习并提取图像特征的特点,并且已经在图像识别、搜索引擎等领域获得了很大的发展。其利用局部感受

到稿日期:2017-05-10 返修日期:2017-07-07 本文受国家自然科学基金(U1304603),河南省高等学校重点科研项目(17A520016),郑州大学优秀青年教师发展基金(1521337044)资助。

石 磊(1967—),男,博士,教授,硕士生导师,主要研究方向为高性能计算、Web 数据挖掘,E-mail:ielshi@zzu.edu.cn(通信作者);王亚敏(1990—),女,硕士生,主要研究方向为人工智能、大数据处理,E-mail:1245593616@qq.com;曹仰杰(1976—),男,博士,副教授,硕士生导师,主要研究方向为高性能计算、大数据处理;卫 琳(1968—),女,博士,副教授,主要研究方向为 Web 数据挖掘。

野和权值共享等方法增强了网络的泛化能力和鲁棒性,从而降低了网络结构的复杂度并减少了权值数量。

对于车型识别,国内外的研究人员尝试了各种方法。早期的方法都是依赖人工设计特征来进行识别的,如 SIFT^[5], HOG^[6]等,这些特征提取的方法不仅运算量比较大,而且在实践中还需要对输入的原始图像进行预处理。而卷积神经网络的输入可以是原始图像,能够使机器自主地学习图像特征,从而避免了人工提取特征带来的复杂性,提高了运算效率。

拍摄的图像在进行参数缩放、旋转平移等外观变化时,会受到车辆外观的复杂表示形式以及摄像机拍摄的角度、距离和环境光照等因素的影响,识别难度较大。而卷积神经网络对图像的平移、比例缩放以及倾斜等其他视觉形变具有高度的不变性,能够有效地克服车辆外观多变带来的影响。

单通道的 CNN 使用灰度图以及 RGB 通道的某一种通道,在学习训练的过程中不能有效地利用颜色信息;传统的 CNN 只提取图片的局部特征,从而丢失了图像的一些重要信息。因此,提取图像的局部特征和全局特征在图像的识别中都具有同等重要的作用。本文提出了一种深度卷积神经网络结构,把最后得到的特征输入到 SVM 分类器中进行识别,在该结构中,通过对输入的多种图片进行特征提取与组合,能够提取更多的图片信息,提高了图像识别的准确率。

2 深度卷积神经网络

2.1 总体流程

本文所构建的深度卷积神经网络结构包括网络的训练以及测试两部分,具体流程如图 1 所示。网络的训练图像有两种:原始图像和经过梯度处理的图像。将两种图像输入相同的网络结构进行学习训练,不断地进行迭代并反馈结果,同时调节网络的参数,使整个网络朝着最优的方向发展。

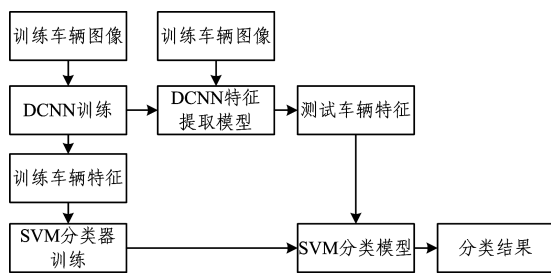


图 1 深度卷积神经网络的总体流程图

Fig. 1 Overall flow chart of deep convolution neural network

2.2 深度卷积神经网络的结构

本文基于深度学习框架 Caffe 中的 AlexNet 结构提出了深度卷积神经网络的模型,并把最终特征输入到 SVM^[7]分类器中以完成对目标的识别。

2.2.1 AlexNet 模型

AlexNet 是深度学习框架中一个重要的模型,共包含 8 层结构,不包含输入层,Layer1—Layer5 为卷积层,Layer6—Layer8 为全连接层,最后一层也是全连接层,网络参数的具体设置如表 1 所列。

表 1 AlexNet 网络的参数设置
Table 1 Parameter setting of AlexNet network

	Type	Filter Size	Stride	padding	OutputSize
InputData	Color image	—	—	—	224 * 224 * 3
Conv1	—	7 * 7	4	0	55 * 55 * 96
Pool1	Max pooling	3 * 3	2	0	27 * 27 * 96
Conv2	—	5 * 5	1	2	27 * 27 * 256
Pool2	Max pooling	3 * 3	2	0	13 * 13 * 256
Conv3	—	3 * 3	1	1	13 * 13 * 384
Conv4	—	3 * 3	1	1	13 * 13 * 384
Conv5	—	3 * 3	1	1	13 * 13 * 384
Pool5	Max pooling	3 * 3	2	0	6 * 6 * 256
Fc6	—	—	—	—	4096
Fc7	—	—	—	—	4096
Fc8	—	—	—	—	1000

Softmax with loss

2.2.2 深度卷积神经网络模型

本文算法是基于 AlexNet 算法提出的深度卷积神经网络,其结构如图 2 所示。

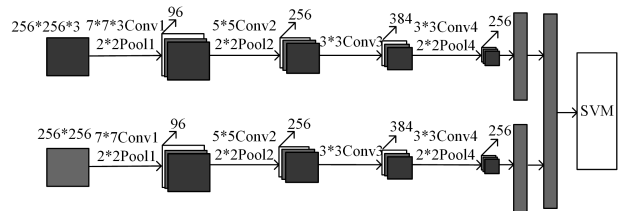


图 2 深度卷积神经网络的结构

Fig. 2 Structure of deep convolution neural network

虽然传统的 CNN 对平移、比例缩放、倾斜等其他视觉形变具有高度的不变性,但是如果输入图像的尺寸发生微小的改变,则仍会使网络识别的最终结果发生变化^[9]。为了增强 DCNN 模型的鲁棒性,将输入的训练图像分为两种:原始图像和经过梯度处理的图像。将两种图像输入相同的网络结构进行学习训练,把得到的特征维度进行融合,最后将融合的特征输入 SVM 分类器中进行分类识别。

当输入原始图像的网络结构时,由于原始图像是彩色的,因此输入图片是三通道的,Conv1 卷积层的卷积核也是三通道的。将卷积核的不同通道分别与输入图像对应的通道进行卷积操作,在以大小为 256 * 256 * 3 的图像作为输入的网络结构中,第一个卷积层(Conv1)使用了 96 个大小为 7 * 7 * 3 的卷积核,并且在步长 Stride 为 4 的情况下对 256 * 256 * 3 的图像进行了卷积,也即使用了 7 * 7 的卷积核在 3 个通道上进行了卷积操作,对图像在间隔为 4 个像素的采样上进行了卷积操作。得到图像的基本特征后,进行一次 ReLU 规则和 Norml 变换后通过下采样层 Pool1 采样后输出,下一层网络的输入即为这一层网络输出的特征图。输出特征图的大小为 27 * 27,数量为 96 个。

其中,多通道的卷积结构与单通道的卷积结构不同,图 3 和图 4 分别给出单通道和多通道的卷积层结构。

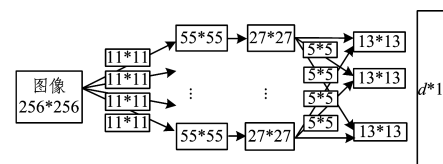


图 3 单通道的卷积层结构

Fig. 3 Convolution layer structure of singel channel

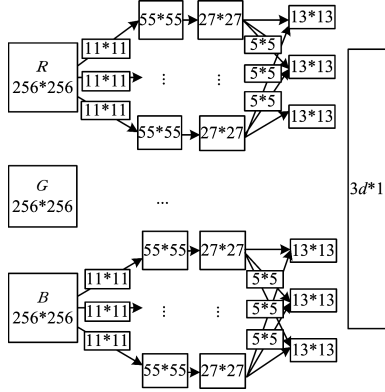


图 4 多通道的卷积层结构

Fig. 4 Multi-channel convolution layer structure

1) 单通道上多个卷积核的卷积过程

通过使用一个卷积核提取特征得到的特征图是不充分的,例如使用 96 个特征图便可以学习 96 种特征,那么输出的结果则为 96 个特征图。

2) 多通道上多个卷积核的卷积过程

以图 5 为例,其描述了在 3 个通道上进行的卷积操作,因有两个卷积核,故生成了两个特征图。需要注意的是,3 个通道上的每一个通道都对应一个卷积核,例如在 w_0 的某个位置 (i, j) 处的值,是通过将 3 个通道上的 (i, j) 处的卷积结果相累加后再取激活函数值得到的,因此最后得到了两个特征图。如果输入 96 个卷积核,那么输出的特征图也为 96 个。

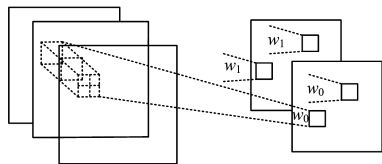


图 5 在 3 个通道上的卷积操作

Fig. 5 Convolution on three channels

在第二个卷积层 Conv2 中,使用了 256 个大小为 5×5 的卷积核,提取的特征图经过下采样 Pool2 层后,得到了 256 个大小为 13×13 的特征图。

在第三个卷积层 Conv3 中,使用了 384 个大小为 3×3 的卷积核,得到了 384 个大小为 13×13 的特征图。

在第四个卷积层 Conv4 中,使用了 256 个大小为 3×3 的卷积核,经过下采样层 Pool4 后,得到的图大小为 $\text{int}(13/2) = 6$,从而得到了 256 个 6×6 的特征图。

全连接层 Fc5 是上一层 Conv4 进行池化后的全连接层,得到的特征维度为 4096。

将梯度处理的图像输入到网络中,第一层使用 96 个 7×7 的卷积核进行卷积操作,其余的都与原始图像作为输入的结构参数相同。

最后将两个网络结构中得到的特征图进行合并,从而得到最终的图像特征维度;将所得到的该特征直接输入到 SVM 分类器中即可识别目标图像。

2.2.3 优化网络参数

在训练样本数目少并且神经网络模型^[10]相对比较复杂

的情况下进行网络训练时,可能会出现过拟合现象。随着训练样本的迭代次数的增加,网络在训练集上的错误率逐渐减小,而在测试集上的错误率却逐渐增大,这是过拟合现象的主要表现。出现过拟合现象的主要原因是:网络对训练数据集进行了过度拟合,但是对除了训练集以外的数据却不能进行很好的拟合。图 6 直观地描述了网络过拟合的现象。

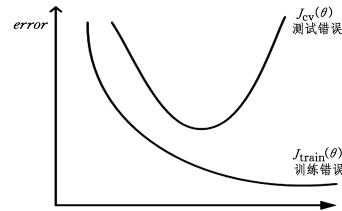


图 6 过拟合现象

Fig. 6 Over-fitting phenomenon

实际的深度网络训练过程中普遍存在过拟合的现象,因此提升网络性能最主要的步骤便是减少网络训练时出现的过拟合现象。本文通过优化网络的参数来降低网络过拟合的程度。

1) batch 值的设定

进行网络训练时,一般根据最大迭代次数或者损失函数的阈值来决定网络停止训练的时间,而在实际的网络训练中,每一次迭代都是随机把训练样本依次分成同样大小的 batch 数据集,再同时把每一个 batch 数据集输入到网络中进行训练学习,直到所有的 batch 数据集都训练完,才能进行下一次的迭代。本文选取不同的 batch 数据集进行网络训练,当网络损失函数的值低于预定的阈值时,停止训练。测试结果取 5 次测试的平均值。

2) 局部响应归一化

局部响应归一化(Local Response Normlization)主要是在卷积神经的网络层对输入的局部区域进行归一化操作,是一种对神经元进行横向抑制的方法。算法的计算公式为:

$$b'_{x,y} = a'_{x,y} / (k + \alpha \sum_{j=\max(0,i-n/2)}^{\min(N-1,i+n/2)} (a'_{x,y})^2)^\beta \tag{1}$$

其中, $a'_{x,y}$ 表示第 i 个卷积核在位置 (x, y) 的输出; $b'_{x,y}$ 表示局部响应归一化后的输出; N 表示卷积层中卷积核的数量; k, α, n, β 是常量,通常将 k 设置为 2, α 设置为 10^{-4} , n 设置为 5。

3) Dropout 的优化

Dropout 的提出降低了网络过拟合的程度,有效地避免了因训练样本不足而导致的网络过拟合的现象。Dropout 层随机选择节点进行训练学习,说明了每次参与训练学习的节点只是网络结构中的一部分,其中大多数的参数是共享的,这等同于使用 Dropout 的功能训练出了多个实用性的模型。例如,在网络的训练过程中将 Dropout 的参数值设置为 0.5,则以 50% 的概率将每一个隐含层的 output 设置为 0,使其对 forward 以及 backward 过程不起任何作用;由于每次训练输入的训练样本对应不同的网络结构,但这些不同的网络结构的权值是共享的,因此求得的参数可以适应不同情况下的网

络结构,从而在一定程度上抑制了网络分类的错误率,改善了网络的性能。

例如,要训练如图 7(a)所示的网络,在开始训练时,随机地选择一些隐含神经元进行删除,那么获得的图 7(b)则保持输入层的数据不变,遵从反向传播的原则开始更新网络中的权值参数。进行多次迭代训练学习,且每一次都按照这种方式来删除隐含层的一些隐含单元,直到网络训练结束。Dropout 防止过拟合的方法主要是通过训练网络中的一部分神经节点来实现,因为训练时间增长,所以很多用于训练学习的神经网络都能得出一个比较准确的分类结果,从而使训练的误差较低,直到满足条件时即可停止对网络的训练。

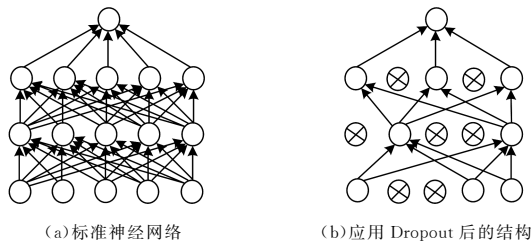


图 7 标准神经网络的结构及对其应用 Dropout 后的结构
Fig.7 Structure of standard neural network and structure after using Dropout

2.2.4 深度卷积神经网络的训练过程

深度卷积神经网络的训练过程与传统的卷积神经网络过程相似,不同点在于全连接层使用了逐层训练的方法。DC-NN 的训练算法的具体流程如下所示。

- 1)采用高斯分布的方式对网络的权值进行初始化,将其中的偏置值设为常量。
- 2)将训练数据集分为两种:①经过归一化处理的 256×256 的彩色图像;②经过梯度处理的 256×256 的灰度图像。
- 3)从训练集中任意抽取一个 batch 块作为训练的样本,并将其作为输入的图片。
- 4)将训练样本输入到网络中进行前向传播,并经过每层的网络计算得到输出的结果。
- 5)计算出网络的实际输出与网络预测值之间的误差,如果误差小于预定的阈值,则停止对网络的训练,否则继续对网络进行训练。
- 6)按极大值的方式对误差进行网络的反向传播,并且逐次进行网络权值的更新。
- 7)重新从 2)开始,继续对网络进行训练。

DCNN 的具体测试流程如下所示:

- 1)采用已经训练好的参数对网络的权值进行初始化。
- 2)从测试集中任意选取一个测试样本,并将其输入到网络中。
- 3)将测试样本输入到网络中进行前向传播,并经过每层的网络计算得到输出的结果。
- 4)将网络的输出结果与已经标记的测试样本进行对比,通过判断输出的结果是否正确,统计出分类识别的结果。
- 5)重新从 2)开始,直到测试完所有的样本为止。

3 实验与结果分析

3.1 测试数据

基于从高速公路监控上截取的车辆图像数据集进行测试实验,将图像分为两种类型:1) 256×256 的彩色图;2) 经过梯度处理的大小为 256×256 的灰度图。该数据集中包括 5 种类型的车辆,共计 8000 幅图像,实验中任意选取车辆图像集中 80% 的图像作为训练样本,剩下的 20% 作为测试样本。本文主要在 Ubuntu+Caffe+MATLAB 环境下来进行实验。

3.2 实验设定

将原始图像作为输入的网络结构称为网络 1,将经过梯度处理的图像作为输入的网络结构称为网络 2,分别对这两个网络结构进行训练,并比较这两个网络结构的识别率;然后融合两个网络得到的特征后进行识别;最后比较网络的识别率。

3.3 实验结果

首先,在数据上采用分类算法对一些常见的特征进行分类,其正确率如表 2 所列。

表 2 采用不同分类算法对常见特征进行分类的结果
Table 2 Classification results of common features by classification algorithm

分类算法	LDA ^[11] +NN	SVM ^[12]	Random forests ^[13]
Color	0.929	0.824	0.931
LBP ^[14]	0.781	0.773	0.817
SIFT ^[15]	—	—	0.802
HOG ^[16]	—	—	0.574

从实验结果可以看出,相比于其他特征,基于颜色特征进行识别时更有效,而传统的 CNN 输入的是灰度图,DCNN 结构输入的都是三维的彩色图,因此 DCNN 结构的识别精确率更高。

本文算法与其他实验算法的比较情况如表 3 所列。CNN 算法直接利用图像进行卷积操作来提取图像的特征;而本文算法则是使用不同的图像作为输入,然后把得到的特征融合,因此特征更准确,识别率更高。

表 3 本文方法与其他实验方法的对比
Table 3 Comparison between our method and others

方法	识别率/%
CNN ^[17]	92.5
DCNN ^[18]	93.4
MSCNN	96.9

不同网络与网络组合的识别率对比如表 4 所列。

表 4 不同通路情况下的识别率对比
Table 4 Comparison of recognition rates in different paths

网络	识别率/%
网络 1	93.5
网络 2	91.4
网络 1+网络 2	96.9

对于车辆数据集而言,网络 1 的识别率最高,因为网络 1 输入的是原始的彩色图像,提取的特征信息更多。网络 2 的

识别率有所下降,因为该网络输入的是经过梯度处理的灰度图,提取特征的过程中会丢失一些重要的信息。网络1与网络2结合后的识别率最高,因此DCNN提取的特征具有的泛化能力和鲁棒性更强。

结束语 本文对多种车型识别方法进行了研究与探索,并针对传统的深度卷积神经网络对低分辨的图像特征提取不明显的问题,利用深度学习的框架提出了一种基于深度卷积神经网络的车型识别模型。该模型结合输入图像的尺寸、复杂性等特点对深度学习框架进行了改进;通过输入不同类型的图片来增加提取特征的信息,并将得到的特征融合后输入到SVM分类器中进行识别。本文提出的模型对输入图像的识别率较高,更加有效地提高了图像的识别率以及鲁棒性,因此本文方法的结构学习能力比较强,具有一定的应用价值。

下一步将探索和设计卷积神经网络的其他结构,更进一步利用其自学能力,并通过提取更丰富的特征信息来提高分类识别的精确率。

参考文献

- [1] TAKAMOTO M, ISHIKAWA H, SHIMIZU K, et al. New measurement method for very low liquid flow rates using ultrasound[J]. *Flow Measurement & Instrumentation*, 2001, 12(4): 267-273.
- [2] WARRIACH E U, CLAUDEL C. Poster abstract: a machine learning approach for vehicle classification using passive infrared and ultrasonic sensors[C]// *International Conference on Information Processing in Sensor Networks*. 2013: 333-334.
- [3] TANG S J, WU H W. A Robust Face Recognition Method Based on Snake Model and Collaborative Representation Classification[J]. *Natural Science Journal of Xiangtan University*, 2016, 38(2): 104-108. (in Chinese)
唐守军, 吴洪武. 基于Snake模型和协作表示分类的鲁棒人脸识别方法[J]. *湘潭大学学报*, 2016, 38(2): 104-108.
- [4] LIU R, LIU J Q, LIU H. Gesture feature extraction and recognition based on acceleration track[J]. *Application Research of Computers*, 2017, 34(3): 924-927. (in Chinese)
刘蓉, 刘家祺, 刘红. 基于加速度轨迹图像的手势特征提取与识别[J]. *计算机应用研究*, 2017, 34(3): 924-927.
- [5] HUA L Q, XU W, WANG T, et al. Vehicle Recognition Using Improved SIFT and Multi-View Model [J]. *Journal of Xi'an Jiaotong University*, 2013, 47(4): 92-99. (in Chinese)
华莉琴, 许维, 王拓, 等. 采用改进的尺度不变特征转换及多视角模型对车型识别[J]. *西安交通大学学报*, 2013, 47(4): 92-99.
- [6] TAIGMAN Y, YANG M, RANZATO M, et al. DeepFace: Closing the Gap to Human-Level Performance in Face Verification [C]// *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2014: 1701-1708.
- [7] RAHATI S, MORAVEJIAN R, KAIEMI E M, et al. Vehicle recognition using contourlet transform and SVM [C]// *5th International Conference on Information Technology: New Generations*. IEEE, 2008: 894-898.
- [8] ZHANG J. Research on Image Retrieval Based on Fusion Feature of AlexNet[D]. Chongqing: Chongqing University of Posts and Telecommunications, 2016. (in Chinese)
张俊. 基于AlexNet融合特征的图像检索研究[D]. 重庆: 重庆邮电大学, 2016.
- [9] GOODFELLOW I J, WARDE-FARLEY D, MIRZA M, et al. Maxout networks [J]. *Computer Science*, 2013, 28(3): 1319-1327.
- [10] YAN F, WANG L Z, SHI L, et al. Digital Image Watermarking Algorithm Based on DWT and PNN[J]. *Natural Science Journal of Xiangtan University*, 2016, 38(3): 89-93. (in Chinese)
闫凤, 王立中, 石磊, 等. 基于DWT和PNN的数字图像水印算法[J]. *湘潭大学自然科学学报*, 2016, 38(3): 89-93.
- [11] ANDREW A M. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods [J]. *Kybernetes*, 2001, 32(1): 1-28.
- [12] CHEN L F, CHEN L, LIU B X. Research of SVM Multiclass Model Based on Granular Computing & Huffman Tree[J]. *Computer Science*, 2016, 43(1): 64-68. (in Chinese)
陈丽芳, 陈亮, 刘保相. 基于粒计算的哈夫曼树SVM多分类模型研究[J]. *计算机科学*, 2016, 43(1): 64-68.
- [13] ZHAO H X, XU F, CHEN J Y. Face Recognition Based on Multi-Scale LBP [J]. *Computer Applications and Software*, 2012, 29(1): 257-259. (in Chinese)
赵怀勋, 徐锋, 陈家勇. 基于多尺度LBP的人脸识别[J]. *计算机应用与软件*, 2012, 29(1): 257-259.
- [14] HUANG F C, HUANG S Y, KER J W, et al. High-performance SIFT hardware accelerator for real-time image feature extraction [J]. *IEEE Transaction on Circuits and Systems for Video Technology*, 2012, 22(3): 340-351.
- [15] DALAL N, TRIGGS B. Histograms of Oriented Gradients for Human Detection[C]// *IEEE Computer Society Conference on Computer Vision & Pattern Recognition*. IEEE Computer Society, 2005: 886-893.
- [16] DENG L. Deep Convolutional Neural Networks for Vehicle Classification [D]. Chengdu: Southwest Jiaotong University, 2015. (in Chinese)
邓柳. 基于深度卷积神经网络的车型识别[D]. 成都: 西南交通大学, 2015.
- [17] PI S Y, TANG H, XIAO N F. Fully Convolutional Deep Learning Model Based Graspable Object Etection [J]. *Journal of Chongqing University of Technology (Natural Science)*, 2018, 32(2): 166-173. (in Chinese)
皮思远, 唐洪, 肖南峰. 基于全卷积深度学习模型的可抓取物品识别[J]. *重庆理工大学学报(自然科学)*, 2018, 32(2): 166-173.