

洪顺贺, 铁治欣, 胡宸滔, 等. 一种改进多标签分类结构的实时车牌识别模型[J]. 智能计算机与应用, 2024, 14(6): 207-212.
DOI: 10.20169/j.issn.2095-2163.240631

一种改进多标签分类结构的实时车牌识别模型

洪顺贺¹, 铁治欣^{1,2}, 胡宸滔³, 丁成富⁴

(1 浙江理工大学 计算机科学与技术学院, 杭州 310018; 2 浙江理工大学 科技与艺术学院, 浙江 绍兴 312369;
3 浙江理工大学 信息科学与工程学院, 杭州 310018; 4 聚光科技(杭州)股份有限公司, 杭州 310052)

摘要: 本文提出一种改进的多标签分类结构的实时车牌识别模型, 通过压缩-激活注意力机制在通道方向关注车牌图像中的字符, 并且融合网络的深浅层特征, 在提取语义的同时又能兼顾到字符定位, 提高字符定位精度, 同时采用深度可分离卷积取代传统卷积, 减少计算量, 提高车牌识别速度。最后, 在 CCPD 数据集中选取最能体现光照强度、车牌倾斜和车牌模糊等复杂场景的 3 个子数据集进行测试实验, 测试结果达到 94.0% 的平均准确率和 395.6 FPS 的识别速度, 这表明本文提出的模型能够在复杂场景下达到快速识别车牌的效果。

关键词: 车牌识别; 深度学习; SE; DSC

中图分类号: TP391.41

文献标志码: A

文章编号: 2095-2163(2024)06-0207-06

An improved multi-label classification structure for real-time license plate recognition model

HONG Shunhe¹, TIE Zhixin^{1,2}, HU Chentao³, DING Chengfu⁴

(1 School of Computer Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China; 2 Keyi College, Zhejiang Sci-Tech University, Shaoxing 312369, Zhejiang, China; 3 School of Information Science and Technology, Zhejiang Sci-Tech University, Hangzhou 310018, China; 4 Focused Photonics (Hangzhou) Inc., Hangzhou 310052, China)

Abstract: This paper proposes an improved multi-label classification structure for real-time license plate recognition model designed using deep learning technology. The model utilizes the Squeeze and Excitation attention mechanism to focus on license plate characters in the channel direction and combines features from the network's deep and shallow layers. This approach realizes character localization and semantic feature extraction, improving character positioning accuracy. Additionally, the model adopts Depthwise Separable Convolution in place of traditional convolution to reduce computational complexity and improve license plate recognition speed. Finally, the proposed model is tested on three subsets of the CCPD dataset that demonstrate complex scenarios, such as varying illumination, plate inclination, and blurriness. The test results show an average accuracy of 94.0% and a recognition speed of 395.6 FPS, indicating the proposed model's effectiveness in quickly identifying license plates in complex environments.

Key words: license plate recognition; deep learning; SE; DSC

0 引言

近年来,人们生活水平显著提高,全国机动车保有量也随之快速增长,与车辆相关的交通拥堵问题、闯红灯现象也有所增加。一个智能化的交通管理系统可以帮助交警部门实现自动化交通管理,从而减少交通拥堵和交通事故的发生。车牌识别技术是智

能交通管理系统的重要组成部分,通过道路上的摄像头,对车辆进行实时监控和识别,一旦发现车辆超速或者闯红灯,就可以根据车牌号获得相关责任人信息,从而促进警方及时处理违法行为,优化道路交通运行,是实现智能化交通管理的核心技术。

传统的车牌识别方法主要分为车牌字符分割和字符识别两个步骤。武云飞^[1]采用形态学算法对

基金项目: 浙江省公益技术研究项目(2015C31024); 教育部产学研合作协同育人项目(220502645275342)。

作者简介: 洪顺贺(1996-),男,硕士研究生,主要研究方向: 车牌识别; 胡宸滔(2000-),男,硕士研究生,主要研究方向: 图像与视频处理技术。

通讯作者: 铁治欣(1972-),男,博士,副教授,硕士生导师,主要研究方向: 数据挖掘,嵌入式系统,电力系统自动化等。Email: tiezx@zstu.edu.cn

收稿日期: 2023-04-12

车牌中的字符进行分割,再使用模板匹配进行车牌字符识别,能够在 1 088 张图片中达到较高的识别正确率。吕文强等学者^[2]提出一个基于车牌最大相邻字符间隙的字符分割方法对车牌上的字符进行分割,能够在字符粘连等环境中识别车牌。刘丽丽^[3]提出了模板匹配联合特征多种算法相结合的综合应用方法,有效解决了相似字符识别容易出错的问题。但是传统方法非常容易受到车牌图像中光照不均、噪声等因素的影响,导致部分车牌字符识别出现错误,深度学习技术的出现为复杂场景中的车牌图片识别任务提供新的途径。张彩珍等学者^[4]使用双向长短期记忆网络取代卷积循环神经网络结构中的传统循环神经网络对车牌进行识别,不用分割车牌字符就能识别车牌号码,有效提高车牌识别平均准确率。高昕葳^[5]使用 Faster-RCNN 网络进行车牌定位,生成车牌提取框提取车牌,并且使用 VGG16 网络模型识别车牌字符,能够在复杂环境下对车牌进行准确识别。

本文采用深度学习研究方法研究复杂场景下的车牌识别问题,提出一种改进的多标签分类结构的实时车牌识别模型(Improved Multi-label Classification Structure for Real-time License Plate Recognition

Model, IMCRLPR),将压缩-激活(Squeeze and Excitation, SE)注意力机制和深度可分离卷积(Depthwise Separable Convolution, DSC)融合到车牌识别模型中,能够达到在复杂场景下快速、准确识别车牌的效果。

1 模型结构

本文提出 IMCRLPR 车牌识别模型,在车牌字符定位模块设计了一个轻量化通道定位网络(Lightweight Channel Positioning Network, LCPN),融合 SE 注意力从通道方向提取到车牌字符位置的权重信息和深度可分离卷积(DSC)提取到的特征信息,有效提高车牌识别的精度。IMCRLPR 主要由车牌特征提取、车牌字符定位、车牌字符识别三个模块组成,其总体框架如图 1 所示。先将输入的车牌图片进行预处理,将图像的大小调整为 30×120×3。在特征提取模块中,对图片进行特征提取,并进行下采样操作,图片被转化为 30×7×128 的特征向量。然后由车牌字符定位网络精确定位车牌字符并提取字符特征。最后,在车牌字符识别模块对提取到的车牌特征进行分类,得到车牌识别结果。

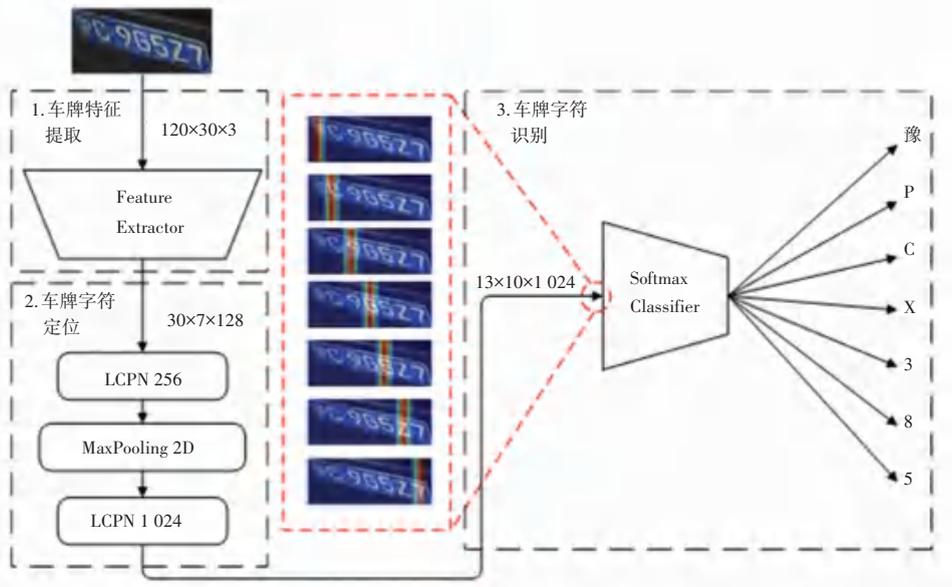


图 1 IMCRLPR 车牌识别模型总体框架图

Fig. 1 Overall structure diagram of IMCRLPR license plate recognition model

1.1 车牌特征提取

车牌特征提取是车牌字符识别的基础,一般使用卷积神经网络来提取图像中的车牌字符特征。目前性能较好的卷积神经网络(比如 ImageNet、MobileNet 等)的网络深度较深,能够提高提取到的

特征的丰富程度,但是也需要更多的计算资源和更长的计算时间,而车牌图片的组成比较简单,只有几个字符,若使用这些网络提取特征会降低模型推理速度。并且由于网络输入图片的尺寸为 120×30,所以提取特征过程中在高度方向上最多只能进行 5 次

下采样。而 ImageNet 和 MobileNet 等网络在设计之初是为了用于目标分类和图像分类等任务, 输入的图片尺寸较大, 下采样次数过多, 无法满足研究需求。因此, 就设计了一个由 3 个卷积层和 2 个最大

池化下采样层组成的网络来提取车牌特征, 该网络只进行 2 次下采样, 而且结构比较精简, 其整体结构如图 2 所示。



图 2 特征提取网络结构图

Fig. 2 Structure diagram of feature extraction network

1.2 车牌字符定位

如图 1 左下部所示, 车牌字符定位模块由 LCPN 256、MaxPooling2D 和 LCPN 1024 组成。其中, LCPN 256 和 LCPN 1024 的模型结构完全相同, 由字符定位分支和浅层特征提取分支组成, 结构如图 3 所示。LCPN 256 和 LCPN 1024 的不同点在于浅层特征提取分支中 DSC 的输出通道数不同, LCPN 256 的输出通道数为 256, LCPN 1024 的输出通道数为 1 024。

反映了该通道在整个特征图中的重要性。该值可由式(1)来计算求出:

$$S_c = F_{sq}(x_{in}) = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H x_{in}(i, j) \quad (1)$$

其中, S_c 表示压缩后的输出特征向量; x_{in} 表示一个具有 C 个通道的输入特征图; W 和 H 分别表示输入特征图的宽和高。

(2) 激活 (Excitation) 操作。主要用于生成权重, 对每个通道的特征图进行加权, 主要包含 2 步:

① 第一步, 先使用一个全连接函数来减少压缩操作的输出向量的通道数, 从而减少注意力矩阵计算操作的计算量, 再使用 $Relu6$ 函数进行激活, 这一过程可被描述为式(2)~式(3):

$$E_{C/r} = Relu6(FC_1(S_c)) \quad (2)$$

$$Relu6 = \min(\max(0, x), 6) \quad (3)$$

其中, $E_{C/r}$ 表示第一步减通道操作的输出特征向量; r 表示维度降低的比例系数; FC_1 表示用于减通道的全连接函数。 $Relu6$ 函数是改进的 $Relu$ 函数, 当输入值大于 6 时, $Relu6$ 可以将输出值全置为常数 6, 因此可以在一定程度上缓解梯度爆炸现象的发生。

② 第二步, 使用一个全连接函数将计算得到的注意力权重向量的通道数提升到与输入向量的通道数相同, 方便后续加权操作, 然后使用 $HardSwish$ 函数激活, 得到一个数值介于 0~1 之间的注意力权重向量, 推得的计算公式如下:

$$E_c = HardSwish(FC_2(E_{C/r})) \quad (4)$$

$$HardSwish = \begin{cases} 0, & x \leq -3 \\ \frac{x(x+3)}{6}, & -3 < x < 3 \\ x, & x > 3 \end{cases} \quad (5)$$

其中, E_c 是第二步提升通道操作的输出特征向量, FC_2 是用于提升通道的全连接函数。 $HardSwish$ 函数是 $Sigmoid$ 函数的改进函数, 在计算过程中只需要进行一个分支判断和一次乘法运算, 能够避免

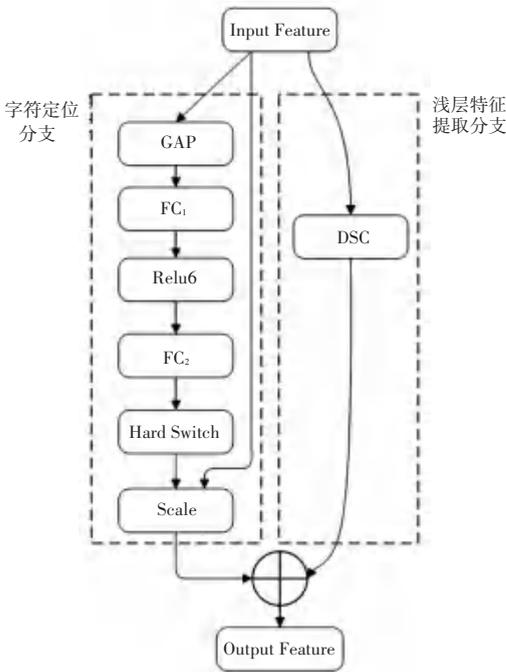


图 3 LCPN 结构图

Fig. 3 LCPN structure diagram

在 LCPN 的字符定位分支, 使用 SE 注意力从通道方向对车牌图片中不同位置的特征赋予不同的权重, 突出字符所在位置的特征信息。在 SE 中有 3 个操作, 分别是压缩、激活、特征重标定操作。对此将给出研究阐释如下。

(1) 压缩 (squeeze) 操作。主要通过全局平均池化 (Global Average Pooling, GAP) 对每个通道的特征图进行压缩, 将其转换为一个标量值, 这个标量值

指数计算过程,减小运算的开销,提高计算速度。

(3)特征重标定(feature recalibration)操作。使用注意力权重向量 \mathbf{E}_c 的各通道权重值与SE模块的输入矩阵 \mathbf{x}_{in} 进行逐通道相乘,突出原向量中的部分特征,可由式(6)进行描述:

$$\mathbf{x}_{out} = F_{scale}(\mathbf{x}_{in}, \mathbf{E}_c) = \mathbf{E}_c \cdot \mathbf{x}_{in} \quad (6)$$

其中, \mathbf{x}_{out} 表示特征重标定操作的输出特征矩阵, F_{scale} 表示特征重标定的矩阵相乘函数。

在LCPN的浅层特征提取分支,使用深度可分离卷积(DSC)来提取输入该模块的浅层特征信息,使网络在定位字符的同时更多保留在网络浅层的原始的特征,防止模型过深或者特征图太小导致的模型退化、精度下降,提高网络的泛化能力。DSC是一种改进的卷积方式^[6-7],通过把卷积过程分为深度卷积和逐点卷积两个步骤,将传统卷积过程解耦,在保持模型准确性的同时大大提高了模型的速度和效率。

深度卷积(Depthwise Convolution)用于对输入的特征向量进行逐通道卷积。把输入特征向量的每个子通道视为一个单独的特征向量进行卷积,以减小计算量,然后得到与输入特征向量通道数量相同的单个卷积结果。

逐点卷积(Pointwise Convolution)可以用来进一步优化深度可分离卷积的性能。通过采用一组大小为 1×1 的卷积核来将深度卷积步骤的输出通道合并为最终输出,可以实现通道间的组合和降维并且能够增加模型的非线性能力。

经过以上2个分支的计算后,LCPN还需要将通道注意力(SE)和深度可分离卷积(DSC)所提取的特征进行融合,这样就能够定位字符的同时又能兼顾到浅层特征的提取,提高字符定位精度。

1.3 车牌字符识别

如图1中最右部分所示,从1.2节得到车牌字符的精准定位后,车牌字符识别模块只需要使用Softmax分类器对得到的字符特征进行分类就能得到车牌识别结果,因此输入到车牌字符识别网络的特征对字符位置的定位是否准确关系到最终识别结果的对错。为了观察输入到车牌字符识别网络的特征是否准确定位,研究提取车牌字符识别网络输入特征矩阵中权重最大的特征值,然后对特征矩阵进行归一化,这样就可以得到神经网络对特征矩阵中“最感兴趣”的特征,最后将特征矩阵还原到车牌原图大小并且将特征矩阵与车牌原图进行重叠,车牌字符识别网络输入特征图可视化效果如图1中部红

色虚线框所示。

2 实验与分析

为了验证本文提出的车牌识别方法在复杂场景中的有效性,研究抽取CCPD数据集中最能体现光照强度、车牌倾斜和车牌模糊等复杂环境的3个子数据集(CCPD-DB、CCPD-Rotate和CCPD-Challenge)进行实验。

2.1 实验数据集

CCPD^[8]是一个大型的、多样化的中国城市车牌开源数据集,提供290k个带有详细注释的独特车牌图像。每张图像中仅包含一个车牌,每个车牌由7个字符组成。第1个字符为省份(共31个类别,不包含台湾省)、第2个字符为字母,剩下的5个字符中的每一字符为字母或数字(所有出现的字母不包含‘I’和‘O’,数字和字母共34个类别)。在做模型对比实验时,随机选取子数据集CCPD-base中的一半数据作为训练集,另外一半用作验证集。测试数据集是CCPD-DB、CCPD-Rotate和CCPD-Challenge这3个子数据集,因为这3个子数据集最能体现光照强度、车牌倾斜和车牌模糊等自然场景对车牌识别网络性能的影响。

2.2 实验设置

本文的实验在GeForce RTX 2060GPU上进行,实验中使用的神经网络框架为Tensorflow1.15.2。为了防止过度拟合,研究使用颜色抖动、随机裁剪、随机缩放、随机平移和随机旋转等数据增强策略。车牌识别模型训练过程中输入图像大小调整为 30×120 ,使用分类交叉熵损失函数和RMSprop优化器,batch size设置为1024,初始学习率为 1×10^{-2} ,epoch不设固定值,当连续5个epoch损失没有减少时,将学习率乘以衰减系数0.9,当连续30个epoch没有得到更低的损失时,就结束训练。车牌识别模型在测试时输入图像的大小调整为 30×120 ,batch size设置为5。

2.3 实验及分析

在本节中,采用一系列实验来验证本文提出的车牌识别方法的有效性。为了评估车牌字符定位网络(LCPN)对识别准确率的影响,使用控制变量法,删除LCPN 1024模块,只保留LCPN 256模块,控制LCPN 256模块中的DSC元素的有无和SE网络的有无来研究其对车牌识别精度和速度的影响。其中,当模型中没有DSC时,使用一个与DSC参数相同的传统卷积来代替,实验结果见表1。表1中,

“√”表示实验中使用的网络包含此列中对应的结构,“×”表示实验中使用的网络不包含此列中对应的结构。通过观测模型预测的精度和速度来评估各

个模块的影响,并且通过计算经过每个卷积层所需要的浮点计算量(Floating Point Operations, *FLOPS*)来统计本模型计算成本。

表 1 LCPN256 模块中各元素对车牌识别精度和速度的影响

Table 1 The influence of each element in the LCPN256 module on the accuracy and speed of license plate recognition

DSC	SE	Accuracy				Calculating Costs/ GFLOPS	Inference Speed/ FPS
		DB	Rotate	Challenge	Overall Accuracy		
×	×	87.6	85.6	74.8	83.9	163.4	239.9
√	×	86.9	86.2	74.5	83.6	101.5	403.0
×	√	91.1	90.0	80.8	88.3	163.5	243.8
√	√	91.2	89.4	80.7	88.1	108.5	395.6

观察表 1 中的第 1 行和第 2 行可以看出,在 LCPN256 中将传统的卷积操作替换为深度可分离卷积后,模型的整体识别精度会下降 0.3%,但是模型的计算成本下降了 38%,从而使模型的推理速度提升了 66.6%。这是因为 DSC 相较于传统的卷积,把整个卷积过程解耦为逐通道卷积和逐点卷积,减小了运算量,提高运算速度。因此在 LCPN 模块中使用 DSC 来代替传统的卷积,这样在仅牺牲小部分精度的情况下,能够实现推理速度的大幅度提高。

观察表 1 中的第 1 行和第 3 行可以看出,在 LCPN256 中使用 SE 注意力模块能够将模型的整体识别精度提高 5.2%,并且也没有明显增加模型的计算成本。这是因为 SE 注意力的主要运算是 1 个全局平均池化和 2 个涉及低维的全连接层,这些都不会带来太大的运算开销,提高运算速度。并且 SE 注意力能够在通道方向将与字符相关的特征进行加权,因此能够提高模型的单个字符识别准确性,从而提高车牌识别整体精度。

影响识别性能的另一个因素是 LCPN 模块的有无,在其他条件都相同的情况下,对 LCPN 256 模块和 LCPN 1024 模块进行依次删除,研究其对车牌识别精度的影响,实验结果见表 2。

表 2 LCPN 模块的有无对车牌识别精度的影响

Table 2 The influence of the presence or absence of LCPN module on the accuracy of license plate recognition

LCPN 256	LCPN 1024	Accuracy			
		DB	Rotate	Challenge	Overall Accuracy
×	×	87.6	85.6	74.8	83.9
√	×	91.2	89.4	80.7	88.1
×	√	94.8	93.6	84.2	91.8
√	√	96.0	98.7	85.4	94.0

将表 2 中的第 1 行、第 2 行和第 3 行进行比较,第 1 行中的网络没有加入 LCPN 模块,第 2 行只加入了 LCPN 256 模块,第 3 行只加入了 LCPN 1024 模块,由此可以看出 LCPN 模块可以提高模型的识别精度。这是因为 LCPN 模块可以精确定位车牌图片中的字符位置,使网络能够正确提取车牌字符特征,进而提高车牌识别精度。

将表 2 中的第 2 行与第 3 行进行比较,第 2 行中的 LCPN 256 模块在进行特征融合的时候输入的是 256 维的特征向量,第 3 行中的 LCPN 1024 模块在进行特征融合的时候输入的是 1024 维的特征向量,实验结果表明在模型中加入 LCPN 1024 对精度的提高比加入 LCPN 256 好。这是因为更多维度的特征向量会包含更多的特征,模型也就能学习到更多的信息,从而带来识别精度的提高。

将表 2 中的第 2 行和第 3 行与第 4 行进行比较,第 2 行和第 3 行都只加入了一个 LCPN 模块,第 4 行叠加了 2 个 LCPN 模块,实验结果表明叠加 LCPN 模块也会提升模型的识别正确率。这是因为叠加 LCPN 模块可以在原来的基础上再次提高对车牌中字符定位的精度,提高车牌识别的精度。

3 结束语

针对自然场景中复杂条件下的车牌检测识别问题,本文提出一种改进的多标签分类结构的实时车牌识别模型 IMCRLPR。该模型在车牌特征提取部分设计一个层数较浅的特征提取网络,避免因为网络过深而引起计算量增加。在车牌字符定位部分设计了一个基于特征融合的轻量化通道定位网络:采用轻量化的深度可分离卷积替代传统卷积,减小计算量,提高运算速度;加入通道注意力机制 SE 模

块,从通道方向对车牌图片中不同位置的字符特征进行关注;故该模块能够帮助模型精准定位复杂场景中车牌内的字符位置。在车牌字符识别部分,使用 *softmax* 分类器对字符进行分类,从而识别车牌字符,得到车牌号码。最后,在 CCPD 数据集中选取最能体现光照强度、车牌倾斜和车牌模糊等复杂场景的 3 个子数据集 (CCPD-DB、CCPD-Rotate 和 CCPD-Challenge) 来对本文提出的模型进行测试,因为模型采用轻量化的设计,并且网络结构简单,所以提出的模型在以上数据集的测试中达到 94.0% 的平均准确率和 395.6 FPS 的识别速度,这表明 IMCRLPR 鲁棒性好、识别速度快,能够完成复杂场景下的车牌识别任务。

参考文献

- [1] 武云飞. 车牌识别系统中的字符分割和识别算法研究[J]. 安阳师范学院学报, 2020(5): 32-35.
- [2] 吕文强, 王启爱. 一种有效的提高车牌字符分割率的方法[J]. 信息化研究, 2017, 43(4): 42-45.
- [3] 刘丽丽. 模板匹配联合特征法在车牌字符识别中的应用[J]. 工程建设和设计, 2022(2): 62-66.
- [4] 张彩珍, 李颖, 康斌龙, 等. 基于深度学习的模糊车牌字符识别算法[J]. 激光与光电子学进展, 2021, 58(16): 259-266.
- [5] 高昕威. 基于深度学习的汽车车牌识别算法研究[J]. 机电工程技术, 2021, 50(10): 164-166.
- [6] HU Jie, SHEN Li, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
- [7] HOWARD A G, ZHU Menglong, CHEN Bo, et al. Mobilenets: Efficient convolutional neural networks for mobile vision applications[J]. arXiv preprint arXiv:1704.04861, 2017.
- [9] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified real-time object detection[C]//Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA; IEEE, 2016: 779-788.
- [10] GOU Chao, WANG Kunfeng, YAO Yanjie, et al. Vehicle license plate recognition based on extremal regions and restricted Blotzmann machines [J]. IEEE Transactions on Intelligent Transportation Systems, 2016, 17(4): 1096-1107.
- [11] XU Zhenbo, YANG Wei, MENG Ajin, et al. Towards end-to-end license plate detection and recognition: A large dataset and baseline [C]//Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany; IEEE, 2018: 255-271.