

# 基于级联卷积神经网络的车牌定位

傅 鹏,谢世朋

(南京邮电大学 通信与信息工程学院,江苏 南京 210003)

**摘要:**针对多车辆、低分辨率等复杂环境下的车牌定位问题,提出了一种基于人眼视觉特性的车牌识别方法。通过模仿人眼视觉原理,利用级联卷积神经网络分层提取目标区域特征,逐步缩小搜索区域的方法,实现车牌的精准定位。首先通过运动目标检测算法定位出目标运动热点区域;然后使用卷积神经网络识别热点区域中的车辆;最后使用卷积神经网络从定位的车辆图片中识别车牌。数据集采集于多个交通路口的天网摄像头,然后对5 000幅图像,约15 000个目标进行人工标注,同时对训练图片进行随机变换,从而提高训练的有效性。实验结果表明,通过提取运动区域可提升卷积神经网络运行的速度和识别的精度。相比于传统车牌识别算法,提出的方法极大地提高了复杂场景下的车牌识别率,同时在处理高分辨率的图片时具有更高的车牌定位率。

**关键词:**车牌定位;运动目标检测;视觉特性;卷积神经网络

**中图分类号:**TP301.6

**文献标识码:**A

**文章编号:**1673-629X(2018)01-0134-04

**doi:**10.3969/j.issn.1673-629X.2018.01.029

## License Plate Location Based on Cascaded Convolution Neural Network

FU Peng, XIE Shi-peng

(School of Communication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

**Abstract:** Aiming at the problem of license plate positioning in complex environments such as multi-vehicle and low resolution, we present a license plate recognition method based on human vision. By imitating the visual principle of human eyes, the precise positioning of license plate is realized by the approach of extraction of target region characteristics through cascade convolution neural network and gradually narrowing the search area. Firstly, the target motion region which we are interested in is located by the motion detection. Then, the vehicle identification is performed on the hot spot region by convolutional neural network. Finally, license plates are located in vehicle picture. Training pictures are collected in 20 different traffic junctions of the skynet camera images, as well as nearly 5 000 images and about 15 000 targets labeled by manual. At the same time, the labeled images are transformed randomly to improve the effectiveness of the training. According to the experiments, the extraction of motion region enhances the speed and recognition precision of convolutional neural network, and greatly improves the license plate recognition rate in complex scenes compared to the traditional license plate recognition algorithm. Moreover, it performs better in dealing with high-resolution pictures.

**Key words:** license plate location; moving target detection; human visual characteristics; convolution neural network

### 1 概述

近年来,车牌识别受到了广泛的关注和研究<sup>[1-2]</sup>,并大量应用于高速公路、停车场、公路自动收费站、居民小区等环境中。其中车牌定位是整个车牌识别技术的核心,一直是受关注的热点。谭同德等<sup>[3]</sup>提出一种基于开源视觉库 OPENCV 的车牌定位方法;利用车牌的颜色特征,将原图像分别在 HSV 颜色空间和 RGB 颜色空间下处理得到两幅二值图像,根据这两幅二值

图像的纹理特点,通过在 HSV 颜色空间下得到的二值图像定位出车牌的上下边界,再按照定位出的上下边界坐标从 RGB 颜色空间下的二值图像中水平切割出车牌区域。实验表明该方法在卡口环境下有较好的定位效果。

叶必锭等<sup>[4]</sup>提出一种基于结构特征和纹理特征的车牌定位方法;对车辆图像进行预处理,检测垂直方向的边缘,执行数学形态学操作,并利用车牌的结构特征

收稿日期:2017-02-20

修回日期:2017-06-27

网络出版时间:2017-10-19

基金项目:江苏省科技重点研发计划-产业前瞻与共性关键技术(BE2016001-4);教育部-中国移动科研基金(MCM20150504)

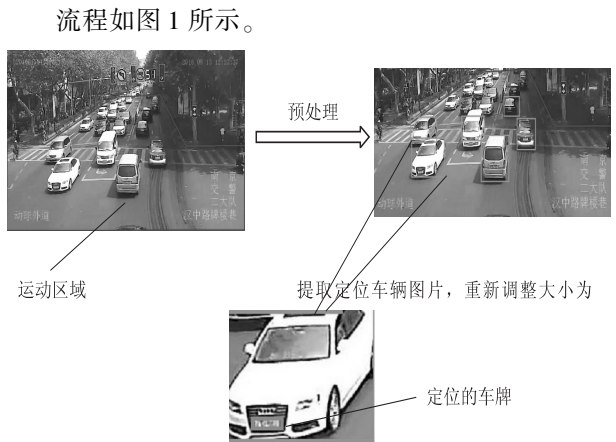
作者简介:傅 鹏(1993-),男,硕士生,研究方向为图像处理;谢世朋,副教授,研究方向为图像处理。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20171019.1625.058.html>

粗定位车牌区域;提取车牌的纹理特征并构建特征向量,通过贝叶斯分类器精确定位车牌区域。实验表明该方法不受车牌大小、位置等因素的限制,定位率高,速度快,具有较好的鲁棒性。罗斌等<sup>[5-6]</sup>提出了基于角点回归的全卷积神经网络的车牌定位。该算法不仅能够标准在标准交通卡口数据上进行有效定位,而且在光照条件低下、车辆倾斜严重等一些复杂环境下,同样具有较高的鲁棒性。但是,以上方法在车牌清晰、车牌大小合适的情况下较适用,但是在交通路口天网摄像头拍摄到的画面中,由于车辆数量较多、目标车牌较小、光线影响等,很难在监控视频中利用车牌自身的边缘和颜色特征定位到车牌。

深度学习<sup>[7-9]</sup>作为传统机器学习的延伸和扩展,已经在众多领域得到广泛应用。卷积神经网络作为深度学习的一种,具备深度学习自动学习并提取特征的特点,利用局部感受野和权值共享提高计算速度,泛化能力强,识别效率高,广泛用于图像识别等方面。Redmon 等将卷积神经网络和人类视觉系统原理相结合,提出了 YOLO 网络:将分类问题作为回归问题处理并直接将原图像输入网络,取得了很好的效果。YOLO 模型处理图像速度达到了 45 fps (frames per second),而 Fast YOLO 网络处理图像的速度能达到 155 fps。但是 YOLO 网络对小物体识别效果不佳,所以在复杂环境下,直接通过 YOLO 网络定位车牌效果并不理想。

人类视觉对物体的认知,是先将其作为统一的整体,而后才以部分的形式认知物体<sup>[10-11]</sup>。当需要从复杂的场景中寻找并识别车牌时,首先会在场景中寻找到车辆的位置,然后再从车辆上寻找到车牌,从而识别车牌字符。基于这种思想,提出了将运动区域检测和卷积神经网络相结合的车牌定位方法,用于定位识别交通路口天网摄像头下的车辆牌照。首先定位车辆的运动区域,然后在运动区域中定位车辆,最后从定位到的车辆中提取车牌信息。



万方数据 1 检测系统流程

## 2 网络详述

### 2.1 输入视频

由于该网络定位于处理复杂场景下的车牌定位问题,所以输入视频为复杂场景下天网摄像头拍摄的視頻,难点在于多车辆、低分辨率、室外环境等。但是由于视频中帧间存在大量的冗余信息<sup>[13]</sup>,若将其全部输入网络将得到大量的重复信息,导致网络效率下降,所以在视频中间隔特定的时间抽取部分帧作为网络的输入,一方面可以避免处理重复信息,另一方面可以提高处理速度。

### 2.2 预处理

由于车辆处于运动热点区域<sup>[14]</sup>,因此可以首先检测视频中车辆的运动区域,该步骤可有效删除视频中的部分冗余信息,去除与车辆无关的区域,从而进一步减少了车牌搜索区域,提高了网络的识别效率。实验结果表明,经过预处理的图片识别率高于未经预处理的图片识别率。

### 2.3 车辆识别

文中网络从整个图像中提取特征,从而对目标进行预测,并改善了 YOLO 网络对小物体的识别效果。网络将图像分为  $S \times S$  的网格,如果物体位于某个网格中,则在该网格中检测该物体。

每个网格预测  $B$  个边界框的位置,每个边界框包含 5 个预测值:  $x, y, w, h, Q$ 。 $(x, y)$  表示边界框的中心,  $w$  和  $h$  分别表示预测边界框的宽度和高度,  $Q$  代表了预测的边界框和真实物体位置的交叠率。其中

$$Q = \text{Pr}(\text{Object}) * \text{IOU}_{\text{pred}}^{\text{truth}} \tag{1}$$

其中,  $\text{Pr}(\text{Object})$  为边界框包含物体的概率;  $\text{IOU}_{\text{pred}}^{\text{truth}}$  为预测边界框和真实背景 的交叠率。如果该网格中没有物体,则  $Q$  应该为 0。

同时每个网格还对预测类别的概率进行预测,记为  $\text{Pr}(\text{Class}_i | \text{Object})$ 。该概率和网格包含的物体有关,由于车牌在整个图像中较小,所以仅考虑每个网格中只存在一种目标类别,即每个网格只对一类目标进行预测。每个边界框预测类别  $C$  的概率为:

$$\text{Pr}(\text{Class}_i | \text{Object}) * \text{Pr}(\text{Object}) * \text{IOU}_{\text{pred}}^{\text{truth}} = \text{Pr}(\text{Class}_i) * \text{IOU}_{\text{pred}}^{\text{truth}} \tag{2}$$

该值不仅显示了类别  $C$  出现在该边界框内的概率大小,还体现出边界框的位置与物体准确位置的 距离。

网络最后一层不仅预测类别概率还预测边界框的位置。其中,将表示位置的宽度和高度通过原图像的宽度和高度归一化在 0 到 1 之间。同理边界框中心点  $x, y$  也通过该方法进行归一化,使其值归一化在 0 到 1 之间。

在网络的最后一层采用线性激活函数,在其他层使用如下的激活函数:

$$\varphi(x)=\begin{cases}x,x>0\\0.1x,otherwise\end{cases}\tag{3}$$

因为误差平方和函数容易优化并且能实现平均精度的最大化,所以卷积网络使用已优化的误差平方和函数作为损失函数。但是误差平方和存在以下问题:误差平方和函数错误地将边界框位置错误等级等同于分类错误等级。由于大部分的网格中没有包含任何物体,所以这些网格的得分  $Q$  接近于 0,这会影响存在物体的网格的梯度,可能会导致模型不稳定,使训练过早地发散。为了解决上述问题,文中网络加重了位置预测误差造成的损失,减少了不包含物体的网格的  $Q$  值的损失。因此为其分别设置系数  $\lambda_{\text{coord}}=5,\lambda_{\text{noobj}}=0.5$ 。由于大面积的边界框定位容错率高于小面积边界框定位容错率,所以网络预测的是边界框的高度和宽度的平方根而不是直接预测边界框的高度和宽度。

网络在一个网格中预测很多个边界框,而对于每个物体只需一个边界框,所以选择在真实背景下当前 IOU 值最高的边界框作为该物体的定位框。这使得每个定位框在预测物体的大小、物体的类别等方面变得越来越准确。

基于上述理论,在训练阶段使用了如下的多部分损失函数:

$$\begin{aligned} &\lambda_{\text{coord}}\sum_{i=0}^{s^2}\sum_{j=0}^B1_{i-j}^{\text{obj}}[(x_i-\hat{x}_i)^2+(y_i-\hat{y}_i)^2]+\\ &\lambda_{\text{coord}}\sum_{i=0}^{s^2}\sum_{j=0}^B1_{ij}^{\text{obj}}[(\sqrt{w_i}-\sqrt{\hat{w}_i})^2+(\sqrt{h_i}-\sqrt{\hat{h}_i})^2]+\sum_{i=0}^{s^2}\sum_{j=0}^B1_{ij}^{\text{obj}}(c_i-\hat{c}_i)^2+\\ &\lambda_{\text{noobj}}\sum_{i=0}^{s^2}\sum_{j=0}^B1_{ij}^{\text{noobj}}(c_i-\hat{c}_i)^2+\\ &\sum_{i=0}^{s^2}1_i^{\text{obj}}\sum_{c\in\text{classes}}(p_i(c)-\hat{p}_i(c))^2\tag{4} \end{aligned}$$

其中,  $1_i^{\text{obj}}$  表示物体出现在网格  $i$  中;  $1_{ij}^{\text{obj}}$  表示网络通过第  $j$  个边界框预测网格  $i$  中的物体。

如果在一开始就采用较高的学习率,由于梯度不稳定,模型会发散,所以学习率从  $10^{-3}$  到  $10^{-2}$  缓慢增加。为了避免过拟合,采用 Dropout 和数据集扩充方法。Dropout 在第一层后随机 (rate=0.5) 删除一些隐藏层神经元,从而避免层与层之间相互适应。数据集扩充方法采用随机缩放并改变原图像大小,同时在 HVS 彩色空间中,以 1.5 倍的因子随机调整图像的曝光度和饱和度。

2.4 车牌定位

在经过卷积网络定位到车辆位置后,得到单个车辆的图片,并重新调整大小为  $225\times225$  像素。在得到的图片中,车牌大小与图片大小的比例在一定的范围内,如图 2 所示。车牌定位部分的卷积神经网络结构

和车辆定位部分相同。训练采用 20 000 幅车牌图片,具体细节和车辆定位部分网络相同,不再赘述。将车辆图片再次输入预训练好的车牌定位卷积网络,输出定位到的车牌。

3 实验结果与分析

将文中方法与传统的车牌定位方法进行对比分析,识别率从平均单幅图片只能检测到 0.131 个车牌提升 4.312 个车牌,识别率大大提升。



图 2 不同环境下的对比识别效果

预处理对检测车辆的影响如表 1 所示。

表 1 预处理对检测车辆的影响

指标	经过预处理	未经预处理
平均每帧检测时间/s	1.006 5	1.078 6
平均检测到车辆数	5.471	4.346
平均每辆车定位所用时间/s	0.183 9	0.248 1

从表 1 可以看出,经过预处理之后的图片平均每帧检测时间略有减少,这是由于在预处理过程中截取了图像中的运动热点区域,排除了大量的冗余信息,增大了每一个网格中包含目标物体的概率。识别车牌数量从 4.346 提升到 5.471,提升近 26 个百分点,所以在视频帧输入之前进行预处理能取得非常好的效果。

算法对比如表 2 所示。

文中方法平均每帧检测时间明显少于基于颜色特征和边缘检测的车牌定位算法,一方面是由于 YOLO 网络运行速度较快,另一方面是由于在定位出的车辆图片中车牌部分特征较原图明显,所以定位速度快。定位到的车牌数目也多于传统车牌定位算法,大约是传统车牌定位算法的五倍,车牌定位的效率远远大于传统车牌定位算法。当然,定位到的车辆数量大于能够定位到的车牌数目,一方面是由于车辆定位不完整,没有包含全部车牌,另一方面是由于图片的分辨率过低,不能从已经定位到的车辆上识别出车牌。

表 2 文中方法和基于颜色特征和边缘检测的车牌定位算法的对比

方法	平均每帧检测时间/s	平均每帧检测到的车辆数	平均每帧定位车牌数	平均每车牌定位所用时间/s
文中方法	1.100 1	5.471	3.612	0.304 5
基于颜色特征和边缘检测的车牌定位算法	2.413	无	0.623	3.873

表 3 比较了不同分辨率的图片的车牌检测效果。其中测试集采用超清摄像机拍摄的 200 幅图片,通过 OPENCV 函数将图片的分辨率从 3 264×1 836 转化为 1 280×720,该操作保证图片拍摄的内容和角度完全相同,不会因为分辨率的不同而导致图片内容的缺失。

表 3 不同分辨率下的车牌检测效果

不同分辨率图片	平均每帧检测时间/s	平均每帧检测到的车牌数目
1 280×720	1.100 1	3.612
3 264×1 836	1.342 4	4.891

从表 3 可以看出,高分辨率的图片检测耗时有所增加,但是检测到的车牌数目也相应增加。虽然现在天网监控分辨率以 1 280×720 为主,但是随着硬件的提升,更高分辨率的监控也在逐渐部署,所以文中方法在摄像头硬件提升后有更大的应用空间。

4 结束语

由于在复杂的环境下,特别是在分辨率有限、车辆数目多而且车牌不清晰的环境中,传统算法无法准确定位车牌。受人类视觉特征启发,文中首先通过运动区域检测缩小车辆的搜索区域,然后通过卷积网络定位车辆,最后缩小车牌搜索区域。实验结果表明,在复杂的场景下尤其是在多车辆场景下,该方法有较高的识别率。但是该方法也存在由于车辆的定位误差导致无法进行车牌定位的问题,因此下一步将提高车辆定位的完整性,从而提高车牌定位率。

参考文献:

[1] YANG F,MA Z. Vehicle license plate location based on histogramming and mathematical morphology [ C ]//Fourth IEEE workshop on automatic identification advanced technologies. [ s. l. ]:IEEE,2005:89-94.

[2] LI G,ZENG R,LIN L. Research on vehicle license plate location based on neural networks [ C ]//International conference on innovative computing,information and control. [ s. l. ]:IEEE Computer Society,2006:174-177.

[3] 谭同德,王三刚. 基于 OpenCV 的车牌定位方法 [ J ]. 计算机工程与设计,2013,34(8):2816-2820.

[4] 叶必锭,李 霆. 基于特征多样性的车牌定位方法 [ J ]. 自动化与信息工程,2012,33(2):17-22.

[5] 罗 斌,郜 伟,汤 进,等. 复杂环境下基于角点回归的全卷积神经网络的车牌定位 [ J ]. 数据采集与处理,2016,31(1):65-72.

[6] 郜 伟. 复杂环境下车牌定位的研究与应用 [ D ]. 合肥:安徽大学,2016.

[7] 陈先昌. 基于卷积神经网络的深度学习算法与应用研究 [ D ]. 杭州:浙江工商大学,2013.

[8] KRIZHEVSKY A,SUTSKEVER I,HINTON G E. ImageNet classification with deep convolutional neural networks [ C ]//International conference on neural information processing systems. [ s. l. ]:Curran Associates Inc. ,2012:1097-1105.

[9] SZEGEDY C,LIU W,JIA Y,et al. Going deeper with convolutions [ C ]//IEEE conference on computer vision and pattern recognition. [ s. l. ]:IEEE Computer Society,2015:1-9.

[10] 周昌雄,曹丰文,崔 鸣,等. 基于格式塔心理学原理的几何活动轮廓模型 [ J ]. 中国图象图形学报,2008,13(5):924-929.

[11] ROCK I,PALMER S. The legacy of gestalt psychology [ J ]. Scientific American,1991,263(6):84-90.

[12] 柴 毅,阳小燕,尹宏鹏,等. 基于运动区域检测的运动目标跟踪算法 [ J ]. 计算机应用研究,2010,27(12):4787-4789.

[13] 肖永豪,余英林. 基于视频对象的自适应去帧/插帧视频处理 [ J ]. 华南理工大学学报:自然科学版,2003,31(8):6-9.

[14] 伏思华,张小虎. 基于序列图像的运动目标实时检测方法 [ J ]. 光学技术,2004,30(2):215-217.