

基于级联多任务深度学习的卡口识别引擎研究

何 霞 汤一平 袁公萍 陈 朋 王丽冉

(浙江工业大学信息工程学院 杭州 310023)

摘 要 针对在将卡口非结构化视频图像数据转化为智能结构化信息的过程中存在环境的复杂性、需求的多样性、任务的关联性和识别的实时性等问题,提出了一种级联多任务深度学习网络的卡口识别引擎方法,其通过充分利用分割、检测、识别等任务之间的相互联系实现了高精度的、高效的、同步实时的卡口车辆多种基本信息的识别(车型、品牌、车系、车身颜色以及车牌等识别任务)。首先,利用深度卷积神经网络自动完成车型的深度特征学习,在特征图上进行逻辑回归,从卡口道路复杂背景中提取出感兴趣区域(包括多车辆对象);然后,利用多任务深度学习网络对提取出来的车辆对象实现多层次的多任务识别。实验结果表明,提出的方法在识别精度和效率上都明显优于传统计算机视觉方法和现有的基于深度学习的识别引擎技术,该方法对车型、品牌、车系及车牌的识别与检测精度均达到 98% 以上,检测效率提升了 1.6 倍。

关键词 卡口识别引擎,深度学习,级联网络,多任务深度学习,卷积神经网络

中图分类号 TP391.4

文献标识码 A

DOI 10.11896/j.issn.1002-137X.2019.01.047

Study on Bayonet Recognition Engine Based on Cascade Multitask Deep Learning

HE Xia TANG Yi-ping YUAN Gong-ping CHEN Peng WANG Li-ran

(School of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China)

Abstract Aiming at the complexity of environment, the diversity of requirements, the relevance of tasks and the real-time of identification in the process of converting the unstructured video data of bayonet into the intelligent structured information, this paper proposed a method of bayonet recognition engine based on cascade multitask deep learning. This method makes full use of the relationship between segmentation and detection recognition tasks to achieve high-precision, efficient, synchronous and real-time recognition of a variety of basic information of bayonet vehicles (motorcycle types, brands, series, colors and license plates etc.). First, the deep convolutional neural network is used to automatically extract the depth feature and the logical regression is performed on the feature map to extract the interested region from the complex background (including multi-vehicle object). And then the multitask deep learning network is used to achieve multilevel multitask recognition for the extracted vehicle objects. Experimental results show that the proposed method is superior to the traditional computer vision method and the existing recognition engine technology based on deep learning in terms of recognition accuracy and efficiency, and the accuracy of recognizing and detecting the motorcycle types, brands, series and license plates is more than 99% respectively, and the detection efficiency is increased by 1.6 times.

Keywords Bayonet recognition engine, Deep learning, Cascade network, Multitask deep learning, Convolutional neural network

1 引言

随着社会经济的快速发展,机动车日益成为人们日常生活必不可少的交通工具,同时也成为了不法分子从事非法活动的必须工具。目前各省、市级高速公路和主干线、城市出入口及主要交通要道均部署了卡口设备(如高清摄像机),对过往车辆进行实时图像采集,但当前的极大多数卡口一般都是基于车牌识别技术的,嫌疑车辆一旦使用假牌、套牌、无牌以

及不断更换车牌的方式,便可逃避现有卡口对其的跟踪和识别。如何充分挖掘视频图像信息,实现车辆类型等特征的检索和识别(识别出车型、品牌、车系、车身颜色),无疑对于违章稽查、肇事逃逸追捕、嫌疑车辆分布式监控、套牌车辆鉴别,以及加快刑侦破案效率都具有非常重要的现实意义和十分广阔的应用前景^[1]。

目前人工智能时代的基于图像的车辆识别方法大致可以分为两类:基于模型的方法和基于外观的方法。基于模型的

到稿日期:2017-10-16 返修日期:2018-01-10 本文受国家自然科学基金(61070134,61379078)资助。

何 霞(1993—),女,硕士生,主要研究方向为计算机视觉、深度学习,E-mail:178332747@qq.com;汤一平(1958—),男,教授,博士生导师,主要研究方向为全方位视觉传感器及应用、计算机视觉、机器学习,E-mail:typ@zjut.edu.cn(通信作者);袁公萍(1992—),男,硕士生,主要研究方向为计算机视觉与深度学习;陈 朋(1992—),男,硕士生,主要研究方向为机器学习、深度学习;王丽冉(1993—),女,硕士生,主要研究方向为机器学习、深度学习。

方法^[2]计算车辆的三维参数(如长度、宽度和高度)来实现车型识别。基于外观的方法^[3]通过提取车辆对象的外形特征(如 Sift^[4], Sobel^[5], Hog^[6]等)结合后续的分类器实现车型等的识别,虽然取得了较好的识别率,但是在识别精度和鲁棒性等方面还存在着不少不尽人意的地方。其存在的主要问题是:1)需要人工来完成特征提取,即需要具备相关先验知识;2)受自然环境的影响太大,各类检测算法要求的条件太苛刻,且车辆种类繁多,区别性特征微小,导致算法的鲁棒性差。

近年来,深度学习在计算机视觉领域得到了迅速发展,其能利用大量的训练样本和隐藏层逐层深入地学习图像的抽象信息,更全面直接地获取图像特征。数字图像是以矩阵来描述卷积神经网络,从局部信息块出发,进而描述图像的整体结构,故在计算机视觉领域的深度学习方法中大多采用卷积神经网络来解决图像特征提取问题。围绕着提高检测精度和检测速率两项重要指标,卷积神经网络技术从 R-CNN^[7], Fast R-CNN^[8]逐步发展到 Faster R-CNN^[9],具体表现为进一步的精度提升、加速、端到端及更实用化,几乎覆盖了分类、检测、分割、定位各个领域。

目标定位检测是图像理解中的一个基本问题。目前深度学习中普遍采用的方法是 R-CNN 及其改进版 Fast R-CNN 和 Faster R-CNN。它们将目标检测问题分解为两个级联任务:1)从复杂背景图像中提取出前景对象的可能性区域;2)对生成的可能性区域进行分类识别。虽然以上方法很好地解决了检测与识别的统一,但还有很大的提升空间。

目前,在智能交通行业中车辆基本信息的识别准确率还不高,车身颜色、品牌和车系的识别率还不能满足实战的要求。其原因主要有如下几点:1)没有很好地利用检测定位和识别多任务之间的关联性;2)没有充分利用各任务之间的递进级联关系,如:车辆精准定位→颜色、车型、品牌和车系识别→车牌精准定位→车牌识别及车牌颜色识别。

多任务深度学习网络的图像识别往往优于单任务深度学习网络,这是由于多任务在学习过程中存在任务之间的关联性,即任务之间存在信息的共享,这也是多任务深度学习的必要条件。在同时训练多个任务时,网络利用任务之间的共享信息增强系统的归纳偏置能力和分类器的泛化能力。

在深度卷积神经网络中引入多任务学习机制,可以利用多个相关任务的权值共享来减弱过拟合对神经网络的影响。文献[10]在字符检测网络中添加分割及字符细粒度识别两个辅助任务,取得了很好的识别效果。文献[11]提出了一种面部关键点检测的多任务人脸识别方法,也获得了较好的识别效果,提高了人脸识别的精度。

在卡口识别引擎研究中,本文提出一种级联多任务深度学习网络。在构建网络时,需要探讨以下几个问题:

1)如何有效化解深度卷积神经网络的学习能力与车型数据集偏小的矛盾。

深度卷积神经网络适用于解决数据量大的复杂问题,如 ILSVRC 竞赛中的图像分类问题,其是一个包含 120 万幅图像、1000 个物体类别的分类问题,图像数量大、类别多,采用卷积神经网络比传统方法具有更大的优势。但是将其应用于卡口识别时,图像的样本数量和类别数量都很难达到以上要求。因此,本文致力于发挥深度卷积神经网络在车型等多任

务识别中的作用,解决数据量受限且类别数量少的问题。

2)如何兼顾深度卷积神经网络的检测精度与检测速率的矛盾。

通常,网络的深度与检测精度成正相关,与检测速率成负相关。如何提高车型识别的精度和速率也是本文需要解决的一个问题。

3)在构建多任务学习网络时致力于如何有效利用任务之间的正相关性,减少它们之间的负相关性,实现参数的共享,从而提高精度与效率。

卡口识别引擎中包括车型识别、车辆品牌识别、车系识别、车身颜色识别以及车牌检测等多个任务。在构建多任务学习网络时,需要将具有相关性的任务进行一定的特征融合,如对于车型、品牌和车系识别任务,将车身颜色识别和车牌检测任务独立训练。

针对上述问题,本文的主要贡献如下:

1)为了解决车型数据集小、难以发挥深度学习网络优势以及训练难收敛的问题,本文采用迁移学习机制^[12],在训练 ImageNet^[13]得到的模型的基础上进行参数初始化,以提高收敛速度,改善识别效果。

2)针对 Faster R-CNN 的 RPN(Region Proposal Net, RPN)网络生成可能性区域建议框中仍然包含了很大比例的背景区域的问题,本文在 Faster R-CNN 网络的基础上添加一个二分类识别网络,以对 RPN 网络输出可能性区域进行识别,剔除背景区域,只将前景区域进行后续的训练,最终降低背景干扰的影响。

3)联合卡口的实际需求与多任务学习的优势,本文提出一种多任务学习识别方法,实现多个任务相互学习。任务之间存在参数共享,同时也相互独立。本文方法在卡口数据集上取得了很好的实验效果,识别的精度与速率都优于现有方法。

2 级联多任务卷积神经网络的结构

本文提出的级联多任务深度卷积神经网络的整体框架如图 1 所示,主要由两个网络模型构成。第一个模型是由 Faster RCNN^[9]网络与添加的二分类识别网络构成,二分类识别网络如图 1 中的虚线区域所示。与 Faster R-CNN 网络相比,该网络利用新添加的二分类识别网络将 RPN 网络输出的可能性区域中的背景区域剔除,以减小干扰,提高精度。第二个模型是对 Faster R-CNN 网络的识别网络进行修改。在二分类识别网络后构建一个多层次、多标签、多特征融合的逐层递进的多任务学习网络,如图 1 中的实线区域所示。

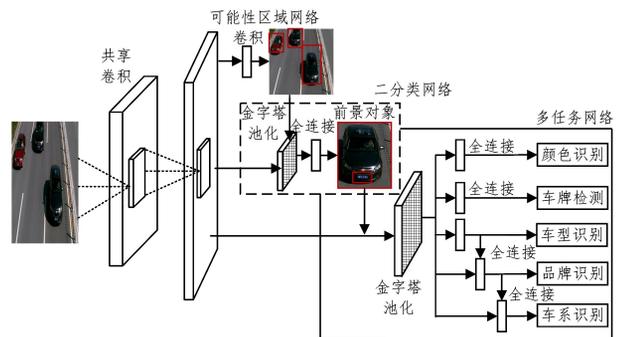


图 1 级联多任务卷积神经网络的结构

Fig. 1 Structure of cascade multitask convolutional neural networks

2.1 目标区域检测网络

RPN 网络将一个任意尺度的图像作为输入,输出矩形目标建议框的集合,每个框包含 4 个位置坐标变量和一个得分。为了生成区域建议框,首先,使输入图像经过卷积共享层生成特征图;然后,在特征图上进行多尺度卷积操作,具体实现为在每一个滑动窗口的位置使用 3 种尺度和 3 种长宽比,以当前滑动窗口的中心为中心,分别使用一种尺度和长宽比,则可以在原图上映射得到 9 种不同尺度的候选区域,如对于大小为 $w \times h$ 的共享卷积特征图,则总共有 $w \times h \times 9$ 个候选区域;最后,分类层输出 $w \times h \times 9 \times 2$ 个候选区域的得分,即对每个区域是目标/非目标的估计概率,回归层输出 $w \times h \times 9 \times 4$ 个参数,即候选区域的坐标参数。

训练 RPN 网络时,给每个候选区域分配一个二进制的标签,来标注该区域是不是目标。具体操作如下:1)对于与某个真正目标区域(Ground Truth, GT)有最高的交集并集之比(Intersection-over-Union, IoU)重叠的候选区域,给予正标签;2)对于与任意 GT 包围盒有大于 0.7 的 IoU 重叠的候选区域,给予正标签;3)将负标签分配给所有 GT 包围盒的 IoU 比率都低于 0.3 的候选区域;4)介于两者之间的舍弃。

根据上述操作,遵循 Faster R-CNN^[18]中的多任务损失,最小化目标函数。一张图像的损失函数定义为:

$$L(\{p_i\}, \{t_i\}) = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* L_{reg}(t_i, t_i^*) \quad (1)$$

其中, i 是第 i 个候选区域的索引, p_i 是候选区域为第 i 类的概率。如果候选区域的标签为正,则 p_i^* 为 1;如果候选区域标签为 0,则 p_i^* 为 0。 t_i 是一个向量,表示预测的包围盒的 4 个参数化坐标, t_i^* 是对应的 GT 包围盒的坐标向量。 N_{cls} 和 N_{reg} 分别为分类损失函数和权重参数。分类损失函数 L_{cls} 是两个类别(目标和非目标)的对数损失:

$$L_{cls}(p_i, p_i^*) = -\log[p_i^* p_i + (1-p_i^*)(1-p_i)] \quad (2)$$

位置回归损失函数 L_{reg} 的定义为:

$$L_{reg} = (t_i, t_i^*) = R(t_i - t_i^*) \quad (3)$$

其中, R 是鲁棒的损失函数(smooth L1)。

$$smooth_{L1}(x) = \begin{cases} 0.5x^2, & \text{if } |x| < 1 \\ |x| - 0.5, & \text{otherwise} \end{cases} \quad (4)$$

2.2 级联网络结构的设计

目标对象的识别问题可以定义为:目标对象在图像中的位置,实现目标检测,再对每个对象进行具体的识别。以上解决方案与人类看到事物的注意机制相一致,首先获取场景中的大致印象,然后专注于感兴趣的地区,即是什么。一个理想的可能性目标检测网络应该生成尽可能少的可能性目标区域,同时覆盖几乎所有的目标对象。卷积神经网络具备强有力的深层特征抽象能力,RPN^[9]网络能够捕获不同物体之间的相似性。然而,前景对象区域与背景区域也存在显著性区别,如色彩特征、边缘特征及封闭性特征等。因此 RPN 网络的实际输出则是目标对象的可能性区域,其中不可避免地包含了大量的背景区域。

为了解决以上问题,本文在 RPN 网络后引入一个额外的二分类网络,即对前景与背景对象区域进行划分,预先提出背景区域的干扰因素,如图 1 中虚线区域网络所示。在这样的

一个级联网络结构中,RPN 网络对输入图像学习相应的纹理特征,并生成大量的可能性区域;而后续的二分类识别网络的输入为 RPN 网络的输出,其对每个区域的细节特征进行学习。因此,级联的方法弥补了 RPN 网络的不足。

级联结构的优点有两方面:1)额外添加的二分类识别网络进一步提高了建议质量并缩小了对象的背景区域,使建议更符合任务要求;2)来自不同对象的可能性区域相互融合,使得信息互补利用。

二分类网络包括一个金字塔池化层、一个全连接层以及一个损失函数层。损失函数层采用 softmax 配合对数似然代价函数训练最后一层的特征,以实现图像分类。softmax 损失函数的定义如下:

$$L_S = - \sum_{i=1}^m \log \frac{e^{W_j^T x_i + b_j}}{\sum_{j=1}^n e^{W_j^T x_i + b_j}} \quad (5)$$

其中, x_i 为第 i 深度特征, W_j 为最后一个全连接层中权重的第 j 列, b 是偏置项, m, n 分别为处理的样本数量与类别数。

2.3 多任务网络结构的设计

多任务学习(MultiTask Learning, MTL)是一种归纳迁移机制,旨在通过联合学习多个相关任务的共同特性来提高泛化能力。而卡口识别引擎正好需要多个任务的识别,且各个任务之间的相关度很高,如车型识别、车辆品牌识别与车系识别。因此,多任务网络学习方法非常适用于本文的卡口识别引擎中,既能充分利用相关任务之间的相关性,使得学习到的模型精度更高,泛化能力更强,且多个任务之间权值参数共享,最大程度地提高识别效率,实现多个任务同步识别。

在卡口识别引擎中,如果利用深度学习方法将各个识别任务(颜色、车型、品牌、车系和车牌检测)都相互独立设计、运作,就会使得整体算法低效、耗时。为此,本文建立一种多层次、多任务的深度学习方法,包括车型识别、车牌检测、车身颜色识别、车辆品牌识别以及车系识别。

然而,训练一个多任务深度学习网络的过程是非常不易的,这是因为不同级别的信息有不同的学习困难和收敛速度。多任务联合训练过程如下:假设总任务数为 T ,将第 t 个任务的训练数据记为 (x_t^i, y_t^i) ,其中 $t \in (1, 2, \dots, T)$, $i \in (1, 2, \dots, N)$, N 为总训练样本数。 $x_t^i, y_t^i \in \mathbb{R}$,分别为第 i 个样本的特征向量和标注标签。那么多任务目标函数可以表示为:

$$\arg \min_{\{w^t\}_{t=1}^T} \sum_{t=1}^T \sum_{i=1}^N L(y_t^i, f(x_t^i; w^t)) + \lambda \Phi(w^t) \quad (6)$$

其中, $f(x_t^i; w^t)$ 是输入特征向量 x_t^i 和权重参数 w^t 的映射函数, $L(\cdot)$ 为式(5)所示的损失函数, $\Phi(w^t)$ 为权重参数的正则化值, λ 为正则化系数因子。

3 实验研究及分析

本文设计的多任务深度学习方法的基础为 Faster R-CNN^[18]方法,针对其不足以及结合在智能交通应用领域的实际需求,对其进行以下改进:1)提出级联网络,消除背景区域的干扰,使得学习到的特征更精确;2)设计多任务网络,使得多个任务共享学习,提高精度及效率。接下来,为了实验的严谨性和清晰性,分别单独地对两种改进方案进行实验验证。

3.1 软硬件条件

深度学习框架的选择与搭建,目前可选择的环境有 Caffe,

Torch, Theano 和 TensorFlow。本文选择了 Caffe 环境,其主要考量是 Caffe 环境与 Python 语言的绑定可进行良好的训练或调整前馈模型, Fast R-CNN 和 Faster R-CNN 内部有 Caffe, 便于配置。实验环境如表 1 所列。

表 1 实验环境

Table 1 Experimental environment	
名称	参数
系统	LINUX64 Ubuntu14.04
框架	Caffe
语言	Python, C++
CPU	Intel Core™i7-2600 3.4GHz×8
GPU	TITAN X (Pascal)
RAM	8.00GB

3.2 性能评价指标

如图 2 所示, I 表示所有标签, TP 表示该图像被正确预测到的标签集合, FP 表示预测结果中不属于该图像的标签集合, FN 表示未被预测出来的标签集合。

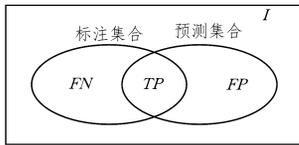


图 2 评价准则原理图

Fig. 2 Principle diagram of evaluation criteria

准确率(Precision Rate, P)、召回率(Recall Rate, R)、单个类别平均准确率(Average Precision, AP)、所有类别平均准确率(Mean Average Precision, MAP)的计算公式如下:

$$P = \frac{TP}{TP + FP} \quad (7)$$

$$R = \frac{TP}{TP + FN} \quad (8)$$

$$AP = \int_0^1 P(R) d(R) \quad (9)$$

$$MAP = \frac{1}{m} \sum_{k=1}^m (AP)_k \quad (10)$$

其中, m 为总类别数。

3.3 级联网络实验及分析

为了验证在本文方法, RPN 网络后添加一个二分类的级联网络的有效性, 选择 BIT-Vehicle^[14] 进行实验, 本节给出实验结果及分析。

BIT-Vehicle 数据库是车型识别中的常用数据库, 其中的图像都是来自于实际路段的卡口高清图像, 与实际情况非常吻合。数据集的具体内容和实验分配如表 2 所列。

表 2 实验样本数据

Table 2 Experimental data			
类别	训练	测试	合计
轿车	2961	2960	5921
SUV	695	695	1390
卡车	436	436	872
客车	279	278	557
面包车	442	441	883
微汽车	238	238	476

注: 图像尺寸为 1920 * 1080, 总数为 9850 张

的特征提取方法以及现有最先进的深度学习方法进行对比, 实验结果如表 3 所列。

表 3 BIT-Vehicle 数据集上平均准确率的对比结果

Table 3 Comparison of average accuracy on BIT-Vehicle dataset

(单位: %)

方法	客车	卡车	SUV	轿车	微汽车	面包车	MAP
SS+Hog+SVM	55.43	73.14	73.36	69.00	56.21	65.39	65.42
EB+Hog+SVM	68.86	73.76	73.56	72.01	67.31	61.00	69.41
EB+KNN	84.64	76.31	54.31	96.86	44.12	71.27	71.25
EB+Sift+Softmax	87.50	78.57	65.80	90.68	51.26	67.42	73.54
Zhen 等 ^[14] *	98.00	90.00	84.00	91.00	83.00	84.00	88.11
ZF ^[18]	90.84	90.67	89.88	90.80	90.06	88.60	90.14
VGG16 ^[18]	90.91	90.52	89.53	90.86	89.24	90.26	90.22
本文 ZF	99.72	97.38	89.20	90.81	90.33	89.82	92.88
本文 VGG16	99.96	90.35	89.67	90.87	90.08	90.20	91.86

注: * 代表此行数据引用文献[14]的结果

相比于传统特征提取方法, 首先利用 Selective Search^[15] 和 Edge Boxes^[16] 两种不同的方法测试图像中的可能性目标区域, 然后分别对可能性目标区域进行 Hog 特征提取和 Sift 特征提取, 最后使用 KNN、支持向量机 SVM、逻辑回归 softmax 分类器实现车型识别。由表 3 的第 2—5 行数据可知, 在传统的特征表达的方法中, 最高平均精度为 73%, 采用 Sift 特征能取得较好的实验结果。由表 3 的第 2—3 行数据可知, 对文中数据集进行车型检测时, Edge Boxes 优于 Selective Search, 使得检测到的可能性目标区域更接近真实标签, 从而使总体识别精度更高。而本文所提方法的精度高达 92.88%, 具有显著的优势。

相比于其他最先进的的方法, 由表 3 的前 7 行数据可以得出, 深度学习方法的识别精度明显高于传统方法, 其平均精度高达 90%。在深度学习方法中, 采用一种多任务结构的 Faster R-CNN 方法的识别结果高于文献[14]中的两阶段深度学习方法。比较表 3 中第 6—7 行数据可得出, 在该数据集中, 随着网络的加深, 精度并未有明显的增加。最后, 由表 3 中的所有数据可知, 本文在原有方法上进行了改进, 使得精度提升约 2.5%。值得注意的是, 在改进的方法中, 浅层 ZF^[9] 网络的识别精度优于深层 VGG16^[9] 网络。

BIT-Vehicle 数据集上平均速率的对比结果如表 4 所列。

表 4 BIT-Vehicle 数据集上平均速率的对比结果

Table 4 Comparison of average speed on BIT-Vehicle dataset

方法	识别平均速度/s
SS+HOG+SVM	9.617
EB+HOG+SVM	1.81
EB+KNN	4.85
EB+Sift+Softmax	1.14
Zhen 等 ^[14]	—
ZF ^[9]	0.051
VGG16 ^[9]	0.094
本文 ZF	0.085
本文 VGG16	0.134

在识别效率方面, 与传统的识别方法相比, 本文方法使速率至少提升了 12 倍。而相比于单任务的 Faster R-CNN 方法, 本文方法的识别速率有所降低。

综上, 相比于提取单一特征的传统方法, 如 Hog 特征、Sift 特征, 本文的深度特征学习具有明显的优势。使用单一

为了充分验证本文所提方法的正确性, 将其分别与传统

特征的车型识别方法存在特征表达不足的问题,这是导致识别精度低的一个非常重要的因素。而深度卷积神经网络是一种自上而下、自低向高的逐层学习策略,能够学习到车辆对象的各级特征,从而使得识别精度较高。相比于文献[14]的两阶段深度学习网络方法与未改进的深度学习方法^[9],本文方法的识别精度更佳。

3.4 多任务网络结构及分析

CompCars^[17]数据集中包含了车型、品牌、车系以及车身颜色标签,为了进一步验证本文方法中多任务网络的有效性,本文选择了该数据集中的卡口图像部分进行实验,本节给出实验验证结果及分析。

为了验证车辆的细粒度识别以及多任务网络的识别精度,需要使用相对于 BIT-Vehicle 数据集更加细分的数据集。CompCars 数据集中包含网络收集图像与卡口抓拍图像两部分。本文选择了卡口图像部分数据集(见图 3),其中总包含约 4.5 万张正面车辆卡口抓拍图像、车身颜色标签 11 类、车辆品牌 69 类、车系 281 类,该数据集适合本文的多任务卷积神经网络识别性能的验证。



图 3 CompCars 卡口数据集样例

Fig. 3 Examples of CompCars bayonet dataset

此外,考虑到该数据集的质量,对其做了如下修正:

1)原数据集中包含白天与夜晚采集的车辆图像,而对于部分晚间图像,由于光线过低,导致无法看清车型。因此,将数据集中的 HSV 图像模型中 V 分量低于 25 的图像删除,如图 3(c)所示,以提高原数据集的数据质量。

2)实验中发现原数据集中存在标签标注错误的情况,对此进行了更正。此外,将车系标签 Venucia D50 与 Venucia R50 进行了合并处理,原因是这两个系列的车辆在外观上完全相同,其区别在于发动机等内在参数配置。将颜色标签中的未识别类都按已有的颜色标签重新进行标注,最终标注颜色标签 10 类。

3)数据集中并未包含粗分类的车型标注(如轿车、SUV、面包车等),为此,逐一对该数据集添加了车型粗分类标注,主要分为轿车、SUV 和面包车 3 类,其他车型图像不存于数据集中。

最后,本文多任务网络实验的数据具体为:车型类别包括轿车、SUV 和面包车 3 类,车辆品牌类别包括常见 69 类不同品牌车辆(如:奥迪、宝马、大众、现代等),车系类别包括 280 类(如:奥迪 Q3, Q5 等),颜色类别包括 10 类。总图像数为 30938 张,训练图像为 22093 张,约占总样本的 70%,测试图像为 8845 张,约占样本的 30%。文中把修整后的数据集称为 CompCars*。

考虑到表 3 的实验结果,使用级联网络后浅层网络的结果可能更佳,因此分别对比了不同深度的网络模型实验结果。

在单任务网络模型的实验结果中,如表 5 中第 2, 4, 6 列所示,与 ZF 模型和参数网络 ResNet-50 相比,VGG16 深层网络模型在各个任务上的精度都显著提升,尤其是车型等简单任务。

在 ZF 与 VGG16 两个不同深度网络中进行多任务对比实验,如表 5 中的第 3, 5, 7 列所示,ZF 网络与 ResNet-50 的多任务模型的精度显著性下降。其原因在于,浅层网络中由单任务转变为多任务,由于多个任务之间相互学习,权重共享,使得学习约束变大,在多任务中出现欠拟合现象。在网络参数足够的 VGG16 深层网络模型中,采用多任务网络结构的精度比单任务网络更高。

综上所述,本文提出的多任务网络结构方法比最先进的现有深度学习方法有更好的识别效果。此外,采用多任务网络结构能将多个识别模型统一为一个模型,减小整体模型的复杂度,从而提升整体识别效率。例如,在单任务的 VGG16 网络中,平均每个对象的识别速率为 $v=100\pm 15$ ms,则总体时间为 $5*v$,而本文的多任务网络的整体速率为 $v'=200\pm 15$ ms,相比之下,效率提高近 1.6 倍。

表 5 在 CompCars* 数据集上多任务网络与单任务网络的平均准确率

Table 5 Average accuracy of multitask network and single task network on CompCars* dataset

(单位:%)

方法	单任务	多任务	单任务	多任务	单任务	多任务
	ZF	ZF	ResNet-50	ResNet-50	VGG16	VGG16
颜色识别	84.70	65.25	86.01	85.54	86.60	87.40
车型识别	93.88	95.47	97.27	95.10	97.78	99.07
品牌识别	99.30	96.10	99.45	99.50	99.60	99.32
车系识别	98.60	93.79	98.70	98.59	98.74	98.57
车牌检测	90.89	90.50	99.91	100.00	100.00	100.00

结束语 本文从卡口识别引擎的实际需求出发,致力于提高识别精度与效率,对现有的深度卷积神经网络进行了改进,提出了一种应用于卡口识别引擎的级联多层次、多任务的深度卷积神经网络学习方法。

该方法利用深度卷积神经网络提取图像特征,可以自学地提取出目标对象的深层特征和可区分性特征,解决了传统的人工提取特征难的问题。同时,利用深度卷积神经网络代替费时的穷举策略,来找出目标的可能性区域,加快了目标检测的精度和效率。相比于其他策略搜索方法,使用卷积操作的 RPN 网络,可使检测和识别更加高效地结合,从而进一步提高检测效率。此外,在车辆精准定位的基础上再进行车牌、车标的检测与定位,极大地提升了车牌、车标的识别精度。

本文的主要贡献如下:首先,在现有最先进的深度学习网络结构 Faster R-CNN 的基础上进行改进,在 RPN 网络后级联一个二分类识别网络,消除背景区域的干扰,使得学习特征具有区分度,进而提高识别精度;其次,进一步改进识别网络,将网络中的单任务识别网络重新设计为一种多层次、多任务的识别网络。从实验结果分析来看,本文设计的网络结构的识别精度更高,识别速度更快,避免了多个任务独立运行而引起的低效问题,同时充分利用相关任务之间的相互学习,提高深度学习网络的整体性能。

目前,作为反映车辆基本特征的车辆颜色识别主要受到环境中光的影响,导致识别准确率不高,无法满足实际需求。今后的研究工作应考虑到在同一光照条件下,车辆颜色和车牌颜色会出现相同程度的色偏。由于车牌颜色是已知的,因此可以通过检测出车牌颜色的色偏来矫正车辆颜色,借助于级联多层次、多任务的深度卷积神经网络来提升车辆颜色的识别率。此外,将考虑在同一深度卷积神经网络中集成更多的识别任务,如驾驶员的安全带检测、车辆年检标检测等任务,使其满足实际应用中多需求的目标要求。

参 考 文 献

- [1] GU R S. Research on vehicle detection and identification for video at traffic checkpoints [D]. Nanjing: Nanjing University, 2014:11. (in Chinese)
顾人舒. 交通卡口车辆检测与自动识别技术研究[D]. 南京:南京大学,2014:11.
- [2] GUPTA S, MASOUD O, MARTIN R F K, et al. Detection and classification of vehicles [J]. *Intelligent Transportation Systems*, 2002, 3(1): 37-47.
- [3] JIANG M, LI H. Vehicle classification based on hierarchical support vector machine [M]// *Computer Engineering and Networking*. Berlin: Springer International Publishing, 2014: 593-600.
- [4] LOWE D G. Distinctive Image Features from Scale-Invariant Keypoints [J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
- [5] NEGRI P, CLADY X, MILGRAM M, et al. An oriented-contour point based voting algorithm for vehicle type classification [C]// *Proceedings of IEEE International Conference on Pattern Recognition*. New York: IEEE Press, 2006: 574-577.
- [6] GE F X, SHI Y, SUN B, et al. Sparse representation based classification by using PCA-SIFT descriptors [C]// *Proceedings of IEEE International Conference on Information Science and Technology*. New York: IEEE Press, 2014: 429-432.
- [7] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]// *Proceedings of the IEEE conference on computer vision and pattern recognition*. New York: IEEE Press, 2014: 580-587.
- [8] GIRSHICK R. Fast R-CNN [C]// *Proceedings of the IEEE International Conference on Computer Vision*. New York: IEEE Press, 2015: 1440-1448.
- [9] REN S Q, HE K M, GIRSHICK R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks [C]// *Proceedings of the IEEE International Advances on Neural Information Processing Systems*. New York: IEEE Press, 2015: 91-99.
- [10] HE T, HUANG W, QIAO Y, et al. Text-Attentional convolutional neural network for scene text detection [J]. *IEEE Transactions on Image Processing*, 2016, 25(6): 2529-2541.
- [11] ZHANG Z, LUO P, CHEN C L, et al. Facial Landmark Detection by Deep Multi-task Learning [C]// *Proceedings of the IEEE International Conference on European Conference on Computer Vision*. New York: IEEE Press, 2014: 94-108.
- [12] HOOCHANG S, ROTH H R, GAO M, et al. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning [J]. *IEEE Transactions on Medical Imaging*, 2016, 35(5): 1285-1298.
- [13] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. Imagenet classification with deep convolutional neural networks [C]// *Proceedings of the IEEE International Conference on Neural Information Processing Systems*. New York: IEEE Press, 2012: 1097-1105.
- [14] ZHEN D, WU, Y W, PEI M T, et al. Vehicle Type Classification Using a Semisupervised Convolutional Neural Network [J]. *Intelligent Transportation Systems IEEE Transactions*, 2015, 16(4): 2247-2256.
- [15] UIJLINGS J R R, SANDE K E A V D, GEVERS T, et al. Selective Search for Object Recognition [J]. *International Journal of Computer Vision*, 2013, 104(2): 154-171.
- [16] DOLLAR P, ZITNICK C L. Structured forests for fast edge detection [C]// *Proceedings of the IEEE International Conference on Computer Vision*. New York: IEEE Press, 2013: 1841-1848.
- [17] YANG L, LUO P, CHEN C L, et al. A large-scale car dataset for fine-grained categorization and verification [C]// *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*. New York: IEEE Press, 2015: 3973-3981.