

车辆再识别技术综述

刘凯, 李滢东, 林伟鹏

(北京交通大学计算机与信息技术学院, 北京 100044)

摘要: 车辆再识别是指给定一张车辆图像, 找出其他摄像头拍摄的同一车辆, 可将车辆再识别问题看作图像检索的子问题。在真实交通监控系统中, 车辆再识别可以起到对目标车辆进行定位、监管、刑侦的作用。随着深度神经网络的兴起和大型数据集的提出, 提升车辆再识别的准确度成为近年来计算机视觉和多媒体领域的研究热点。从不同角度对车辆再识别方法进行了分类, 并从特征提取、方法设计和性能表现等方面对车辆再识别技术进行了概述、比较和分析, 对车辆再识别技术面临的挑战及发展趋势进行了预测。

关键词: 车辆再识别; 深度学习; 特征表达; 度量学习

中图分类号: TP520

文献标识码: A

doi: 10.11959/j.issn.2096-6652.202002

A survey on vehicle re-identification

LIU Kai, LI Yidong, LIN Weipeng

School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

Abstract: Given a vehicle image, vehicle re-identification aims to find the same vehicle caught by other cameras, it can be regarded as a sub-problem of image retrieval. In the real traffic surveillance system, vehicle re-identification can play a role in locating, supervising and criminal investigation of target vehicles. With the rise of deep neural networks and the release of large-scale dataset, improving the accuracy and efficiency of vehicle re-identification has become a research focus in the field of computer vision and multimedia in recent years. The vehicle re-identification methods from different perspectives were classified, and the overview, comparison and analysis in terms of feature extraction, design and performance were given in detail, and the challenges and future trends of vehicle re-identification were predicted.

Key words: vehicle re-identification, deep learning, feature representation, metric learning

1 引言

车辆再识别问题是指在一个特定范围内的交通监控场景下, 判断非重叠区域内拍摄的车辆图像是否属于同一辆车的检索问题。目前, 国内外在交通发达的区域都安装了监控摄像头, 如何更好地利用监控摄像头进行交通监管和刑侦, 是一个非常值得研究的问题。近年来, 计算机视觉和多媒体领域的研究人员主要对车辆检测^[1]、细粒度的车辆分类^[2]、车辆跟踪^[3]、驾驶员行为建模^[4]等方面进行研究,

车辆再识别技术并未得到广泛关注。

随着深度神经网络在计算机视觉和多媒体领域的发展, 车辆再识别技术引起了学术界和工业界的广泛关注, 并成为深度学习领域的热点研究问题。与以往的近邻重复图像检索 (near-duplicate image retrieval, NDIR) 问题^[5]不同的是, 在车辆再识别任务中, 摄像头的位置不同会产生光照变化、视角变化及分辨率的差异, 这导致同一车辆在不同视角下产生自身差别 (intra-class difference) 或不同车辆因型号相同形成类间相似 (inter-class similar-

收稿日期: 2020-01-02; 修回日期: 2020-02-18

通信作者: 李滢东, ydli@bjtu.edu.cn

基金项目: 国家自然科学基金资助项目 (No.61672088, No.61790575)

Foundation Items: The National Natural Science Foundation of China (No.61672088, No.61790575)

ity),这就使得车辆再识别任务的研究人员面临巨大的挑战。

车辆再识别任务按照发展的历程可分为两个阶段。第一个阶段是在 Liu 等^[6]提出 VeRi 数据集之前,这一时期的车辆再识别任务缺乏大规模高质量的数据集,同时学术界又没有对车辆再识别任务的数据集和评估指标进行严格定义,所以大部分数据集不满足再识别任务的要求,同时期主要以手工特征为主,以各类探测器^[7]为辅助设备完成车辆再识别任务。第二个阶段是深度学习在计算机学术界兴起之后,许多研究人员和机构发布了大型车辆再识别数据集(如 VeRi^[6]、VehicleID^[8]等),并对车辆再识别任务的评价指标进行了统一,因此大量的车辆再识别方法被提出,大部分方法都基于深度学习的流行框架,并在准确度和效率上都远超使用传统手工特征的方法。

目前国内外关于车辆再识别问题的研究非常多,但相关的综述文章却几乎没有。Khan 等^[9]于 2019 年在 *Computer Vision and Image Understanding* 上发表了关于车辆再识别的综述性论文,这是当时车辆再识别领域唯一的一篇综述性文献。但是,该文献存在以下局限性:一是没有给出车辆再识别问题的具体描述,其中提到的许多文献中的方法都不属于再识别的范畴;二是该文献提到的大多数工作都是基于传感器的车辆识别方法进行的,缺乏对基于二维彩色图像和深度图像特征的方法的总结和比较;三是该文献对方法的分类方式有一定的局限性,仅简单地分为基于传感器的方法和基于视觉的方法。

本文对车辆再识别的问题边界、研究挑战和处理流程进行了梳理,对车辆再识别技术的研究进展进行了综述。第 2 节给出车辆再识别的问题描述,并介绍解决方案的一般流程。第 3 节对现有的车辆再识别研究进行分类,对传统的车辆再识别方法进

行简短地总结与回顾,从解决问题的着手点出发,详细介绍基于深度学习的车辆再识别方法。第 4 节介绍具有代表性的车辆再识别数据集,并对取得优异性能的方法进行比较和分析。第 5 节对车辆再识别技术面临的挑战和研究趋势进行总结和展望。

2 问题描述

2.1 车辆再识别的问题来源

虽然车牌识别是区分不同车辆最简单直接的方法,但是在很多情况下仅依靠车牌信息无法完成再识别任务。

首先,交通监控系统存在拍摄视角变化、天气影响、光照变化和图像分辨率低等^[10]问题,从而导致车牌像素模糊、车牌信息不完整,这会在很大程度上降低车牌识别的准确率。其次,在一些交通场景下,车牌有可能被遮挡、移除甚至伪造,这极大地提升了寻找目标车辆的难度。因此,在刑侦、安防等场景中,以车辆的外观信息为基础进行车辆身份的检索和匹配成为一个具体任务,并且具有重大的研究意义和良好的应用前景。

2.2 车辆再识别的问题描述

不同于车辆检测、跟踪或分类等问题,可将大型交通监控场景下的车辆再识别问题看作一个近邻重复图像检索问题^[5]。具体来说,给定目标车辆在特定区域内的一张图像,车辆再识别的任务是找到目标车辆在其他摄像头下被拍摄到的图像,期间只使用车辆的外观信息和辅助信息(如车辆型号、拍摄时间、地点信息等)。此外,检索图像的视角、拍摄时间、天气等客观因素都不受限制(可以和给定图像不同)。

典型的车辆再识别流程包括车辆图像或视频的获取、车辆检测、特征提取与表达、相似性度量

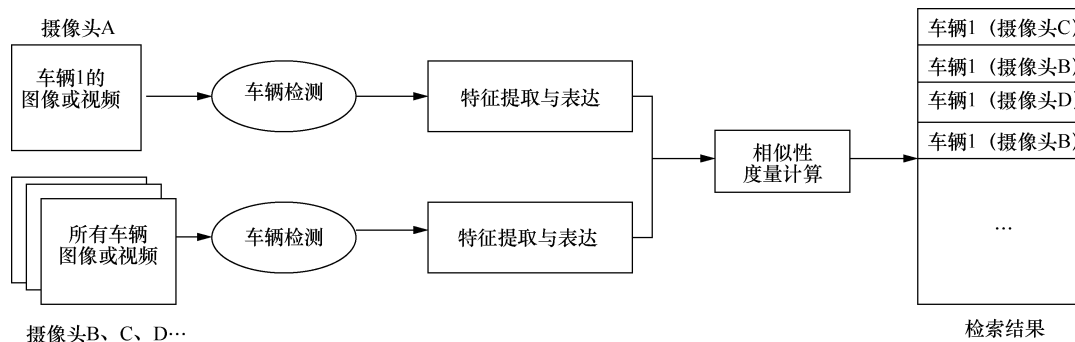


图1 典型的车辆再识别流程

计算和检测结果的展示,如图1所示。现有的车辆再识别方法一般默认车辆图像的获取和车辆检测已经完成,由数据集直接提供清晰完整的车辆图像。

具体地,数据集一般分为训练集和测试集,训练集用于模型的训练和学习,测试集又进一步分为查询集(probe set)和图库集(gallery set)。在测试方法性能时,将所有测试集的图像输入模型,并进行特征提取,以某种度量方式(如欧氏距离)计算查询集的每一张图像与所有图库集的图像的特征距离,距离越小表示车辆之间的相似程度越高,按距离从小到大依次排列图库集中的图像序列,得到当前查询集图像的再识别检索结果,具体如图2所示。

3 车辆再识别方法分类

20世纪90年代以来,大量基于传感器或人工设计的车辆再识别方法被提出,这些方法设计复杂,需要的硬件设备较多,同时没有统一的评价标准和数据集,因此很少有研究者对其进行整理和综述。近十几年来,监控摄像头被广泛安装在城市、郊区以及高速公路上,大量的车辆监控图像被实时采集并存储,对不同区域出现的目标车辆进行检索成为现实需求,因此广义上的车辆再识别方法主要是指基于二维彩色图像的车辆再识别方法。下面对国内外主要的车辆再识别方法进行分类综述。

表1为现有的车辆再识别方法的总体分类情况。根据实现方式的不同,将这些方法大致分为基



图2 车辆再识别的检索过程

表1 车辆再识别方法的分类情况

分类依据	方法类别	参考文献	优点	缺点
实现方式	基于传感器的方法	参考文献[11-21]等	不需要训练学习	复现难度高; 识别率低; 需要大量硬件
	基于人工设计特征的方法	参考文献[22-33]等	不依赖特殊硬件; 可解释性强	时间复杂度高; 受光照变化、视角变化和遮挡的影响较大
	基于深度学习的方法	参考文献[34-68]等	识别率高; 实时性强; 可以解决光照变化、视角变化和遮挡问题	训练学习时间长; 特征抽象, 可解释性差
解决车辆再识别问题的着手点	利用多维度信息的方法	参考文献[39-48]等	对车辆特殊外观敏感	容易误判辨识度区域; 易受视角、光照变化影响
	基于度量学习的方法	参考文献[49-54]等	识别率高; 对难样本的识别效果好	训练学习时间长
	解决视角变化引起的外观偏差	参考文献[55-58]等	对车辆的视角变化有鲁棒性; 识别率高	训练学习时间长; 系统实时性难以得到保证
	利用时序地理信息的方法	参考文献[59]等	对难样本的识别效果好; 方法新颖	需要车辆时序和位置标签; 对数据集要求高
是否需要车辆身份标签	有监督学习方法	参考文献[34-65]等	识别率高; 符合发展潮流; 在不同数据集下的泛化能力强	训练学习时间长; 需要大型数据集支持
	无监督学习方法	参考文献[66-68]等	无需人工标注; 符合实际需求; 拓展性强	识别率较低; 时间复杂度高

于传感器的方法、基于人工设计特征的方法、基于深度学习的方法三大类。基于传感器的方法是最早针对车辆身份进行识别的方法,这类方法一般需要搭建传感器或感应器,并将其作为信息输入来源,虽然方法简单,但是容易受信号强弱、天气情况和交通状况等客观因素影响。基于人工设计特征的方法无需多余的硬件设备,仅需要车辆的二维彩色图像便可正常工作,此类方法大多采用计算机视觉领域里一些有效的特征设计手段来对车辆图像进行表达,识别率较低但特征的可解释性强。基于深度学习的方法能较好地克服视角变化和光照变化带来的影响,识别率非常高,是进行车辆再识别研究的比较理想的选择。

根据解决车辆再识别问题时选择的着手点,还可以将车辆再识别方法分为利用多维度信息的方法、基于度量学习的方法、解决视角变化偏差的方法和利用时序地理信息的方法,当前大多数车辆再识别方法处理问题的思路都可以归为这4类。另外,根据是否需要车辆身份标签又可将车辆再识别方法分为有监督学习方法和无监督学习方法。

实际上,一个好的车辆再识别方法往往综合利用不同计算机视觉任务中的思想,因此难以从单个角度去严格划分。为了更好地对现有车辆再识别方法进行梳理和总结,接下来以方法的实现方式为主线,对车辆再识别方法进行综述,同时以解决问题为着手点,对流行方法进行详细介绍。

3.1 基于传感器的方法

使用各种传感器(或感应器)对车辆进行探测并推断身份是最基本的、最早出现的车辆再识别方法。但是,正如引言部分提到的,这些方法在提出时没有合格的数据集和明确的问题定义,因此大部分方法都不符合严格意义下的车辆“再”识别,故只做简单介绍。

对于每个被传感器探测到的车辆来说,这类方法通常利用特定的手段提取车辆特征,通过特征匹配来判断车辆身份。最早出现的车辆再识别方法^[11-12]基于各种硬件探测器(如红外线、超声波等)提取车辆的特征信息。之后,许多利用传感器或感应器的车辆识别方法被提出,其中, Sanchez 等^[13]提出了使用无线磁感应器来探测车辆身份的方法;参考文献[14]设计了一种三维磁感应器来探测车辆的多维特征,并能够从感应器中获取时间信息,用于训练高斯极大似然分类器。参考文献[15]使用无

线磁传感器提取车辆特征信息,用于实时预测车辆在多个路口的行驶时间。

感应线圈是在交通场景中最常用的获取数据的工具,可以监测多种车辆属性(如速度、体积和车辆占地面积),城市的主干道路和高速公路上一般都部署了感应线圈。参考文献[16]提出了一种基于感应线圈的实时车辆再识别方法 RTREID-2M; Ndoye 等^[17]利用感应线圈提供的数据提取车辆特征,并估计车辆行程时间,从而完成车辆再识别任务。Ndoye 等^[18]还提出了一种改进车辆再识别算法的信号处理框架; Ali 等^[19]提出了一种用于车辆再识别和变道监控的多感应线圈检测系统。感应线圈对车辆速度变化敏感,但是以上方法在实现时都假设车辆是匀速行驶的,这显然不符合现实情况。

随着一些新兴传感器或技术(如全球定位系统(GPS)、无线射频识别(RFID)和手机)的面世,一些方法探索了基于信标的车辆跟踪和监视系统,以完成车辆再识别任务。Prinsloo 等^[20]提出了一种基于无线射频识别标签的车辆再识别算法,适用于各类收费站; Mazloumi 等^[21]提出了一种基于 GPS 的车辆行驶时间估计方法,用于解决车辆再识别问题。

基于传感器的方法大多需要安装大量的硬件设备,实验环境较为苛刻,难以复现。此外,许多方法易受客观环境影响,如天气情况、信号强弱、交通拥挤程度和车辆行驶速度等,这些都将不同程度地降低传感器的灵敏度。同一时期也没有统一的性能评估标准,因此这类方法不能算是理想的车辆再识别方法。

3.2 基于人工设计特征的方法

3.2.1 基于 3D 建模的方法

研究人员就如何提取具有身份辨识度的描述符进行了许多探索,其中基于车辆 3D 建模的方法被证明是一种有效的方法。车辆图像具有许多诸如厂商标志、车型和零件的信息,研究人员从不同的角度对车辆特征进行人工提取,并进行 3D 建模,如 Zhang 等^[22]使用基于 3D 模型的匹配方法对同一型号的车辆进行识别,但在识别粒度上属于车辆型号识别; Woessler 等^[23]提取了车辆顶面的三维模型和颜色信息,以此为描述符完成了车辆再识别。Guo 等^[24]通过 3D 建模的方法估算了参考车辆和目标车辆的姿态和外观,在标准化 3D 空间中提取车辆

特征,对车辆进行几何不变的比较,以完成再识别任务。

Zapletal 等^[25]提出使用颜色直方图和方向梯度直方图的线性回归方法对车辆进行再识别,采取参考文献[26]和参考文献[27]中的方法对图像中的车辆建立 3D 边界框,并以边界框为参照空间,把车身侧面和前面的图像进行拉伸和拼接,形成一个信噪比非常高的合成图像。在此基础上将合成图像划分成若干个网格,分别计算出网格内的颜色分布直方图和方向梯度值,然后将所有网格特征进行拼接,形成两个高维特征向量,最后取均值得到平均特征向量。在识别过程中,线性回归器对两幅图像的平均特征向量进行回归验证,以完成车辆再识别。

总的来说,基于 3D 建模的方法可以提取多视角下的车辆特征,在外观信息获取的维度上优于其他人工设计的方法,但时间复杂度高、实现效率和识别精度较低,无法适用于海量车辆图像场景。

3.2.2 基于手工特征的方法

车辆图像本身存在的一些属性信息(例如颜色和纹理)也可以用于识别车辆身份,最早的一些基于手工特征的方法便是利用不同的方法获得图像的属性,以进行身份判别。参考文献[28]利用包含相同车辆和不同车辆的图像特征集来训练身份分类器,这些图像特征包括车辆位置、边缘对比度等;Feris 等^[1]提出了一种由多种特征描述符组成的大规模特征池来对车辆图像进行表达,并基于此构建了一个检索系统,依靠特征池进行车辆型号检索。参考文献[29]提出了一种无监督车辆匹配算法,该算法将车辆再识别定义为二分类问题,通过计算两个非重叠区域下的车辆是同一身份的概率,对车辆进行匹配。Zheng 等^[30]提出了从大型图像数据库中搜索车辆的匹配算法和重排序方法。

Liu 等^[6]提出了基于纹理特征的 BOW-SIFT 方法和基于颜色特征的 BOW-CN 方法。BOW-SIFT 方法利用 SIFT 算子^[31]提取车辆的局部纹理特征,使用 BOW 模型对 SIFT 算子进行字典编码以得到纹理特征向量,其用来表示车辆的纹理特征。BOW-CN 方法则采用 Color Name (CN)^[32]作为局部颜色描述子,类似于 BOW-SIFT 方法,将图像分割成大小一致的块,在每一块中计算出 CN 算子,使用 BOW 模型对 CN 算子进行字典编码,最终得到颜色特征向量。参考文献[16]中使用了行人再识

别中对光照变化具有鲁棒性的手工特征局部最高出现率表达(local maximal occurrence representation, LOMO)法进行车辆再识别,按照参考文献[33]的描述,每一张车辆图像经过 LOMO 特征表达形成特征向量,特征向量用于计算车辆之间的身份相似度从而完成再识别任务。上述几种方法都以特征之间的距离表示图像之间的相似性,距离越短,两车身份相同的概率越高。

总的来说,基于手工设计特征的车辆再识别方法在早期取得了一定的进展,积累了许多成功的经验。但是由于车辆颜色和型号有限,许多车辆(尤其是同款车型)过于相似,因此这类方法的性能容易受到限制。受视角固定、分辨率低、光照情况不理想等客观因素的影响,也很难从手工特征中获得车辆的辨识性信息。

此外,这类再识别方法的性能往往取决于手工特征设计者的先验知识和调参能力。由于参数个数受限,一个有效的手工特征往往需要很长的时间才能产生,且泛化能力较弱。

3.3 基于深度学习的方法

在卷积神经网络(convolutional neural network, CNN)被提出以后,深度学习在计算机视觉领域开始流行,它弥补了传统手工特征表达能力不足的缺点,同时也更适合图像识别和检索任务。目前车辆再识别领域内性能最好的模型都是基于深度学习的方法创建的,故本文着重总结和分析这一类方法。

在学习过程中,此类方法通常以构建一个具备辨识车辆身份能力的神经网络为主要目的。经过训练,网络能将车辆图像投射到特征空间中,并以特征空间中的度量相似性(如欧氏距离)表示车辆之间的相似性,特征距离小的车辆图像中的车辆将被看作同一辆车。

基于深度学习的车辆再识别方法有两个核心任务。一是网络结构的选择。不同的研究人员对车辆外观信息的理解和对辅助信息的利用有所差别,因此会选择不同的网络结构对车辆特征进行提取。二是损失函数的设计。在优化网络的过程中,能否设计出合理有效的损失函数决定了最终性能表现的好坏。损失函数不仅影响神经网络的优化过程,而且决定了大型数据集能否被完整和有效地利用。简单来说,对网络结构和损失函数的设计是目前基于深度学习的车辆再识别任务中的研究重点。

3.3.1 利用车辆多维度信息的方法

与行人再识别^[34-38]技术相比, 车辆再识别更具挑战性。首先, 大型车辆再识别数据集的获取和标注难度较大, 一些数据集不仅对车辆身份进行了标注, 还对车型、颜色等车辆属性进行了记录; 其次, 同车型的车辆(如奥迪 A6L)具有相同(或相似)的外观, 如果没有车牌信息的帮助, 即使用人眼去区分不同身份的车辆也相当困难。挖掘车辆的辨识性特征成为解决问题的关键, 车身的特殊外观和数据集提供的属性标签都能用于提升神经网络的敏感性, 车身贴图、保险杠、备胎、车内装饰甚至刮擦痕迹都是非常明显的辨识性外观信息。精准挖掘每个车辆实例特有的外观信息, 并利用多维度信息对神经网络的训练过程进行监督, 是提升网络辨识能力的一种方法。

参考文献[6]提出了将手工特征和深度学习特征融合的属性和颜色特征融合(fusion of attributes and color features, FACT)方法, 首先分别对测试集图像进行 BOW-SIFT 纹理特征、BOW-CN 颜色特征和 GoogleNet 语义特征的提取, 然后分别计算这 3 种粒度下的特征之间的相似性得分, 并按照 0.1、0.2、0.7 的权重进行加权求和, 得到车辆图像之间的相似性得分, 以此为准进行车辆再识别排序。

随后 Liu 等人^[39]又提出了先进的车辆再识别(progressive vehicle re-identification, PROVID)方法, 在 FACT 方法的基础上加入 NFST (null Foley-Sammon transform)^[40]形成了 NuFACT 方法, 利用车牌信息和时间-地理信息对结果进行重排。NFST 方法最初用于解决人脸识别里的小样本尺寸问题, Zhang 等^[41]在行人再识别研究中提出了 Kernelized NFST, 把多重特征映射到一个具有区分度的零空间中, 该方法在行人再识别研究中获得了不错的结果。因此参考文献[39]提出将 FACT 方法中使用的 BOW-SIFT 特征、BOW-CN 特征和 GoogLeNet 语义特征拼接成一个外观特征, 使用 Kernelized NFST 将外观特征嵌入零空间并进行相似性计算, 同样也得到了理想的车辆再识别性能。除此之外, 参考文献[39]和参考文献[40]还在外观相似性计算后对结果进行了基于车牌和时间-地理信息重排, 对不同维度的信息进行了分析和利用, 是一个比较综合的车辆再识别方法。

参考文献[42]提出了一种多任务深度学习框架, 基于多维度信息同时完成车辆分类和相似性排

序, 以达到识别同一车辆的目的。在车辆分类中利用车辆身份和其他车辆属性(如型号、颜色)对车辆进行分类。在相似性排序中, 该方法提出将车辆的二元关系(是否同一身份)进一步细分为多种维度, 依次为: 同一车辆、同一模型同一颜色、同一模型不同颜色、不同模型不同颜色。基于这种多维度关系, 该方法提出了两种排序方法: 嵌入多粒度约束, 使成对排序从二元关系转为多元关系; 采用似然损失函数对不同粒度的图像列表进行排序。实验结果表明, 该方法在提供车型和颜色信息的数据集上取得了不错的结果。

为了扩大同车型实例之间的细微差异, 研究者们也采取了感兴趣区域(region of interest, ROI)预测或注意力模型的设计思路。He 等^[43]提出了一种简单有效的部分正则化方法, 利用预训练的 YOLO^[44]探测器检测感兴趣区域, 将零件级(车窗、车灯、厂牌别)的多维度约束引入典型的车辆再识别框架中, 以提升局部特征在学习过程中的影响力, 同时也能增强神经网络对细微差异的感知能力。利用多维度信息来扩大局部特征影响力的思想在车辆再识别任务中至关重要。此外, 同车型车辆之间的细微差异是车辆再识别性能瓶颈的重要突破口, 这一点不应被忽略。Zhang 等提出了利用局部区域指导的注意力模型——局部区域指导的注意力网络(part-guided attention network, PGAN), 先利用目标检测模型提取车辆图像的关键部位区域(厂牌、车灯、年检标识、个性装饰等), 并将其作为网络学习过程中的候选搜索区域。PGAN 包含一个局部注意力模块, 其目的是通过学习候选搜索区域的权重来挖掘最重要的局部区域, 如此便可以在学习中强调车辆图像中最具辨识性的部分, 同时降低无效的局部区域对再识别性能的负面影响。通过对整体特征和局部指导特征的联合优化, PGAN 能够得到非常出色的再识别结果。

当现有数据集能提供的多维度信息有限时, 对已有数据集进行额外标注或提出携带多维度信息的数据集是另一种解决方案。

Wang 等利用无人机收集了全新的空景车辆再识别数据集, 除了车辆 ID 以外, 还提供了颜色、车型、特殊外观(保险杠、备胎、天窗等)的标注信息, 并基于这些丰富的多维度标注信息提出了车辆再识别方法。该方法分为 3 个部分: 首先是一个多目标分类器, 该分类器利用深度学习对车辆

图像进行特征提取,并以上述标注信息为监督信号分别进行 ID 分类、颜色分类和型号分类任务的训练;其次是一个利用 YOLOv2^[45]的特殊区域检测器,用来检测车辆的特殊辨识区域;最后将分类器网络得到的特征与检测器得到的权重热点图进行权重池化,得到加权特征,使用三元组损失函数对整个网络进行训练。这种方法最大程度地利用了多维度信息,利用现有的损失函数和流行的神经网络即可得到十分可靠的再识别结果。

Tang 等^[46]利用高度随机化生成数据模型提出了新的合成车辆数据集 CityFlow-ReID,标注信息包括关键点、车身朝向、车身颜色、车型,同时还提出了一个基于已知的车身姿态的多任务再识别 (pose-aware multi-task re-identification, PAMTRI) 方法。该方法主要包括车辆热点图的获取、车身姿态预测和多任务学习 3 个步骤,在训练过程中, PAMTRI 同时利用真实数据和生成数据,从预训练的图像分割网络中得到车辆热点图,并将其附在图像原始的 RGB 三通道后,作为姿态预测和多任务学习的输入。姿态预测和多任务学习分别使用 HRNet^[47]和 DenseNet^[48]作为基础网络,由 HRNet 得到的关键点坐标和确信得分将和 DenseNet 输出的深度学习特征拼接,形成最终用于再识别和分类任务的特征向量。

利用多维度信息的车辆再识别方法适用于中小型数据集,结合额外的标注信息或采取注意力机制,使网络关注车身携带的辨识性特征,即可实现较高的识别准确率,并且能明显简化网络的学习过程。

3.3.2 基于度量学习的方法

度量学习 (metric learning) 是图像检索、车辆 (行人) 再识别、人脸识别等计算机视觉任务中常用的一种方法。通常来说,度量学习旨在学习一个特征空间,在空间中来自同一类别的样本距离相近,不同类别的样本距离较远。车辆再识别技术利用特征之间的相似性来估计车辆是同一身份的概率,选择适当的度量学习方法能够有效提升特征的区分性和表达性,使模型更容易辨别车辆身份。

Liu 等^[8]提出的深度相关性距离学习 (deep relative distance learning, DRDL) 方法是首个基于度量学习的车辆再识别方法。作者提出查询图像与候选图像之间有两种区分粒度:同一车型和同一 ID。DRDL 方法背后的思想是:若两个外观相似的车

辆不是同一型号的,那么它们一定不是同一 ID;若是同一型号的,则再利用外观信息进行细粒度区分。

在 DRDL 中,作者在 VGG 网络的基础上设计了一个混合差分网络,将单分支网络扩展成双分支网络,使用车辆身份和车辆型号对神经网络的训练进行监督,使得网络同时具备区分车辆型号和 ID 的能力。传统三元组损失 (triplet loss) 函数中的移动方向错误问题如图 3 所示,对传统三元组损失函数的缺陷和车辆再识别的特性进行分析后,作者提出了成对簇损失 (coupled clusters loss) 函数。和三元组损失相比,该损失主要的进步在于:在优化过程中正样本点总是向着聚类中心点的方向移动,而不是向着随机的正样本点移动;在计算损失函数时每次接收的样本是一个训练批次中的所有样本,而非仅仅 3 个样本。因此每一次反向传播对网络的优化都更加合理,也使得自身更容易收敛,解决了传统三元组损失函数出现的局部优化和原点选择敏感的问题。

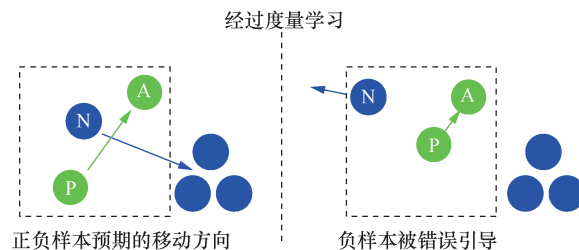


图 3 传统三元组损失函数中的移动方向错误问题

三元组损失函数被大量用于度量学习任务,旨在拉近每个三元组中原点 A 和正样本 P 的距离,并尽可能拉开原点 A 与负样本 N 的距离,但其存在不少缺点。

首先,在一次优化过程中,传统的三元组损失函数仅关注 3 张图像,无法有效地利用大规模数据集进行度量学习。其次,三元组损失函数容易出现移动方向错误的情况,如图 3 所示,在一个三元组的样本进行度量学习时,为了和原点 A 保持距离,负样本 N 被推向了远离自身所属集合的方向,更为严重的是,负样本 N 有可能不会出现在其他三元组中,导致错误无法被修正,因此损失函数的收敛速度变慢。而这种情况越多,性能折损就越严重。最后,在训练集中,由于三元组都是随机采样形成的,采样过程不受限制,导致部分三元组本身就满足或很轻易就满足损失函数的要求,因此这部分样本在

网络训练过程中不进行反向传播, 在训练过程中几乎不起作用。

为了避免以上情况发生, 参考文献[49]提出了新的三元组采样方法: 由于移动方向错误的情况只会出现三元组的负样本 N 上, 因此需要保证负样本 N 出现在后续的三元组中, 并作为原点 A 或者正样本 P , 以构成新的三元组, 迫使该负样本 N 在新的三元关系中以新的身份继续移动, 这样即可减少方向移动错误带来的负面影响; 每个训练批次中的原点 A 和正样本 P 成对出现时, 规定二者在后续的三元组中再次出现并交换身份, 使同类样本在特征空间中进一步接近, 从而避免某些难样本偏离类中心; 在进行采样时, 对最终提取的特征向量进行采样, 而非对数据集的图像进行采样, 避免图像重复经过网络。

在度量学习过程中, 再识别模型需要对图像在特征空间中的位置十分敏感, 这种能力需要网络逐渐去学习和适应, 参考文献[50]提到了一种由粗糙至精细 (coarse-to-fine, C2F) 的结构化特征嵌入方法, 在训练中, 让网络从简单粗略的分类任务开始学习, 逐步提升难度, 使其最终具备准确的身份辨别能力。其核心是一个由粗略至具体的排序损失表达式, 其中包括分类损失项、粗粒度损失项、细粒度损失项和成对损失项。分类损失项采用交叉熵损失, 以车辆型号为标签对网络进行监督训练, 使网络学会对车辆的型号进行分类。粗粒度损失项和细粒度损失项均采用三元组损失, 前者旨在拉近特征空间中同型号车辆的距离, 拉开不同型号车辆的距离, 即进一步区分型号; 后者拉近 ID 相同车辆的距离, 并拉开 ID 不同车辆的距离。成对损失函数项是指使 ID 相同的车辆在特征空间中进一步靠近, 甚至重合。利用上述 4 项损失函数, 该模型能够逐步使神经网络具备类内紧凑、类间分明的特征嵌入能力。

为了解决视角变化引起的类内差异和类间相似的问题, Bai 等^[51]设计了一个组内敏感的三元组嵌入 (group-sensitive triplet embedding) 方法, 使用端到端的方式进行度量学习。在训练中, 将同 ID 车辆的图像分为一类, 并求出类中心点; 通过 K-means 算法把图像聚类为若干组, 并求出组中心点。通过类内偏差损失 (intra-class variance loss) 函数将同 ID 的车辆图像在特征空间中向类中心点移动, 在组内将同一组的图像向组中心点移动, 同

时类与类之间、组与组之间的难样本保持一定距离。完成移动后使用 K-means 重新分组, 重复以上步骤直到损失函数最终收敛。

在度量学习中, 多任务学习的思想也被用于提升再识别的性能, 由于传统三元组损失对原点 A 、正样本 P 和负样本 N 之间的距离关系限制过于简单, 使得损失函数轻易被满足, 故一般会在网络中添加一个分类任务^[51-53], 使用 Softmax 损失和三元组损失一起对网络进行优化, 让不同的损失函数互相约束, 使得度量学习过程更具鲁棒性, 这也成为后续基于度量学习的模型都采取的做法。

受行人再识别^[52-53]中对特征图进行水平分割的训练策略的启发, Chen 等^[54]设计了一种先分割后聚合的分割重连网络 (partition and reunion network, PRN) 模型, 对车辆图像的神经网络特征图在高度、宽度、通道上分别进行了四等分割, 并将最后的全连接层输出的子特征向量进行拼接, 作为交叉熵损失函数的输入。对全局特征也施加了一个三元组损失函数, 使用 ID 标签作为监督信号训练每一个被切割的网络分支。通过这种先分割再聚合的训练策略, 该模型不使用任何额外的标注信息 (除 ID 以外) 就可以捕捉到丰富的车辆局部特征, 并对全局特征进行细节上的补充。

总的来说, 基于度量学习的方法大多都比较简单, 训练时间较短, 且解释性较强。一般采用的是三元组损失函数和交叉熵损失函数, 其再识别性能较好, 是目前车辆再识别领域的研究热点, 但是需要抽象地理解车辆图像在特征空间中的关系, 并在神经网络的训练过程中运用一些技巧。

3.3.3 解决跨视角偏差的方法

在现实世界的交通监控系统中, 不同位置的摄像头视角不同, 拍摄到的车身方向的不同 (如图 4 (a) 所示) 导致同一车辆的外观具有很大的差异, 这给车辆再识别任务带来了巨大的挑战。由视角变化引起的再识别偏差问题, 可进一步细分为同一车辆因视角变化而产生的类内偏差和同一视角下不同车辆因型号相同形成的类间相似, 如图 4 (b) 所示。在深度学习中解决这一问题的途径主要有两种, 一是借助额外的辅助信息, 例如对某一视角的车辆外观特征进行推测、根据车身关键点预测当前视角等, 以获取已知视角下的车辆特征; 二是设计出合理的网络结构和损失函数, 使视角变化引起的偏差对再识别任务的影响尽可能小, 让 ID 相同的

车辆在特征空间中更加紧凑。

Wang等^[10]将车身表面具有辨识性的20个特殊位置做标记并分为前、后、左、右4类,利用堆叠式沙漏网络模型^[55]对20个关键点进行预测,形成关键点映射图,按照其所属面进行累加,形成4个面的关键区域映射图。4个关键区域映射图和全局特征图共同经过卷积层、池化层、全连接层后得到用于相似性判断的特征向量。



图4 因视角变化而产生的类内偏差和类间相似

同样地, Zhou等^[56]提出了视角已知的注意力推断模型,将车身外观分为5类:前面、前侧面、侧面、后侧面、后面。根据生成对抗网络(generative adversarial network, GAN)的思想,设计了一个基于多视角生成网络来生成车辆各种视角的特征向量,从单视角特征学习生成多视角特征,以实现全方位的外观特征对比。由于生成器输入了车辆的全视角外观信息和图像显著区域信息,这种方法在预测其他视角的外观时更具合理性和鲁棒性。

利用已知视角下的车身关键点信息预测车辆在其他视角下的外观信息,虽然解决了部分由于视角缺失引起的再识别偏差问题,但是相同型号的车辆在单一视角下的关键点信息十分相似,若以相似的关键点信息来预测其他面的外观特征,得到的结果可能完全一致,这与解决视角变化问题的初衷有所出入。此外,并不是所有的关键点都能够提供具有辨识性的信息,每个关键点的贡献还是取决于车辆在图像中的方向。为解决这些问题, Khorramshahi等提出了一种注意力网络,其特点是能够根据车辆的方向自适应地选择对再识别任务有贡献的关键点,进行局部辨识性特征提取,从而为全局特征提供互补信息。该网络的结构有两个分支,一个是整体外观学习网络,用于捕捉车辆的全局外观特征;另一个是已知视角下的区域外观学习网络,通过学

习将注意力集中在最具辨识性的关键点,从而捕捉到车辆最具辨识性的区域,最终得到有效的局部特征。在实现以上方法的同时,作者还设计了关键点检测算法和车辆朝向预测模型,这两种算法对于解决车辆再识别中的视角变化引起的偏差都有十分明显的正面作用。

视角信息的生成或推测都需要进行大量的深度学习过程,因此在算力资源和时间资源有限的情况下很难达到好的识别效果,也难以满足真实交通场景精准、可靠、实时识别的需求。

通过预训练模型来获取视角信息是一种解决此类需求的方式。参考文献[57]通过预训练的视角分类器来判别两幅车辆图像是否处于同一视角,在视角已知的条件下将车辆图像在特征空间角度划分为两种情况:相同视角和不同视角。在两种情况下分别进行度量学习,形成S空间(同视角)和D空间(不同视角)两种特征空间,使用两种约束同时优化特征空间:空间内约束使S空间和D空间内的同ID样本更近;跨空间约束拉近D空间下的正样本对,并拉开S空间下的负样本对。对比实验证明,两种约束对提升车辆再识别性能都有明显的贡献。

另一种方式则是采用聚类思想的无监督学习获取视角信息。Lin等^[58]提出了多视角学习(multi-view ranking learning, MRL)方法,采用K-means聚类算法获得车辆的视角标签,并将车辆再识别建模成两个子任务:同视角和跨视角下的车辆再识别。在同视角下,让相同车辆在特征空间中靠近视角中心点,并推远负样本车辆;在跨视角下,让同一车辆簇内的样本在特征空间中靠近,并在不同车辆簇之间保证一定的间隔。该方法的优势在于不需要使用额外的标注信息,即可减少视角变化对车辆再识别任务的消极影响。

总的来说,视角信息对于解决车辆外观的跨域偏差能起到关键性的指导作用,无论是通过生成对抗模型推测其他视角特征,还是使用预训练模型和聚类算法获取视角信息,深度学习模型都能通过训练提升对车辆身份的辨识性和敏感性,此类方法是基于车辆再识别任务的特点(即视角变化丰富)进行设计的,都取得了一定的成功。

3.3.4 利用时间-地理信息的方法

目前以车身外观为主要信息的有监督学习方法在车辆再识别领域比较常用,然而在许多情形下此类方法的再识别难度却非常大,例如同一型号的

车辆没有任何特殊标识, 外观信息完全一致, 这时仅依靠车身外观无法识别车辆身份。更加棘手的问题是, 即便车辆存在一些具有辨识性的特殊标识(如装饰物、年检标志、贴纸等), 在车辆的非前方视角下也无法被摄像头捕获。另外, 车辆距离摄像头较远、光照不足等限制条件也导致车辆图像分辨率较低, 因此以车身外观为监督信号进行再识别的方法性能受到了一定限制。

为了解决这类问题, 研究人员提出了一些利用时间、地理信息的深度学习方法。在参考文献[39]提出的 PROVID 方法中, 作者对数据集中车辆图像之间的时间-地理信息进行了统计与分析, 发现时间间隔越短、地理位置距离越近的两幅图像, 越有可能出自同一车辆。因此在得到基于外观信息方法的结果之后, 作者计算了查询图像和图库图像序列之间的时间-地理相似性, 并将时间-地理相似性作为正则项对结果进行重排, 大幅度提升了再识别的性能。

在现实交通环境中, 大部分车辆会在道路上定速行驶, 但是也有不少车辆会在摄像头无法拍摄的地点长时间停留, 或相隔很长一段时间才再次进入监控范围, 大多数再识别方法忽略了车辆出现的时间-地理关系, 因此难以获得较好的再识别结果。Wang 等^[10]认为每一对摄像头下车辆出现的时间间隔是一个随机变量, 并服从某种概率分布。经过分析, 作者采用对数正态分布对同一车辆经过两个摄像头的的时间间隔进行建模, 并对时序相似度与外观相似度分数进行加权求和, 以表示两幅图像的综合相似度。

除了对时间信息进行定量分析以外, 对地理位置的定性分析也有利于解决此类问题。给定 A、B、C 是一条直线道路上依次出现的 3 个摄像头, 如果一辆汽车在摄像头 A 和 C 下出现, 那么一定也会在 B 下出现。因此, 对于给定的一对在 A 和 C 下出现的车辆图像, 如果在一定的时间间隔内 B 中没有出现类似车辆, 则两幅图像属于同一车辆的概率就很小。参考文献[59]提出将双重卷积神经网络(Siamese-CNN)和长短期记忆(long short term memory)网络融合, 把每一张图像中的车辆外观信息、时间戳和地理位置信息合并成一个视-时-空状态, 在比较车辆的相似度时, 向网络中输入两张图像的视-时-空状态, 该方法将综合视觉外观相似度和时间-地理相似性给出判断结果。在已知起点和终

点的情况下, 该方法首先采取链式马尔科夫随机场(chain MRF)模型提出候选路径, 随后利用双重卷积神经网络计算出候选路径上两个相邻图像的外观相似性和时空相似性, 并输出一条需要验证的路径序列。长短期记忆网络则以这条路径序列作为先验信息进行特征嵌入, 按照时间推移顺序计算相似性得分。

在现实场景下, 摄像头之间的候选路径非常多, 且每条路径上出现的车辆也非常多, 此类算法在执行再识别任务时的时间复杂度十分高, 因此在计算验证路径时, 需要尽可能地减少重复冗余的计算。

与行人再识别中行人出现的随机性、无时序性相比, 车辆再识别中的汽车具有规律性、时序性的特点, 交通监控图像提供的时间和位置信息能够对车辆再识别任务提供很大的帮助, 上述方法都对时间-地理信息进行了合理地分析和利用, 这也是值得该领域内的研究人员思考的一个方向。

值得注意的是, 使用时间-地理信息的再识别方法对数据集的要求十分严格, 需要大量额外的时间戳和地点标注工作。目前能够提供时间-地理信息的只有 VeRi776 数据集, 其规模相对较小。

4 数据集及性能比较

4.1 数据集介绍

在大量基于深度学习的车辆再识别方法被提出之前, 车辆再识别的数据集十分匮乏, 且规模较小, 图像标注信息较少, 多数与车辆相关的数据集主要用于分类任务。有研究人员从 ImageNet^[60]中抽取了包含 1 537 辆不同类型的车辆的图像数据集, 用于车辆分类任务。Yang 等^[61]提出了一个综合的车辆数据集 CompCars, 其中包含两种图像来源, 一种是从网络中截取的 1 687 个汽车模型照片, 另一种是从交通监控系统中获取的图像, 但是所有图像只有车身的前方视角, 且摄像头之间没有时序联系, 无法保证车辆的复现率, 因此该数据集不能满足车辆再识别任务的要求, 主要被用于细粒度分类和车型预测任务。

如前文所述, 一个标准的车辆再识别数据集应该至少满足两点要求: 包含大量从真实交通环境下捕获的车辆图像; 车辆必须有较高的复现率。复现率是指同一车辆需在不同摄像头下以不同视角和背景出现, 以满足再识别任务中车辆的全方位跨视域检索。

VeRi 数据集^[6]由 20 个摄像头拍摄的 40 000 张关于 619 辆汽车的图像组成, 标注信息丰富, 包括边界框、车辆类型、颜色、厂商等, 每一辆汽车被至少 2 个至多 18 个摄像头拍摄, 图像具有不同的视角、光照、分辨率和遮挡情况, 构成了一个真实场景下的高复现率数据集。参考文献[39]对 VeRi 数据集进行了扩充和标注, 提出了著名的 VeRi776 数据集, VeRi776 数据集也是目前使用较广泛的车辆再识别数据集之一。VeRi776 数据集包含了 49 360 张关于 776 辆汽车的图像, 在标注信息上除了 VeRi 数据集已有的类型, 还多出了对车牌、时间戳和地理位置的标注。

VehicleID 是由 Liu 等^[8]提出的大规模车辆再识别数据集, 其中包括 221 763 张关于 26 267 辆汽车的图像, 主要包含前后两种视角, 且每张图像除了车辆 ID、摄像头编号的标注信息以外, 还有车辆型号的详细信息 (共 250 种厂商车型), 为了使车辆再识别方法的性能评测更加全面, VehicleID 将测试集按照车辆图像的尺寸划分为大、中、小 3 个子集。

Vehicle-1M 是中国科学院自动化研究所^[50]提出的车辆再识别数据集, 其中包含 936 051 张关于 55 527 辆汽车的图像, 每张图像中的汽车都被标注了厂商、车型和发售年份 (共 400 种厂商车型)。同样地, Vehicle-1M 的测试集也根据图像尺寸被分为了大、中、小 3 个子集。随着深度学习在车辆再识别技术中的发展与应用, 研究人员对车辆图像数据的要求越来越多, Vehicle-1M 是迄今为止规模最大的车辆再识别数据集, 可为深度神经网络的学习过程提供充足的训练数据。

VRIC 是由 Kanaci 等^[62]提出的真实世界车辆再识别数据集, 之前的数据集在采集过程中为保证图像质量都受到了一定程度的约束, 如最低分辨率限制、最低亮度等, VRIC 在采集过程中尽量保留了实际情况中采集到的车辆图像特点 (如具有多种分辨率、视角变化、光照强弱和遮挡情况), 存在车辆运动模糊的情况, 具有不受限和随机性的特点。它包含 60 430 张关于 5 622 辆汽车的图像, 捕获自 60 个不同的交通摄像头, 拍摄时间包括白天和晚上。训练集有 54 808 张图像 (共 2 811 辆汽车), 其余 5 622 张图像 (共 2 811 辆汽车) 被用来组成测试集。

VERI-Wild 是由 Lou 等^[63]发布的车辆再识别数据集, 其收集于市郊地区, 包含 174 个交通摄像头拍摄的 416 314 张关于 40 671 辆汽车的图像。

VERI-Wild 是在 200 km² 的市郊地区收集得到的, 因此车辆在被捕获的过程中不受过多限制, 且车辆所处场景更加丰富, 车辆图像的采集时间跨度长, 光照和天气的变化十分明显。训练集包括 277 797 张图像 (共 30 671 辆汽车), 测试集包括 138 517 张图像 (共 10 000 辆汽车)。同样地, VERI-Wild 的测试集也根据图像尺寸被分为了大、中、小 3 个子集。

4.2 评价指标

车辆再识别是一个图像检索子问题, 在衡量再识别方法的性能时, 采用的评价指标主要包括平均精度均值 (mean average precision, mAP)、累计匹配性能 (cumulative match characteristic, CMC) 曲线和 Rank-*N* 表格。

(1) 平均精度均值

平均精度均值用于评估再识别方法的总体性能, 表示所有检索结果精度的平均值。如第 2.2 节提到的, 测试集分为查询集和图库集, 首先对每一个查询集图像 q 的检索结果计算平均精度 AP, 如式 (1), k 表示图库集图像序号, n 表示图库集图像总数, N 表示目标车辆的图像总数。 $P(k)$ 表示在检索序列中第 k 位之前的精度, $gt(k)$ 表示第 k 位图像是否是目标车辆, Q 代表查询图像的总数。最终对所有查询集检索结果的平均精度求均值, 即 mAP, 如式 (2) 所示。

$$AP = \frac{\sum_{k=1}^n P(k) \times gt(k)}{N} \quad (1)$$

$$mAP = \frac{\sum_{q=1}^Q AP(q)}{Q} \quad (2)$$

(2) 累计匹配性能曲线

平均精度均值仅能从总体上反映方法的再识别性能, 不能反映不同检索条目下精度的分布情况, 累计匹配性能曲线表示了在前 k 个检索结果中找到正确结果的概率。横坐标表示 k , 纵坐标表示正确率, 将不同方法的性能曲线画在图中可以更直观地比较性能。第 k 位之前的 CMC 如式 (3) 所示, 当图像 q 的正确匹配目标出现在检索序列的第 k 位之前时, $gt(q, k)$ 等于 1。

$$CMC@k = \frac{\sum_{q=1}^Q gt(q, k)}{Q} \quad (3)$$

(3) Rank-*N* 表格

在对不同方法进行性能比较时,若方法之间的性能差异不大,累计匹配性能曲线会出现大部分重叠的情况,从而无法准确地判断性能好坏。为了更简洁地对比不同方法之间的性能差异,一般选择一些关键匹配位置的累积匹配准确率进行比较,其中 Rank-1 和 Rank-5 比较常见,分别表示在结果序列中前 1 张和前 5 张图像正确匹配的概率。

以上 3 种评价指标中,平均精度均值和 Rank-*N* 表格被使用的频率最高,为了尽可能多地比较各个方法,本文选择这两种指标进行性能比较。

4.3 性能比较

目前领域内最常用于性能测试的两个数据集是 VeRi776 和 VehicleID。虽然有更多的大型数据集被相继提出,但是由于提出时间比较靠后,一些较早的方法在较晚发布的数据集(如 Vehicle-1M)上没有充足的实验数据,故本文不在这些数据集上进行性能比较。此外,一些国内的研究方法^[69-71]没有典型数据集上的实验结果,而一些国外的研究方法过于复杂,不适用于大型车辆图像数据集,且实现难度较高,同样无法对这些方法作性能比较。表 2 和表 3 分别列出了本文提到的一些车辆再识别方法在这两个典型数据集上的性能比较结果。一些方法(如 S-CNN+PathLSTM)要用到时间-地理信息,故只能在 VeRi776 数据集上进行实验。还有一些方法

(如 OIFE、PGAN)只在 VehicleID 的大尺寸(Large)子集进行了测试, Large 子集被认为是 VehicleID 数据集里最难识别的部分,因此足够证明这些方法的有效性。

表 2 VeRi776 数据集上流行方法的性能对比

方法	参考文献序号	评价指标		
		mAP	Rank-1	Rank-5
FACT	[6]	19.9	59.6	75.2
XVGAN	[64]	24.7	60.2	77.0
FACT+SNN+STR	[6]	27.7	61.4	78.7
OIFE	[10]	51.4	68.3	89.7
PROVID	[39]	53.4	81.5	95.1
FDA-Net	[63]	55.4	84.2	92.4
VGG+C+T+S	[49]	57.4	86.5	92.8
S-CNN+PathLSTM	[59]	58.2	83.5	90.0
GS-TRE	[51]	59.4	96.2	98.9
VAMI	[56]	61.3	85.9	91.8
AAVER	—	61.1	90.1	94.7
RAM	[65]	61.5	88.6	94.0
VANet	[57]	66.3	89.7	95.9
PNVR	[43]	74.3	94.3	98.7
MRL	[58]	78.5	94.3	99.0
PGAN	—	79.3	96.5	98.3
PRN	[54]	85.8	97.1	99.4

表 3 VehicleID 数据集上流行方法的性能对比(按图像尺寸分类)

方法	参考文献序号	评价指标					
		小尺寸		中尺寸		大尺寸	
		Rank-1	Rank-5	Rank-1	Rank-5	Rank-1	Rank-5
DRDL	[8]	49.0%	73.5%	42.8%	66.8%	38.2%	61.6%
C2F-Rank	[50]	61.1%	81.7%	56.2%	76.2%	51.4%	72.2%
VAMI	[56]	63.1%	83.2%	52.8%	75.1%	47.3%	70.2%
FDA-Net	[63]	64.0%	82.8%	57.8%	78.3%	49.4%	70.4%
VGG+C+T+S	[49]	69.9%	87.3%	66.2%	82.3%	63.2%	79.4%
OIFE	[10]	—	—	—	—	67.0%	82.9%
GS-TRE	[51]	75.9%	84.2%	74.8%	83.6%	74.0%	82.7%
AAVER	—	74.6%	93.8%	68.6%	89.9%	63.5%	85.6%
PNVR	[43]	78.4%	92.3%	75.0%	88.3%	74.2%	86.4%
PGAN	—	—	—	—	—	77.8%	92.1%
PRN	[54]	78.9%	94.8%	74.9%	92.0%	71.5%	88.4%
VANet	[57]	83.2%	95.9%	81.1%	94.7%	77.2%	86.7%
MRL	[58]	84.8%	96.9%	80.9%	94.1%	78.4%	92.1%

随着领域内研究人员的不断思考与创新,各个数据集上的最佳性能不断地被刷新。可以看到,基于深度学习的车辆再识别方法的识别精度远高于人工设计的方法,在两个典型数据集上的性能表现都十分优异。

随着度量学习方法的进步以及时间-地理信息被利用, VeRi776 数据集上的研究方法在识别性能上增长显著,其中以 PRN 为代表的度量学习方法的识别率最高, PRN 方法的平均精度均值达到了 85.84%, 较之前的方法都有大幅度的领先。同时可以看出,基于手工设计特征的方法 FACT 无法很好地完成车辆再识别任务,这显示出手工设计特征对车辆身份表达能力的不足。一些方法(如 DRDL 等)需要用到车辆型号等辅助信息,因此适合在 VehicelID 数据集上进行实验,其中 MRL 在 VehicelID 各测试子集上的综合表现都超越了之前的方法。

综合上述分析和表格数据,可以看出深度学习对于车辆图像具有强大的特征表达能力,能够较好地化解光照变化、视角变化等客观因素带来的困难和挑战,在完成训练后,可以在极短的时间内完成对目标车辆的再识别过程。

总的来说,通过合理地设计网络结构和损失函数,深度学习方法可以准确、快速地在海量图像中找出目标车辆,是可以较好地完成车辆再识别任务的基础框架。

5 结束语

本文结合近年来车辆再识别领域的研究工作,从设计思路对目前的再识别方法进行了分类总结,并分析了不同算法的优点和缺点。

从本领域的发展情况来看,基于手工特征的车辆再识别方法已经成为历史,随着数据集规模的不断扩大,基于深度学习的方法被广泛证明比手工设计特征的方法有效得多,设计一个对车辆身份具有辨识性和敏感性的深度神经网络是目前车辆再识别工作的一个基本要求。

随着神经网络结构的深度化和复杂化以及所用数据集规模的扩大,基于深度学习的方法通常需要更长的时间进行训练才能得到良好的性能;同时,现有的方法在面对车辆再识别中类间相似度高、类内差异大等研究难点时,还无法获得足够高的精确度和鲁棒性。而实际应用场景对再识别算法

的效率以及识别精度都有很高的要求,所以目前车辆再识别技术与其实应用还有一定的差距,未来的研究工作可以从以下几方面展开。

- 目前大多车辆再识别方法属于有监督学习方法,只能在车辆身份和总数已知的前提下进行训练,在同一数据集下的测试集中表现良好,但在跨数据集或现实情况下的性能表现较差。最近在行人再识别方面已经有人提出利用领域自适应来解决这种问题,领域自适应是迁移学习的一种,旨在利用源领域中带标签的样本来解决目标领域的学习问题,其关键在于如何最大限度地减小领域间的分布差异。未来研究可在此方向进行展开,以提升车辆再识别技术在安防监控条件下的应用能力。

- 由于摄像头视角和车型相同等导致的类内差异大、类间相似度高仍然是限制车辆再识别技术取得高识别精度的主要原因,如何提出合适的度量学习方法以提取具有鲁棒性的车辆特征,仍是目前迫切需要解决的问题。

- 车辆再识别领域目前存在精度和效率不可兼得的问题,如何在轻量化的网络结构上取得足够高的识别精度,使得再识别技术能应用到嵌入式系统或移动设备中,也是目前该领域的一大研究难点。

- 现有的车辆再识别的相关研究大多是在车辆图像已经切割好的假设下进行的,但是实际场景中往往需要在一张复杂的图像中找出目标车辆,因此包括目标分割、特征提取和目标识别的端到端车辆再识别系统将是研究热点之一。

- 在现实交通环境中,往往会出现道路拥堵的情况,此时车辆高度密集出现,导致车辆检测框中包含其他车辆的部分外观,大大增加了车辆身份的识别难度。另外,拍摄距离远、设备老化等导致的图像分辨率低的问题也会进一步增加车辆再识别的难度。如何克服复杂环境因素的负面影响,也是车辆再识别研究领域中的一个难点。

参考文献:

- [1] FERIS R S, SIDDIQUIE B, PETTERSON J, et al. Large-scale vehicle detection, indexing, and search in urban surveillance videos[J]. IEEE Transactions on Multimedia, 2011, 14(1): 28-42.
- [2] YANG L, LUO P, CHANGE LOY C, et al. A large-scale car dataset for fine-grained categorization and verification[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015: 3973-3981.
- [3] SUN C C, ARR G S, RAMACHANDRAN R P. Vehicle reidentification

- tion using multidetector fusion[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2004, 5(3): 155-164.
- [4] LI N, JAIN J J, BUSO C. Modeling of driver behavior in real world scenarios using multiple noninvasive sensors[J]. *IEEE Transactions on Multimedia*, 2013, 15(5): 1213-1225.
 - [5] MEI T, RUI Y, LI S. Multimedia search reranking: a literature survey[J]. *ACM Computing Surveys (CSUR)*, 2014, 46(3): 1-38.
 - [6] LIU X, LIU W, MA H. Large-scale vehicle re-identification in urban surveillance videos[C]// *The 2016 IEEE International Conference on Multimedia and Expo (ICME)*. Piscataway: IEEE Press, 2016: 1-6.
 - [7] KWONG K, KAVALER R, RAJAGOPAL R. Arterial travel time estimation based on vehicle re-identification using wireless magnetic sensors[J]. *Transportation Research Part C: Emerging Technologies*, 2009, 17(6): 586-606.
 - [8] LIU H, TIAN Y, YANG Y. Deep relative distance learning: tell the difference between similar vehicles [C]// *The IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2016: 2167-2175.
 - [9] KHAN S D, ULLAH H. A survey of advances in vision-based vehicle re-identification[J]. *Computer Vision and Image Understanding*, 2019, 182: 50-63.
 - [10] WANG Z, TANG L, LIU X. Orientation invariant feature embedding and spatial temporal regularization for vehicle re-identification[C]// *The IEEE International Conference on Computer Vision*. Piscataway: IEEE Press, 2017: 379-387.
 - [11] KLEIN L A, KELLEY M R, MILLS M K. Evaluation of overhead and in-ground vehicle detector technologies for traffic flow measurement[J]. *Journal of Testing and Evaluation*, 1997, 25(2): 205-214.
 - [12] DAVIES P. Vehicle detection and classification[J]. *Information Technology Applications in Transport*, 1986: 11-40.
 - [13] SANCHEZ R O, FLORES C, HOROWITZ R, et al. Vehicle re-identification using wireless magnetic sensors: algorithm revision, modifications and performance analysis[C]// *2011 IEEE International Conference on Vehicular Electronics and Safety*. Piscataway: IEEE Press, 2011: 226-231.
 - [14] CHARBONNIER S, PITTON A C, VASSILEV A. Vehicle re-identification with a single magnetic sensor[C]// *2012 IEEE International Instrumentation and Measurement Technology Conference*. Piscataway: IEEE Press, 2012: 380-385.
 - [15] KWONG K, KAVALER R, RAJAGOPAL R, et al. Arterial travel time estimation based on vehicle re-identification using wireless magnetic sensors[J]. *Transportation Research Part C: Emerging Technologies*, 2009, 17(6): 586-606.
 - [16] JENG S T C, CHU L. Vehicle re-identification with the inductive loop signature technology[J]. *Journal of the Eastern Asia Society for Transportation Studies*, 2013(10): 1896-1915.
 - [17] NDOYE M, TOTTEN V, CARTER B, et al. Vehicle detector signature processing and vehicle reidentification for travel time estimation[C]// *Transportation Research Board Meeting*. [S.l.:s.n.], 2008.
 - [18] NDOYE M, TOTTEN V, KROGMEIER J V, et al. A signal processing framework for vehicle re-identification and travel time estimation[C]// *The 2009 12th International IEEE Conference on Intelligent Transportation Systems*. Piscataway: IEEE Press, 2009: 1-6.
 - [19] ALI S S M, GEORGE B, VANAJAKSHI L. Multiple inductive loop detectors for intelligent transportation systems applications: Ramp metering, vehicle re-identification and lane change monitoring systems [C]// *2013 IEEE Symposium on Computers & Informatics (ISCI)*. Piscataway: IEEE Press, 2013: 176-180.
 - [20] PRINSLOO J, MALEKIAN R. Accurate vehicle location system using RFID, an internet of things approach[J]. *Sensors*, 2016, 16(6): 825-848.
 - [21] MAZLOUMI E, CURRIE G, ROSE G. Using GPS data to gain insight into public transport travel time variability[J]. *Journal of Transportation Engineering*, 2009, 136(7): 623-631.
 - [22] ZHANG Z, TAN T, HUANG K. Three-dimensional deformable-model-based localization and recognition of road vehicles[J]. *IEEE Transactions on Image Processing*, 2012, 21(1): 1-13.
 - [23] WOESLER R. Fast extraction of traffic parameters and re-identification of vehicles from video data[C]// *The 2003 IEEE International Conference on Intelligent Transportation Systems*. Piscataway: IEEE Press, 2003: 774-778.
 - [24] GUO Y, RAO C, SAMARASEKERA S, et al. Matching vehicles under large pose transformations using approximate 3D models and piecewise MRF model[C]// *The IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2008: 1-8.
 - [25] ZAPLETAL D, HEROUT A. Vehicle re-identification for automatic video traffic surveillance[C]// *The IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2016: 25-31.
 - [26] DUBSKÁ M, HEROUT A, SOCHOR J. Automatic camera calibration for traffic understanding[C]// *British Machine Vision Conference (BMVC)*. Guildford: BMVA Press, 2014, 4(6): 8.
 - [27] DUBSKÁ M, HEROUT A, JURÁNEK R. Fully automatic roadside camera calibration for traffic surveillance[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2015, 16(3): 1162-1171.
 - [28] FERENCZ A, LEARNED-MILLER E G, MALIK J. Building a classification cascade for visual identification from one example [C]// *The IEEE International Conference on Computer Vision*. Piscataway: IEEE Press, 2005, 1: 286-293.
 - [29] LIU X, LIU W, MEI T, et al. A deep learning-based approach to progressive vehicle re-identification for urban surveillance[C]// *The European Conference on Computer Vision*. Heidelberg: Springer, 2016: 869-884.
 - [30] ZHENG Q, LIANG C, FANG W, et al. Car re-identification from large scale images using semantic attributes[C]// *The IEEE 17th International Workshop on Multimedia Signal Processing*. Piscataway: IEEE Press, 2015: 1-5.
 - [31] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. *International Journal of Computer Vision*, 2004, 60(2): 91-110.
 - [32] VAN DE WEIJER J, SCHMID C, VERBEEK J. Learning color names for real-world applications[J]. *IEEE Transactions on Image Processing*, 2009, 18(7): 1512-1523.
 - [33] LIAO S, HU Y, ZHU X. Person re-identification by local maximal occurrence representation and metric learning[C]// *The IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway: IEEE Press, 2015: 2197-2206.
 - [34] FU Y, WEI Y, WANG G, et al. Self-similarity grouping: a simple unsupervised cross domain adaptation approach for person re-identification[C]// *The IEEE International Conference on Computer Vision*. Piscataway: IEEE Press, 2019: 6112-6121.

- [35] DENG W, ZHENG L, YE Q, et al. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 994-1003.
- [36] PENG P, XIANG T, WANG Y, et al. Unsupervised cross-dataset transfer learning for person re-identification[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 1306-1315.
- [37] WANG J, ZHU X, GONG S, et al. Transferable joint attribute-identity deep learning for unsupervised person re-identification[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 2275-2284.
- [38] WEI L, ZHANG S, GAO W, et al. Person transfer GAN to bridge domain gap for person re-identification[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 79-88.
- [39] LIU X, LIU W, MEI T. PROVID: progressive and multimodal vehicle reidentification for large-scale urban surveillance[J]. IEEE Transactions on Multimedia, 2018, 20(3): 645-658.
- [40] GUO Y F, WU L, LU H. Null Foley-Sammon transform[J]. Pattern Recognition, 2006, 39(11): 2248-2251.
- [41] ZHANG L, XIANG T, GONG S. Learning a discriminative null space for person re-identification[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 1239-1248.
- [42] YAN K, TIAN Y, WANG Y. Exploiting multi-grain ranking constraints for precisely searching visually-similar vehicles[C]// The 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 562-570.
- [43] HE B, LI J, ZHAO Y, et al. Part-regularized near-duplicate vehicle re-identification[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 3997-4005.
- [44] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2016: 779-788.
- [45] JOSEPH REDMON AND ALI FARHADI. YOLO9000: better, faster, stronger[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 6517-6525.
- [46] TANG Z, NAPHADE M, BIRCHFIELD S, et al. PAMTRI: pose-aware multi-task learning for vehicle re-identification using highly randomized synthetic data[C]// The IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 211-220.
- [47] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 5693-5703.
- [48] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely connected convolutional networks[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 4700-4708.
- [49] ZHANG Y, LIU D, ZHA Z J. Improving triplet-wise training of convolutional neural network for vehicle re-identification[C]// The 2017 IEEE International Conference on Multimedia and Expo (ICME). Piscataway: IEEE Press, 2017: 1386-1391.
- [50] GUO H, ZHAO C, LIU Z. Learning coarse-to-fine structured feature embedding for vehicle re-identification[C]// The 3rd AAAI Conference on Artificial Intelligence. Piscataway: IEEE Press, 2018: 6853-6860.
- [51] BAI Y, LOU Y, GAO F. Group-sensitive triplet embedding for vehicle re-identification[J]. IEEE Transactions on Multimedia, 2018, 20(9): 2385-2399.
- [52] SUN Y, ZHENG L, YANG Y, et al. Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline)[C]// The European Conference on Computer Vision (ECCV). Heidelberg: Springer, 2018: 480-496.
- [53] WANG G, YUAN Y, CHEN X, et al. Learning discriminative features with multiple granularities for person re-identification[C]// The 2018 ACM Multimedia Conference on Multimedia Conference. New York: ACM Press, 2018: 274-282.
- [54] CHEN H, LAGADEC B, BREMOND F. Partition and reunion: a two-branch neural network for vehicle re-identification[C]// The IEEE Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE Press, 2019: 184-192.
- [55] NEWELL A, YANG K, DENG J. Stacked hourglass networks for human pose estimation[C]// The European Conference on Computer Vision. Heidelberg: Springer, 2016: 483-499.
- [56] ZHOU Y, SHAO L, DHABI A. Viewpoint-aware attentive multi-view inference for vehicle re-identification[C]// The IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2018: 6489-6498.
- [57] CHU R, SUN Y, LI Y, et al. Vehicle re-identification with viewpoint-aware metric learning[C]// The IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 8282-8291.
- [58] LIN W, LI Y, YANG X, et al. Multi-view learning for vehicle re-identification[C]// 2019 IEEE International Conference on Multimedia and Expo (ICME). Piscataway: IEEE Press, 2019: 832-837.
- [59] SHEN Y, XIAO T, LI H. Learning deep neural networks for vehicle re-id with visual-spatio-temporal path proposals[C]// The 2017 IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 1918-1927.
- [60] DENG J, DONG W, SOCHER R. ImageNet: a large-scale hierarchical image database[C]// The 2009 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2009: 248-255.
- [61] YANG L, LUO P, CHANGE LOY C. A large-scale car dataset for fine-grained categorization and verification[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2015: 3973-3981.
- [62] KANACI A, ZHU X, GONG S. Vehicle re-identification in context[C]// The German Conference on Pattern Recognition. Heidelberg: Springer, 2018: 377-390.
- [63] LOU Y, BAI Y, LIU J, et al. Veri-wild: a large dataset and a new method for vehicle re-identification in the wild[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 3235-3243.
- [64] ZHOU Y, SHAO L. Cross-view GAN based vehicle generation for re-identification[C]// British Machine Vision Conference (BMVC). Heidelberg: Springer, 2017.
- [65] LIU X, ZHANG S, HUANG Q, et al. Ram: a region-aware deep model for vehicle re-identification[C]// The 2018 IEEE International Conference on Multimedia and Expo (ICME). Piscataway: IEEE Press, 2018: 1-6.

- [66] ANTONIO M P, PALAZZI A, BERGAMINI L, et al. Unsupervised vehicle re-identification using triplet networks[C]// The IEEE Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE Press, 2018: 166-171.
- [67] BASHIR R M S, SHAHZAD M, FRAZ M M. DUPL-VR: deep unsupervised progressive learning for vehicle re-identification[C]// The International Symposium on Visual Computing. Heidelberg: Springer, 2018: 286-295.
- [68] BASHIR R M S, SHAHZAD M, FRAZ M M. Vr-proud: vehicle re-identification using progressive unsupervised deep architecture[J]. Pattern Recognition, 2019, 90: 52-65.
- [69] 王茜, 陈一民, 丁友东. 复杂环境中基于视觉词袋模型的车辆再识别算法[J]. 计算机应用, 2018, 38(5): 83-87.
WANG Q, CHEN Y M, DING Y D. Vehicle re-identification algorithm based on bag of visual words in complicated environments[J]. Computer Application, 2018, 38(5): 83-87.
- [70] 李熙莹, 周智豪, 邱铭凯. 基于部件融合特征的车辆再识别算法[J]. 计算机工程, 2019(6): 12-20.
LI X Y, ZHOU Z H, QIU M K. Vehicle re-identification algorithm based on component fusion feature[J]. Computer Engineering, 2019(6): 12-20.
- [71] 王盼盼, 李玉惠. 基于特征融合和 L-M 算法的车辆再识别方法[J]. 电子科技, 2018, 31(4): 12-15.
WANG P P, LI Y H. Vehicle re-identification algorithm based on component fusion feature[J]. Electronic Science and Technology, 2018, 31(4): 12-15.

[作者简介]



刘凯（1996—），男，北京交通大学计算机与信息技术学院硕士生，主要研究方向为计算机视觉、车辆再识别。



李滄东（1982—），男，博士，北京交通大学计算机与信息技术学院教授、副院长，主要研究方向为大数据分析与安全、隐私保护、智能交通等。



林伟鹏（1995—），男，北京交通大学计算机与信息技术学院硕士生，主要研究方向为计算机视觉、车辆再识别。