



# 基于深度学习的车辆重识别算法综述

宁 轩 何智敏(宁波大学信息科学与工程学院 浙江 宁波 315211)

**摘 要:**随着计算机视觉技术的迅速发展以及人们对于建设智慧城市的渴望,车辆重识别技术取得了不小的突破。它可以助力于搜寻救援、打击犯罪等诸多领域。深度学习及神经网络为该项任务突破传统特征的瓶颈带来了可能,而随着各种大规模数据集的提出,越来越多的学者关注到了此项任务,并成为当下的研究热点。本文对车辆重识别任务的兴起、发展及现状进行了一定的归纳总结,提出了现有技术下的一些不足,并对未来的发展做出了一些思考与预测。

**关键词:**车辆重识别;深度学习;卷积神经网络

**中图分类号:**TP183 **文献标识码:**A

## 0 引言

深度学习的崛起以及适用于各种场景的网络被提出,使重识别算法取得了极大的进展。通过提取输入图片特征,在跨摄像头拍摄的图库中检索与该图片同一身份的任务,即称为重识别,它可以被视作图像检索的子问题。而随着人们对车辆的依赖性增加、摄像头数量的剧增以及人们对建设智慧城市的渴望,车辆重识别任务引起了学者们的巨大关注。车辆重识别任务在现实中可以助力于智能交通的建设,也可以作为城市智慧管理、安防建设的一个有力工具。

而随着各种大规模、高质量车辆重识别数据集的提出,相关研究也进一步的使得车辆重识别这一任务满足于现实中的各种的需求。根据实现方式的不同,可将这些方法大致分为两类:第一类是基于非深度学习的方法,主要依赖于一些外部设备或者利用传统特征方法;第二类是基于深度学习的方法。本文将对基于深度学习的车辆重识别算法进行重点分析。

## 1 问题描述

### 1.1 问题来源

虽然每辆车都有唯一的一个车牌,在不考虑其余视觉信息的情况下,依靠车牌对车辆进行搜索与识别是最简单、快捷的方法。但是在很多的场景下,仅仅依靠车牌是不足以对车辆完成识别的。首先,车辆重识别任务的特性之一是在跨摄像头拍摄的图库中寻找同一身份的车辆。由于拍摄视角的不同,不能保证每张图片上都带有车牌信息。其次,在助力于打击犯罪时,嫌疑人会对车牌信息进行丢弃或更改,导致车牌信息无法利用。因此在安防、救援等场景中以车辆自身视觉信息(如车型、颜色等)作为查询特征在跨摄像头拍摄的图库中进行搜索与匹配成为了一个具有重大应用价值和广阔研究空间的课题,即称为车辆重识别。

### 1.2 问题描述

车辆重识别任务具体来说是利用一张查询(query)图片或者一段视频,在一个非重叠摄像头拍摄的图库(gallery)或者视频库中检索同一身份的车辆,且

这个库中拍摄到的车辆可能受到天气、光照、背景的影响导致质量不同。如何克服这些困难提高重识别的准确率,是目前的研究热点。目前的重识别任务流程主要有:输入查询(query)和图库(gallery)车辆特征;网络提取特征后嵌入高维空间得到向量;进行相似性度量,距离更小的被视作同一辆车的概率更大;展示匹配结果。在训练时,定义一个训练集  $D_{\text{train}} = \{I_k, y_k\}_{k=1}^N$ , 表达的含义为第  $K$  辆车  $y_k$  所有的图片  $I_k$ ,  $N$  代表车辆身份总数。在训练阶段,我们的目的是学习一个函数  $F(\cdot)$  用于提取所有图片  $I$  的特征,即  $f_I = F(I)$ 。在测试阶段时,从测试集  $D_{\text{test}}$  中获取一对图片  $\{I_i, I_j\}$ , 且图片身份在两个数据集中未重叠,我们需要判断两张图片是否属于同一辆车。车辆重识别任务的常见流程如下图 1 所示:

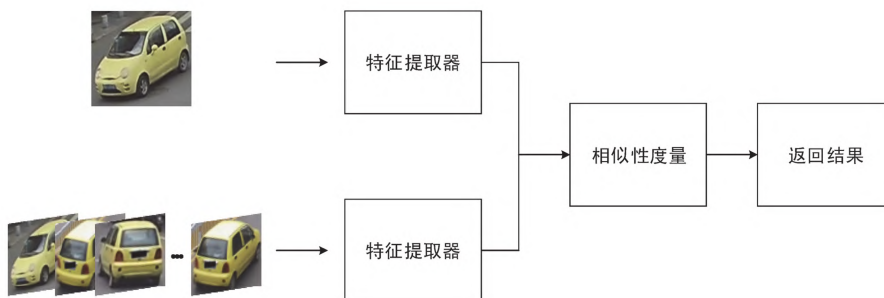


图 1 车辆重识别流程

### 1.3 数据集

由于车辆重识别侧重于车辆的身份检索,因此使用的数据集默认是经过一定的切割后所获得的图片。使用较为广泛的数据集有 VeRi-766<sup>[1]</sup>、VehicleID<sup>[2]</sup> 等。这些数据集中将图片分为训练集和测试集。VeRi-766 数据集中共有 49360 张,由 776 辆车拍摄而成,它所包含的标注信息是现有车辆数据集中最为丰富之一。它将车辆的车型、颜色、车牌和生产厂商等都做了详细且精确的标注,每一辆汽车也由多个摄像头拍摄而成,符合车辆重识别任务跨摄像头拍摄的特征,是一个真实场景下的高质量数据集。在 VehicleID 数据集中包括了由 26267 辆车拍摄而成的 221763 张图片,它所提供的标注包括了车辆身份、摄像头编号、车型等信息,并将测试集根据不同车辆图片尺寸分为了大、中、小三个子集以供测试使用。

### 1.4 评价指标

车辆重识别可以看成是图像检索的子问题,目前该任务主要采用以下三个指标相结合的方式以衡量车

辆重识别方法的性能。

#### (1) 平均精度均值

平均精度均值在意义上表示所有检索结果精度的平均值,是衡量方法总体性能的一个指标。首先对每一个查询集图像  $q$  的检索结果进行一次计算,得到平均精度  $AP$ ,如式(1)。最终对所有查询集检索结果的平均精度求均值,如式(2),得到  $mAP$ 。

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

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

#### (2) 累计匹配性能

平均精度均值是反映重识别方法总体性能的一个

指标,但当需要反映单个检索条目下的精度时,就需要用到累积匹配性能。计算如式(3):

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

#### (3) Rank-N 表格

我们大多采用 Rank-1 或者 Rank-5 以表示返回结果中,相似度排第一和相似度前五的图片中,存在与查询图片同一身份的概率,该指标可对重识别任务进行准确性描述。

## 2 基于深度学习的车辆重识别方法分类

受益于计算机算力的上涨和卷积神经网络的提出,计算机视觉任务可以容纳更多的参数,也可以提取到更多特征用于车辆重识别任务。在基于深度学习的车辆重识别任务中,通常以构建一个能够提取车辆辨识性特征的神经网络为最终目的。该网络通过训练后能将车辆特征投射到高维空间中获得向量,并通过衡



量成对向量间的距离以获得车辆相似性,距离越小的两个向量被视作同一辆车的概率越大。目前基于深度学习的车辆重识别方法性能已经对传统方法达到全面覆盖,因此本文将重点分析这一方法。

## 2.1 融合车身多维度信息的方法

与行人重识别不同的是,具有同一外观(如车型和配色)的车辆在场景中较为常见。而在没有车牌信息的帮助下,人眼都难以区分出两辆车是否属于同一身份,因此如何提取出具有辨识性的特征将成为车辆重识别任务的重点。具有辨识性的特征有诸如车窗贴图、车内物品摆放等,但这些特征区域较小,精准挖掘到这些特征是能够提升车辆辨识度的一个方法。

基于上述考量,Meng<sup>[7]</sup>等人提出将车身拆解,分块进行匹配的思路。作者借助一个车身解析网络将车身特征拆解为车头、车侧、车尾和车顶四个区域,并分别得到四个区域的特征向量。在匹配时两辆车的同一区域进行匹配并获取该区域的共有视觉得分,进而融合四个区域的共有视觉得分作为两辆车的局部特征距离。将局部特征距离与全局特征距离以一定权重结合后即为两辆车的特征距离,该方法取得了不错的效果但依赖于车辆解析网络的性能。Zhao<sup>[8]</sup>等人提出用检测网络预处理在车辆中可能存在辨识性特征的区域。在文中作者共划分出21个检测区域,并且提出了一个带有21个区域标签的车辆重识别数据集,用于训练检测网络。这是第一个标记了这么多精细属性的数据集,通过这些属性,网络能够学习对车辆识别有用的车辆鉴别特征。此外,作者在构建的数据集上使用单阶段检测方法SSD作为检测的基线模型,并基于此进行了修改后取得了良好的效果。Lin<sup>[9]</sup>等人提出了一个可以进行多任务学习的双分支框架,同时进行属性学习和深度排序,以学习更具区分性的全局特征(粗粒度和细粒度),从而生成高质量的候选对象作为初始排序列表。

利用多维度信息的车辆重识别方法只能在数据集较小时使用,因为当数据集过大时其带来的标注开销也会过大。

## 2.2 基于车辆视点转换的方法

由于重识别任务进行的是跨摄像头间的图片检索,因此必须要考虑到视点转换所带来的影响,如何让网络学习到跨视点的鲁棒特征是本方法的重点。Zhou<sup>[10]</sup>等人提出了一个VAMI模型,该模型可以利用

查询图片的单视角特征,生成同一身份车辆的多视角特征。具体来说作者将对抗式训练机制和车辆解析分类器相结合,提出了一个视点感知注意模型,用于从输入图像中获取到注意力图谱,而考虑到该注意力图谱是单视点特征,作者设计了一个基于该注意力图谱的多视图生成网络以推断出不同视点的全局特征。

Zhu<sup>[11]</sup>等人提出了一个四重方向池化方法。具体而言在作者所提框架的第一阶段中,利用基本深度学习架构进行提取特征,进而在第二阶段中对特征图进行四个方向(水平、垂直、对角、反对角)的池化以获得对应方向的特征图,并将这些方向特征图进行空间归一化,串联在一起作为出车辆重识别的四重方向学习特征。

Teng<sup>[12]</sup>等人提出一个多分支网络用于处理不同视点下的车辆图片。借助一个分类器将输入图片分成前、侧、后视图三类,并提高对应分支上的加权重,当图片平行经过三个分支后特征会进行加权融合。

## 2.3 基于度量学习的方法

由于车辆重识别的任务是跨摄像头进行匹配,如何能学习到跨视图间一种鲁棒的特征表示,是该类问题的关键。该特征表示应该对显著的类内(同一身份的车辆图片)变化和较小的类间(不同身份的车辆图片)变化都具有抗干扰性。针对这一特性,Jin<sup>[14]</sup>等人设计了类内排名损失(IRL)和类间排名损失(CRL)以利用同一辆车在不同视角下的外观差异所产生的多中心信息,并提出了多中心排名损失(MCRL)来联合学习类内和类间信息。在该方法中,将车辆图片投影到一个高维空间,在该空间中一辆车的不同视图将被建模成多个视图簇,在每个簇中都有各自可优化的中心,通过优化MCRL最终属于同一车辆的同一视图将被收集到一个簇中。

在度量学习中,现有重识别方法普遍使用多任务学习的思路以提高性能。即使用softmax损失与triplet损失相结合,让两个损失函数之间相互约束使得度量学习过程更具鲁棒性。

## 2.4 利用时间-地理信息的方法

现有的车辆重识别方法普遍集中在利用车辆的外观特征进行识别,此方法在某些情况下会失效,如车辆具有同一车型、同一颜色,并且具有辨识性特征的一些小物体由于拍摄视角不好未被捕获,这会给车辆重识别任务性能的提高带来一定的困难。因此在



PROVID<sup>[15]</sup>方法中,作者认为若两辆车在摄像头下出现的时间间隔越短、地理位置越接近的时候,那么两辆车具有同一身份的概率也会越大。基于该论述提出了一个将重识别的结果进行后处理的策略,即利用时间-地理相似性进行重排的机制。该方法在某些理想环境下大幅提升了车辆再识别的性能。

但是在现实的交通环境中,车辆并不总是匀速行驶的,若某一车辆以非均匀速度行驶,或在某些摄像头盲区进行停留,上述重排机制则会对其失效。因此,Wang<sup>[16]</sup>等人将每一对摄像头下车辆出现的时间间隔以正态分布进行建模,并利用时序相似度和外观相似度,将两者以一定权重结合后的结果作为衡量车辆身份的依据。

使用时间-地理信息的车辆重识别方法在某种意义上也属于车辆多维度信息的方法,并且需要使用大量的标注信息,对数据集的要求较高,目前仅有VeRi-776能够提高该类标注。

## 2.5 实验对比

我们收集了一些经典的算法和最近的一些算法,在VeRi-776数据集上进行实验并比较(表1)。

表 1

Methods	mAP	Top1	Top5
FACT	20.0	59.7	75.3
XVGAN	24.7	60.2	77.0
FACT + Plate - SNN + STR	27.8	61.4	78.8
OIFE	48.0	65.9	87.7
VAMI	50.1	77.0	90.8
Siamese - CNN + Path - LSTM	58.3	83.5	90.0
FDA - Net	55.5	84.3	92.4
VGG + C + T + S	57.4	86.6	92.9
AAVER	61.2	89.0	94.7
RAM	61.5	88.6	94.0
MVAN	72.5	92.6	97.9
PVEN	79.5	95.6	98.4
QD - DLF	61.8	88.5	94.5
SN + (2022TNNLS)	75.7	95.1	98.1
VARID(2022TITS)	79.3	96.0	99.2
DDM(2022TITS)	53.6	72.8	53.6

## 3 总结与展望

本文针对近年来车辆重识别的发展现状和相关工作对其进行了一次归纳与总结。随着技术的发展及数据集的提出,基于深度学习的方法已经突破了基于传统方法所遭遇到的一些瓶颈。而深度学习的方法在当下技术中碰到的主要困难有:为了追求更高的性能,人们将神经网络深度化及使用的数据集规模的扩大,而这通常需要一个更长的训练时间才能够使模型收敛较好;实际应用场景远比数据集中的场景要复杂,如何提高现有的模型不同应用场景下的一个鲁棒性;在面对类间相似度高,类内差异大等问题时,各项指标无法达到较高的水平。因此对以后车辆重识别领域的发展提出以下思考:(1)如何在轻量化的网络结构上取得足够高的指标同时,提高车辆重识别任务的效果,使得该技术能够更完善地应用到嵌入式系统、边缘设备中;(2)如何模拟出现实中可能出现应用场景(如大雾、弱光等),并提出相应的解决方案,使得在这些场景下的重识别指标得到提高;(3)如何提出更为鲁棒的度量学习方法,以解决重识别任务由于类内差异大、类间相似度高和跨摄像头拍摄导致的精度下降问题。

## 参考文献

- [1] 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
- [2] 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
- [3] 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
- [4] NDOYE M, TOTTE 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
- [5] 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
- [ 6 ] 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
- [ 7 ] MENG D, LI L, LIU X, *et al.* Parsing – based view – aware embedding network for vehicle re – identification [ C ]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 7103 – 7112
- [ 8 ] ZHAO Y, SHEN C, WANG H, *et al.* Structural analysis of attributes for vehicle re – identification and retrieval [ J ]. IEEE Transactions on Intelligent Transportation Systems, 2019, 21(2): 723 – 734
- [ 9 ] LIN X, LI R, ZHENG X, *et al.* Aggregating global and local visual representation for vehicle re – identification[J]. IEEE Transactions on Multimedia, 2020, 23: 3968 – 3977.
- [ 10 ] ZHOU Y, SHAO L. Aware attentive multi – view inference for vehicle re – identification [ C ]// Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 6489 – 6498
- [ 11 ] ZHU J, ZENG H, HUANG J, *et al.* Vehicle re – identification using quadruple directional deep learning features[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 21(1): 410 – 420
- [ 12 ] TENG S, ZHANG S, HUANG Q, *et al.* Multi – view spatial attention embedding for vehicle re – identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 31(2): 816 – 827
- [ 13 ] BAI Y, LOU Y, GAO F, *et al.* Group – sensitive triplet embedding for vehicle reidentification[J]. IEEE Transactions on Multimedia, 2018, 20(9): 2385 – 2399
- [ 14 ] JIN Y, LI C, LI Y, *et al.* Model latent views with multi – center metric learning for vehicle re – identification[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 22(3): 1919 – 1931
- [ 15 ] 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
- [ 16 ] 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
- 作者简介:宁轩(1997 – ),男,宁波大学信息科学与工程学院硕士研究生,主要研究方向为人工智能、计算机视觉。■

## 关于与作者签订版权许可使用承诺书的启事

为了适应不断发展的学术期刊现代化出版与传播方式,进一步提高《数据通信》网络传播渠道的影响力,更好地为广大读者、作者、学者和科技工作者提供服务,本刊已加入国内多个数据库,通过传统纸媒及国际互联网等形式进行国内外传播。为保护论文作者与《数据通信》的合法权益,根据《中华人民共和国著作权法》,论文被《数据通信》录用后,本刊需要与被录用稿件作者签订《论文著作权许可使用承诺书》,希望广大作者予以理解、配合。