



基于注意力机制与多尺度融合学习的车辆重识别方法

潘海鹏, 王云涛, 马 淼

(浙江理工大学机械与自动控制学院, 杭州 310018)

摘 要: 在车辆重识别任务中, 通常会出现相机角度变化和场景变化等情况, 导致重识别准确率降低, 为此提出了一种基于注意力与多尺度融合学习的车辆重识别方法, 在多尺度下提取并融合浅层细节信息和深层语义信息。首先, 构造一种深度学习网络, 通过注意力机制学习车辆图像的显著性特征; 然后, 在多个尺度下对描述车辆身份的信息进行提取, 将浅层表达的细节信息和深层表达的语义信息相融合构造空间特征; 其次, 对空间特征进行分解与重组, 得到具有空间鲁棒性的局部特征, 并与全局特征融合, 构造车辆身份重识别特征; 最后, 利用该特征计算不同车辆图像间相似度, 判断是否具有相同的身份。实验结果表明: 在 VeRi-776 数据集上测试得到的 Rank-1 指标达到了 94.0%, mAP 指标达到了 72.2%, 表明该方法在相机角度变化、场景变化等情况下可以有效提高车辆重识别的准确率。

关键词: 车辆重识别; 注意力机制; 多尺度融合; 全局特征; 局部特征; 深度学习网络

中图分类号: TP391

文献标志码: A

文章编号: 1673-3851(2021)09-0657-09

Vehicle re-identification methods based on attention mechanism and multi-scale fusion learning

PAN Hai-peng, WANG Yuntao, MA Miao

(Faculty of Mechanical Engineering & Automation, Zhejiang Sci-Tech University, Hangzhou 310018, China)

Abstract: In the task of vehicle re-identification, camera angle changes and scene changes often occur, which leads to the decrease of re-identification accuracy. Therefore, a vehicle re-identification method based on attention mechanism and multi-scale fusion learning was proposed in this paper. Shallow detail information and deep semantic information were extracted and integrated at multiple scales. Firstly, a deep learning network was constructed to learn the salient features of vehicle images by the attention mechanism. Then, the information describing the identity of vehicles was extracted at multi-scale level, and then the detailed information expressed at the shallow layers and the semantic information expressed in the deep layers were integrated to construct spatial features. Secondly, the spatial features were decomposed and reorganized to obtain local features of spatial robustness, which were merged with global features to construct vehicle identity re-identification features. Finally, these features were utilized to calculate the similarity between different vehicle images to determine whether they are of the same identity. The experimental results showed that the Rank-1 index tested on the VeRi-776 dataset reached 94.0%, and the mAP index reached 72.2%, revealing that our method is able to effectively improve the accuracy of vehicle re-identification in the case of camera angle changes and scene changes.

Key words: vehicle re-identification; attention mechanism; multi-scale fusion; global feature; local feature; deep learning network

收稿日期: 2021-03-26 网络出版日期: 2021-04-27 China Academic Electronic Publishing House. All rights reserved. <http://www.cnki.net>

基金项目: 浙江省自然科学基金项目(LQ19F030014); 浙江理工大学青年创新专项(2019Q035)

作者简介: 潘海鹏(1965—), 男, 河南濮阳人, 教授, 主要从事智能检测、智能控制以及图像信息处理方面的研究。

通信作者: 马 淼, E-mail: mamiao@zstu.edu.cn

0 引言

车辆重识别属于图像检索问题,目的是在不同场景中,由某一场景摄像头拍摄的待识别车辆图像,能够在其他场景中检索出与待识别车辆具有相同身份的车辆。车辆重识别技术在视频监控^[1]、智能交通^[2]等领域有着广泛应用。由于摄像头在拍摄过程中存在光照变化、背景干扰、相机角度变化、场景变化等情况,从而导致车辆重识别的准确率受到极大的影响。车辆重识别中存在准确率较低的问题已引起国内外学者的广泛关注^[3-4]。

现有车辆重识别任务的解决方法主要采用基于手工设计的特征^[5]以及一些探测器加以辅助。例如,Cormier等^[6]将LBP特征与局部方差度量相结合构造了新的特征描述符,并应用于车辆图像的局部网格单元,得到了较好的重识别效果。Charbonnier等^[7]将磁传感器应用于车辆重识别任务中,通过传感器记录车辆的三维磁特征,提高了车辆重识别方法的效果。但是,基于手工设计的特征容易受到光照变化等因素的影响,基于传感器的方法容易受到外界的干扰,导致早期的车辆重识别准确度较低。

近年来,深度学习技术广泛应用于车辆重识别领域^[8],推动了车辆重识别技术的发展^[9],极大地提高了车辆重识别方法的准确率。相比于手工设计的特征,基于深度学习的方法不需要足够的先验知识和人工设计的复杂特征,并且对光照变化、相机角度变化、场景变化具有更好的鲁棒性。Liu等^[10]在使用外观特征和时空特征的双重约束的基础上,利用孪生网络计算车牌的相似度,并且通过高质量数据集VeRi-776进行验证,结果表明该方法有效提高了车辆重识别的准确率。Tang等^[11]使用车辆姿态估计的方法预测车辆的视角,并使用卷积神经网络学习车辆的颜色、身份和类型等信息,得到鲁棒性更好的车辆身份特征,从而提高了重识别算法的准确率。Chu等^[12]提出了VANet网络,将相同及不同视角下的图像分开处理,并通过空间内约束和跨空间约束构造损失函数,进一步提高了算法的性能。Meng等^[13]采用U-Net对车辆图像进行语义分割,将车辆的顶部、前部、后部以及侧面做出区分,提取四个面的特征进行学习,并在不同视角下进行对齐,实现车辆身份的有效识别。Khorramshahi等^[14]引入自适应关键点,通过注意力机制学习车辆的局部特征,并与浅层全局特征融合,获得了良好的重识别效果。

上述基于深度学习的方法,在处理相机角度、场景变化等情况时,存在较多问题,如:缺少对全局特征和局部特征综合学习的能力^[10,12],车辆的类型信息^[11]和车牌信息^[10]难以获取,语义分割使用的标签标注需要的人工与时间成本较高^[13],学习到的车辆外观显著性特征较少^[14]等。因此,本文建立了注意力与多尺度融合网络(Attention and multi-scale fusion net, AMSFnet),在此基础上提出了一种基于注意力与多尺度融合学习的车辆重识别方法。该方法使用注意力机制学习图像的显著性特征,利用多尺度分析融合图像的细节信息与语义信息,从而有针对性地对车辆的全局特征和局部特征进行融合学习,训练过程中无需额外的信息标注,在节约人工与时间成本的同时,提高了车辆重识别的准确率。

1 方法

本文建立了AMSFnet,通过该网络实现车辆重识别。首先,通过注意力机制学习车辆图像的显著性特征;然后,在局部特征分支中,利用空间特征重组计算得到局部特征,进而与全局特征融合得到全局-局部特征;最后,使用该全局-局部特征计算车辆间的相似度,实现车辆身份重识别。

1.1 网络结构

AMSFnet网络结构示意图如图1所示,该网络的输入图像大小为 192×192 ,由ResNet50网络框架^[15]和Squeeze and Excitation networks (SEnet)注意力机制^[16]、多尺度特征融合环节、全局特征和局部特征分支构成。本文提出的AMSFnet建立在ResNet50网络的基础上,在第5组卷积中取消了下采样^[17],在第2至第5组卷积后,分别加入由全连接层和池化层构成的SEnet注意力机制模块。多尺度特征融合环节由4个分支组成,含有相应的卷积层和平均池化层。该部分首先提取4个注意力机制模块处理后的输出,即提取4个不同尺度下的特征。随后通过相应的卷积层调整特征通道数量,并使用平均池化层进行下采样。最后将4个不同尺度的特征拼接融合得到空间特征。

在AMSFnet的局部特征分支中,对多尺度特征融合环节计算得到的空间特征进行处理,并通过空间特征重组、平均池化和重组等步骤得到局部特征。将局部特征与全局特征分支中的全局特征相融合,使用融合后的全局-局部特征实现车辆身份的重识别。

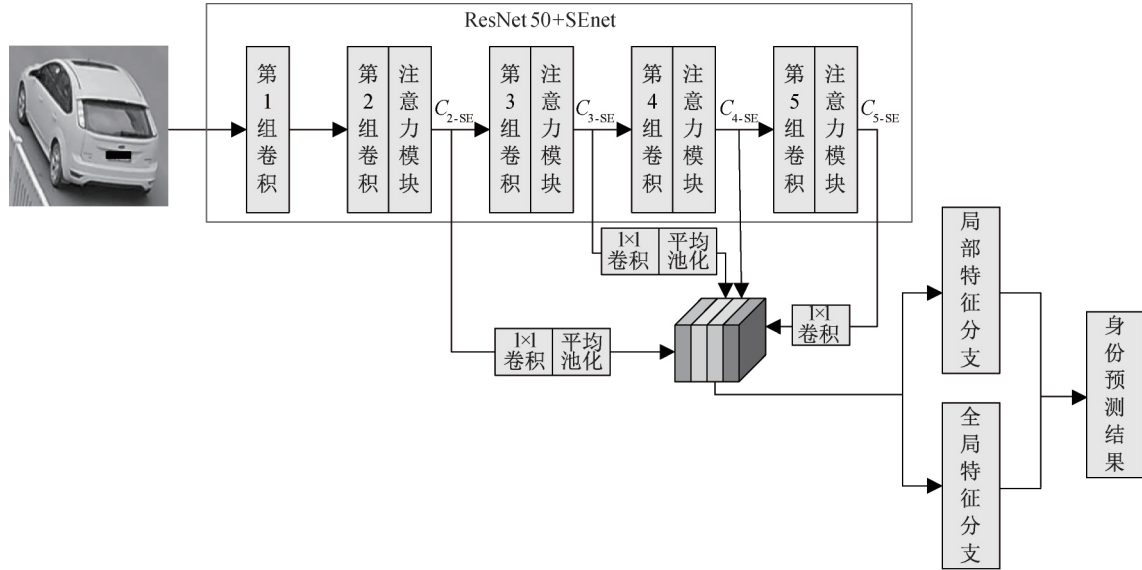


图 1 AMSFnet 网络结构示意图

1.2 注意力机制模块

本文将 SEnet 注意力机制模块添加至 AMSFnet 的第 2 组至第 5 组卷积层之后。该注意力机制模块的结构示意图如图 2 所示,其中输入特征图 F ,大小为 $C \times H \times W$ 。首先输入特征图 F ,由平均池化层计算得到 C 维大小的特征。然后该特征经过两个全连接层 FC1 与 FC2 对特征通道进行处理,全连接层 FC1 将特征的通道数量调整为 $C/16$,经过 ReLU 激活函数后输入全连接层 FC2;FC2 则将特征的通道数量回调为 C ,经过 Sigmoid 激活函数运算后得到特征图 F 的通道权重。最后将通道权重与输入特征图 F 的元素对应相乘,得到注意力机制模块的输出。SEnet 注意力机制模块的计算过程

如式(1)—(4)所示:

$$X_1 = f_{\text{GAP}}(F) \quad (1)$$

$$X_2 = \text{ReLU}(W_1 X_1) \quad (2)$$

$$X_3 = \text{Sigmoid}(W_2 X_2) \quad (3)$$

$$F' = X_3 \cdot F \quad (4)$$

其中: $F \in \mathbf{R}^{C \times H \times W}$ 和 $F' \in \mathbf{R}^{C \times H \times W}$ 分别表示注意力模块的输入与输出; f_{GAP} 表示平均池化层; $X_1 \in \mathbf{R}^{C \times 1 \times 1}$ 表示经过平均池化层处理后的输出; $W_1 \in \mathbf{R}^{C/16 \times C}$ 表示全连接层 FC1 的具体参数; $X_2 \in \mathbf{R}^{C/16 \times 1 \times 1}$ 表示经过 FC1 通道压缩和 ReLU 激活函数计算后的输出; $W_2 \in \mathbf{R}^{C \times C/16}$ 表示全连接层 FC2 的具体参数; $X_3 \in \mathbf{R}^{C \times 1 \times 1}$ 表示通道权重,是经过 FC2 通道回调和 Sigmoid 激活函数计算后的输出。

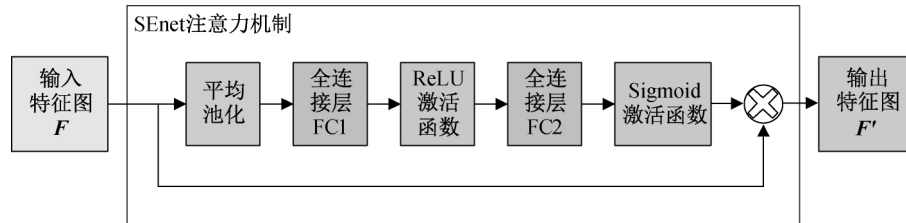


图 2 SEnet 注意力机制模块结构示意图

1.3 多尺度特征融合

多尺度特征是将网络中不同大小、不同深度的特征^[18]进行融合后得到的特征,可以有效提高网络的性能,本文基于 ResNet50 深度学习网络和 SEnet 注意力机制模块构造了多尺度特征融合环节。将第 2 组至第 5 组卷积经过 SEnet 注意力机制模块计算后的输出定义为 C_{2-SE} 、 C_{3-SE} 、 C_{4-SE} 和 C_{5-SE} 。其中 C_{2-SE} 和 C_{3-SE} 能够表征图像中更为细节的信息,而 C_{4-SE} 和 C_{5-SE} 可以表达图像更深层次的语义信息,这

些信息的结合能够有效提高网络对车辆身份的判别能力。AMSFnet 中多尺度特征融合环节分别提取 C_{2-SE} 、 C_{3-SE} 、 C_{4-SE} 和 C_{5-SE} 4 个输出特征,其大小分别为 $256 \times 48 \times 48$ 、 $512 \times 24 \times 24$ 、 $1024 \times 12 \times 12$ 和 $2048 \times 12 \times 12$ 。然后 C_{2-SE} 、 C_{3-SE} 和 C_{5-SE} 通过 1×1 大小的卷积核将特征通道数量调整为 1024,使其通道数与 C_{4-SE} 相同。此外, C_{2-SE} 和 C_{3-SE} 进一步通过平均池化层进行下采样,将特征图大小调整为 $1024 \times 12 \times 12$,使其大小与 C_{4-SE} 和 C_{5-SE} 相同。最后,将

C_{2-SE} 、 C_{3-SE} 、 C_{4-SE} 和 C_{5-SE} 拼接融合,构造车辆身份的空间特征。多尺度特征融合环节的计算过程如式(5)~(9)所示:

$$F_1 = f_{GAP}(f^{1 \times 1}(C_{2-SE})) \quad (5)$$

$$F_2 = f_{GAP}(f^{1 \times 1}(C_{3-SE})) \quad (6)$$

$$F_3 = C_{4-SE} \quad (7)$$

$$F_4 = f^{1 \times 1}(C_{5-SE}) \quad (8)$$

$$F_s = g(F_1, F_2, F_3, F_4) \quad (9)$$

其中: $f^{1 \times 1}$ 表示卷积核大小为 1×1 的卷积层,作用是调整特征通道数量; F_1 和 F_2 表示经过卷积层和平均池化层计算后的特征; F_3 即 C_{4-SE} ; F_4 表示经过卷积层计算后的特征,大小均为 $1024 \times 12 \times 12$; g 表示拼接融合; F_s 表示空间特征,大小为 $4096 \times 12 \times 12$ 。

1.4 空间特征重组

全局特征分支与局部特征分支的结构示意图如图3所示,其中空间特征由多尺度特征融合环节计

算得到,大小为 $4096 \times 12 \times 12$ 。全局特征分支首先使用平均池化将空间特征转化为4096维全局特征。然后该全连接层FC1将该全局特征的特征维数调整为256。最后该全局特征通过全连接层FC2将特征维数放大,得到576维全局特征(576为训练集车辆身份的数量),全局特征分支的计算过程如式(10)~(12)所示:

$$Y_1 = f_{GAP}(F_s) \quad (10)$$

$$Y_2 = W_3 Y_1 \quad (11)$$

$$F_g = W_4 Y_2 \quad (12)$$

其中: Y_1 表示大小为4096维的全局特征; $W_3 \in \mathbf{R}^{256 \times 4096}$ 表示全连接层FC1的具体参数; Y_2 表示经过FC1缩小特征维数后的256维全局特征; $W_4 \in \mathbf{R}^{576 \times 256}$ 表示全连接层FC2的具体参数; F_g 表示经过FC2放大特征维数后的576维全局特征。

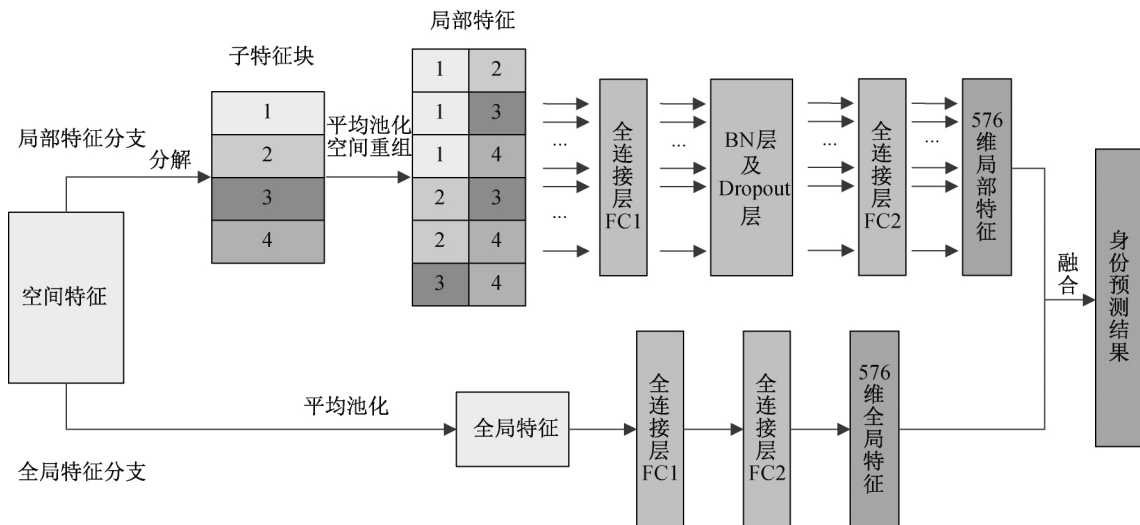


图3 全局特征分支与局部特征分支结构示意图

局部特征分支首先将空间特征分解为4个子特征块,每个子特征块的大小为 $4096 \times 3 \times 12$ 。然后每个子特征块经过平均池化转化成4个大小为4096维的子空间特征,从上至下依次记为1、2、3、4,并将子空间特征两两重组,顺序为(1,2)、(1,3)、(1,4)、(2,3)、(2,4)、(3,4),共6个8192维局部特征。随后将6个局部特征分别输入全连接层FC1,调整为256维的局部特征。此外,为了避免数据过拟合,局部特征分支在全连接层FC1之后使用了Batch Normalization层(BN层)^[19]和Dropout层。

最后,6个256维局部特征分别经过全连接层FC2计算后,得到大小为576维的局部特征(576为训练集车辆身份的数量),局部特征分支的计算过程如式(13)~(16)所示:

$$Z_i = f_{GAP}(F_d), d=1,2,3,4 \quad (13)$$

$$Z_L^k = c(Z_i, Z_j), k=1,2,\dots,6, i < j \quad (14)$$

$$T_L^k = \delta(\beta(W_5 Z_L^k)), k=1,2,\dots,6 \quad (15)$$

$$F_L^k = W_6 T_L^k, k=1,2,\dots,6 \quad (16)$$

其中: F_d 表示空间特征分解后的4个子特征块,大小为 $4096 \times 3 \times 12$; Z_i 表示子特征块经过平均池化计算得到4个4096维子空间特征; c 表示重组操作; Z_L^k 表示8192维局部特征; $W_5 \in \mathbf{R}^{256 \times 8192}$ 表示全连接层FC1的具体参数; β 表示Batch Normalization层; δ 表示Dropout层; T_L^k 表示经过全连接层FC1以及防止过拟合处理后的输出; $W_6 \in \mathbf{R}^{576 \times 256}$ 表示全连接层FC2的具体参数; F_L^k 表示FC2计算得到的576维局部特征。

车辆身份识别结果的计算公式如式(17)所示,

本文通过融合两个分支计算得到的 576 维全局特征和局部特征识别车辆身份结果：

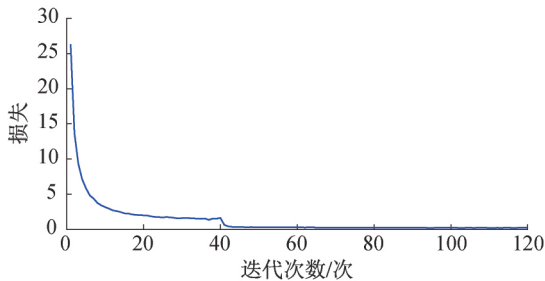
$$ID_{\text{vehicle}} = \max(\sigma(\mathbf{F}_g) + \sum_{k=1}^6 \sigma(\mathbf{F}_L^k)) \quad (17)$$

其中： σ 表示归一化指数函数； ID_{vehicle} 表示车辆身份的识别结果。

2 实验

2.1 数据集

为了验证提出方法的有效性,本文在公开的车辆重识别通用数据集 VeRi-776^[10] 上进行了验证。VeRi-776 数据集由 20 个摄像头拍摄的 776 辆汽车共 51035 张图像组成,分为训练集 D_t 、查询集 D_r 与测试集 D_a 。该数据集的训练集 D_t 由 576 辆汽车组成,共 37778 张图像,用于 ASMFnet 网络参数训练;查询集 D_r 与测试集 D_a 均由 200 辆汽车组成,其中 D_r 有 1678 张图像, D_a 有 11579 张图像。在验证本文提出的车辆重识别方法有效性的实验中,依次取查询集 D_r 中的车辆图像,并与测试集 D_a 中的图像库进行匹配,检索 D_a 中所有与被查询车辆具有相同身份的车辆图像。



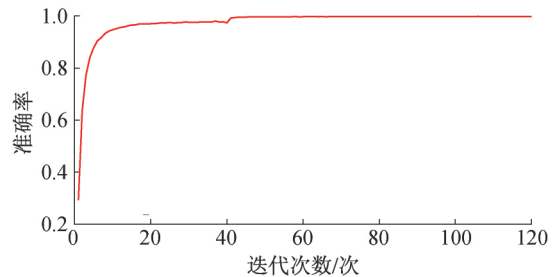
(a) 训练过程中的损失变化曲线

2.2 实验参数与评价指标

ASMFnet 的训练与验证实验基于深度学习框架 PyTorch 实现,采用 RTX2080 进行 GPU 加速,CPU 为 I7-9700KF@3.6 GHz。在训练过程的主要参数见表 1,由于实验使用性能普通的台式机,因此 batch size 大小受限,设为 32;实际实验中发现在迭代次数大于 60 时,训练过程中的损失可趋近于 0,准确率可趋近于 1,如图 4(a)和图 4(b)所示,因此,实验中采用迭代次数 120;全连接层的冗余信息较多,可使用较大的学习率,设为 0.05,卷积层的学习率根据经验设为 0.005。训练集 D_t 的图像输入大小调整为 192×192 ,学习率每经过 40 次迭代调整为原先的 0.1 倍,损失函数采用交叉熵损失函数,数据增强方式采用随机擦除^[20],水平翻转和随机裁剪。

表 1 实验中的主要参数

参数	数值
batch size	32
训练迭代次数/次	120
全连接层初始学习率	0.05
卷积层初始学习率	0.005



(b) 训练过程中的准确率变化曲线

图 4 训练过程中损失及准确率的变化

本文使用车辆重识别领域中的两个常用指标对算法性能进行评价:累计匹配特征(CMC)曲线和平均精度均值(mAP)^[21]。其中,累计匹配特征曲线指标是指重识别结果中置信度最高(最靠前)的 K 张图有正确结果的概率,记为 Rank- K 。通常使用 Rank-1 作为主要评价指标,即第一个重识别结果就匹配成功的概率 R_1 ,本文同样采用 Rank-1 评价算法性能,计算过程如式(18)——(19)所示:

$$N_i = \begin{cases} 1, & \text{第一个重识别结果正确,} \\ 0, & \text{第一个重识别结果错误} \end{cases} \quad (18)$$

$$R_1/\% = \frac{\sum_{i=1}^m N_i}{m} \times 100 \quad (19)$$

其中: R_1 表示 Rank-1 算法性能评价结果; i 表示待

识别车辆的索引; N_i 表示第 i 个待识别车辆的 Rank-1 结果; m 表示待识别车辆的数量。平均精度均值 mAP 是指计算所有样本准确率-召回率曲线下的面积平均值,该指标能够准确反映车辆重识别结果中所有正确结果图像排名靠前的程度。平均精度均值首先需要计算每一个待识别车辆的平均精度 P_a ,平均精度可以用式(20)表示:

$$P_a/\% = \int_0^1 P(R) dR \times 100 \quad (20)$$

其中: $P(R)$ 表示召回率为 R 时的准确率。然后计算平均精度的均值得到 P_m 如式(21)所示:

$$P_m/\% = \frac{1}{m} \sum_{i=1}^m P_a^i \times 100 \quad (21)$$

其中: i 表示待识别车辆的索引。

2.3 实验测试过程

在对比与验证实验中,对查询集 D_q 中的所有车辆图像,提取全局特征分支与局部特征分支中全连接层 FC1 的输出特征,使用欧氏距离进行相似度计算,依次与测试集 D_a 中的所有车辆图像进行相似度计算。其中,全连接层 FC1 的输出特征为 256 维,共有 1 个全局特征和 6 个局部特征,因此,每张图像提取出的特征为 1792 维特征。待相似度计算完毕后依据相似程度高低对测试集 D_a 中的图像进行排序,最终计算出 Rank-1 与 mAP 指标。在本文

的实验环境下,整个测试过程约需 140 s。

2.4 实验结果

图 5 为本文方法在 VeRi-776 数据集^[10]上的车辆重识别效果。图 5 中第一列为待识别车辆,后十列为相似度排序最高的前 10 张车辆图像,即 Rank-1 至 Rank-10 的重识别结果。其中,图 5(a)与图 5(b)为 Rank-1 至 Rank-10 的车辆身份重识别均正确的实例展示;图 5(c)为重识别结果存在错误的实例,其中有方框标注的为错误结果,其余为正确结果。

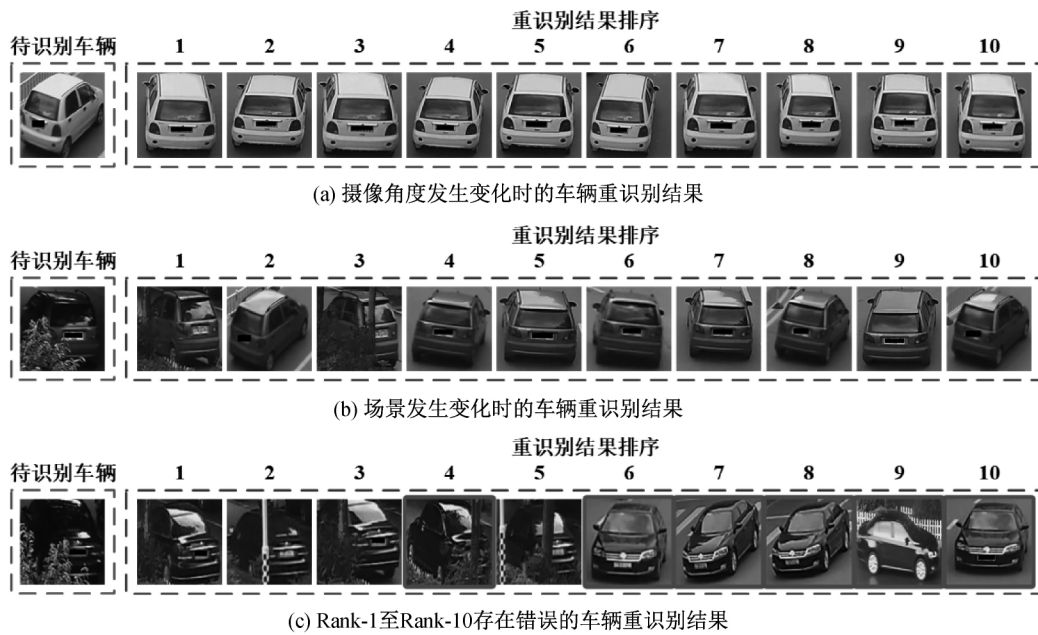


图 5 VeRi-776 数据集的车辆重识别结果

图 5(a)与图 5(b)分别展示了在摄像角度和场景发生变化的情况下,本文提出的方法仍能够准确地重识别车辆图像。但是,在场景较为复杂、遮挡严重及过度曝光等情况下的车辆重识别任务仍然存在很大的挑战性。例如,图 5(c)中待识别车辆场景复杂且存在遮挡及过度曝光,给重识别任务造成了很大的干扰,其中第 4 个查询结果与待识别车辆的外观以及所处环境高度相似,第 6 至第 10 个查询结果与待识别车辆均为黑色外观,导致车辆重识别产生错误。由于图 5(c)中识别车辆的第 1 个查询结果正确,因此该情况下出现的重识别错误不会使 Rank-1 指标下降,但是会导致实验评价指标 mAP 结果降低。

2.5 对比实验与分析

为证明提出的方法能够有效提高车辆重识别的准确率,在 VeRi-776 数据集上,与近几年车辆重识别领域中的几种前沿方法进行了比较,结果见表 2。

表 2 在 VeRi-776 数据集上的对比实验

方法	$R_1/\%$	$P_m/\%$
Siamese+Path	83.5	58.3
AAVER	89.0	61.2
VANet	89.8	66.3
PAMTRI	90.9	65.4
AMSFnet(本文方法)	94.0	72.2

对比方法中, Siamese + Path 方法^[22] 在 ResNet50 的基础上构建孪生网络判断车辆图像间的外观相似度,并利用链马尔可夫模型对路径建模生成车辆图像间的时空相似度,在外观和时空条件的约束下实现车辆重识别。AAVER 方法^[14] 使用 ResNet50 提取车辆的全局特征,并引入自适应关键点作为注意力机制学习车辆的局部特征。VANet 方法^[12] 将相同视角下和不同视角下的图像分开处理,并利用空间内约束和跨空间约束构造损失函数,有效提高了车辆重识别算法的性能。PAMTRI 方

法^[11]使用姿态估计的方法预测车辆的视角,并使用卷积神经网络学习车辆的颜色、身份和类型信息。

本文提出的方法在 VeRi-776 数据集上测试得到的结果见表 2, R_1 为 94.0%, P_m 为 72.2%。而 Siamese+Path 方法使用孪生网络计算车辆外观相似度,缺少对全局与局部特征的全面分析,因此本文方法与其相比,Rank-1 提高了 10.5%, mAP 提高了 13.9%。此外,AAVER 方法虽然将浅层全局特征与局部特征融合,但是其基于自适应关键点的注意力机制只能学习到较少的显著性特征,而本文方法利用 SEnet 注意力机制学习车辆的显著性特征,与 AAVeR 相比能够获取更多的细节信息,有利于车辆的身份重识别,因此本文方法的 Rank-1 提高了 5.0%, mAP 提高了 11.0%。表 2 中 VANet 的方法设计了空间内和跨空间损失函数,有效学习了相同视角下和不同视角下的图像特征,但只改进损失函数很难做到对图像特征的全面学习,因此与其相比本文方法的 Rank-1 高 4.2%, mAP 高 5.9%。而 PAMTRI 引入姿态估计辅助车辆特征的学习,并额外加入车辆的颜色与类型标签帮助车辆身份分类,但是由于没有对多尺度特征进行有效利用,忽略了不同尺度下的细节信息,所以本文方法的 Rank-1 比其高 3.1%, mAP 比其高 6.8%。

从表 2 的数据以及实验结果分析中可以看到,本文方法与近几年的前沿方法相比,无需额外的车辆标签信息,降低了车辆重识别成本,并通过注意力机制与多尺度融合的方法显著提高了车辆重识别的准确率。

2.6 注意力机制有效性验证实验

本文在 VeRi-776 数据集上进行注意力机制有效性验证实验,实验中将 ASMFnet 去掉注意力机制后与 ASMFnet 进行对比,结果见表 3。表中将去掉注意力机制后的网络记为 MSFnet。

表 3 注意力机制有效性验证实验

方法	$R_1/\%$	$P_m/\%$
MSFnet	94.0	70.9
AMSFnet(本文方法)	94.0	72.2

由表 3 可见,AMSFnet 与 MSFnet 相比,Rank-1 相等, mAP 提高了 1.3%。该结果证明了 SEnet 注意力机制可以有效学习车辆身份的显著性特征,使重识别结果中所有正确结果图像的排名更为靠前,从而提高了车辆身份重识别的准确率。

2.7 多尺度特征融合的分支数量有效性验证实验

由于多尺度特征融合分支数量会对算法性能产

生影响,本文在 VeRi-776 数据集上进行了使用不同分支数量的对比实验,验证本文提出的 AMSFnet 网络使用四分支方案的有效性。在本文所构造的全局特征分支和局部特征分支结构基础上,改变多尺度特征融合环节中的分支数量。共分 4 种情况进行对比实验:a) C_{5-SE} 单分支,将 C_{5-SE} 作为空间特征输入全局与局部特征分支;b) $C_{4-SE} + C_{5-SE}$ 双分支,将 C_{4-SE} 和 C_{5-SE} 作为空间特征输入全局与局部特征分支;c) $C_{3-SE} + C_{4-SE} + C_{5-SE}$ 三分支,将 C_{3-SE} , C_{4-SE} 和 C_{5-SE} 作为空间特征输入全局与局部特征分支;d) $C_{2-SE} + C_{3-SE} + C_{4-SE} + C_{5-SE}$ 四分支(即本文提出的 AMSFnet 网络,如图 1 所示),即将 C_{2-SE} , C_{3-SE} , C_{4-SE} 和 C_{5-SE} 作为空间特征输入全局与局部特征分支。

多尺度特征融合的分支数量有效性验证实验结果见表 4。其中,方法 1 多尺度特征融合环节为单分支模式时,实验得到的 R_1 为 93.5%, P_m 为 70.7%。方法 2 为双分支模式,其在方法 1 的基础上加入了 C_{4-SE} 特征,Rank-1 和 mAP 均提高了 0.2%,原因是方法 2 融入了更多的深层语义信息,有助于加强深度学习网络的判别能力。

表 4 多尺度特征融合的分支数量有效性验证实验

序号	方法	$R_1/\%$	$P_m/\%$
1	C_{5-SE}	93.5	70.7
2	$C_{4-SE} + C_{5-SE}$	93.7	70.9
3	$C_{3-SE} + C_{4-SE} + C_{5-SE}$	93.9	72.0
4	$C_{2-SE} + C_{3-SE} + C_{4-SE} + C_{5-SE}$ (AMSFnet)	94.0	72.2

由表 4 可见,方法 3 为三分支模式,与方法 2 相比,在其基础上加入了 C_{3-SE} 特征,Rank-1 提高了 0.2%, mAP 提高了 1.1%,主要原因是 C_{3-SE} 中含有浅层较为细节的信息,有助于网络学习图像的局部细节特征;与方法 1 相比,方法 3 的 Rank-1 提高了 0.4%, mAP 提高了 1.3%,证明了细节信息的加入能够有效学习车辆图像的局部细节特征,从而提高了车辆重识别的准确率。

此外,考虑到 C_{2-SE} 对应着车辆图像更为原始的浅层细节信息,所含局部细节信息更加丰富,如果将其加入多尺度特征融合环节,可以使空间特征所含浅层细节信息与深层语义信息更为均衡,可以更好地提高网络的判别能力。因此,方法 4 四分支模式在方法 3 上的基础上加入了 C_{2-SE} 特征,Rank-1 提高了 0.1%, mAP 提高了 0.2%,这表明丰富的细节信息与语义信息可以进一步提高车辆重识别的准确率。

3 结 论

本文设计了一种基于注意力与多尺度融合学习的车辆重识别网络,将 ResNet50 网络、SEnet 注意力机制和多尺度结合,通过 SEnet 注意力机制有效挖掘车辆图像的显著性特征,并在多个尺度下提取并融合浅层细节信息和深层语义信息,有效提高了车辆重识别网络的判别能力。本文提出的方法在 VeRi-776 数据集上进行了验证实验,Rank-1 指标达到了 94.0%,mAP 指标达到了 72.2%,与现有车辆重识别方法相比有明显优越性。本文所述方法在一定程度上解决了相机角度、场景变化等情况对车辆重识别带来的影响,提高了车辆重识别的准确率。

参考文献:

- [1] 张强,李嘉锋,卓力. 车辆识别技术综述[J]. 北京工业大学学报, 2018, 44(3):382-392.
- [2] Wang H B, Hou J Y, Chen N. A survey of vehicle re-identification based on deep learning[J]. IEEE Access, 2019, 7: 172443-172469.
- [3] 邱铭凯, 李熙莹. 用于车辆重识别的基于细节感知的判别特征学习模型[J/OL]. 中山大学学报(自然科学版): 1-10. (2021-03-16) [2021-04-06]. <https://doi.org/10.13471/j.cnki.acta.snus.2020.03.16.2020B023>.
- [4] Xie Y, Zhu J Q, Zeng H Q, et al. Learning matching behavior differences for compressing vehicle re-identification models [C]// IEEE International Conference on Visual Communications and Image Processing, Macau, China: IEEE, 2020: 523-526.
- [5] 刘凯, 李滢东, 林伟鹏. 车辆再识别技术综述[J]. 智能科学与技术学报, 2020, 2(1):10-25.
- [6] Cormier M, Sommer L W, Teutsch M. Low resolution vehicle re-identification based on appearance features for wide area motion imagery [C]// IEEE Winter Applications of Computer Vision Workshops, Lake Placid, NY, USA: IEEE, 2016: 1-7.
- [7] Charbonnier S, Pitton A C, Vassilev A. Vehicle re-identification with a single magnetic sensor[C]// IEEE International Instrumentation and Measurement Technology Conference Proceedings, Graz, Austria: IEEE, 2012: 380-385.
- [8] He B, Li J, Zhao Y F, et al. Part-regularized near-duplicate vehicle re-identification [C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA: IEEE, 2019: 3992-4000.
- [9] Liu H Y, Tian Y H, Wang Y W, et al. Deep relative distance learning: tell the difference between similar vehicles[C]// IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA: IEEE, 2016: 2167-2175.
- [10] Liu X C, Liu W, Mei T, et al. A deep learning-based approach to progressive vehicle re-identification for urban surveillance [C]// European Conference on Computer Vision (ECCV 2016). Berlin, Germany: Springer, 2016: 869-884.
- [11] Tang Z, Naphade M, Birchfield S, et al. PAMTRI: Pose-aware multi-task learning for vehicle re-identification using highly randomized synthetic data [C]// IEEE/CVF International Conference on Computer Vision, Seoul, Korea: IEEE, 2019: 211-220.
- [12] Chu R H, Sun Y F, Li Y D, et al. Vehicle re-identification with viewpoint-aware metric learning[C]// IEEE/CVF International Conference on Computer Vision, Seoul, Korea: IEEE, 2019: 8281-8290.
- [13] Meng D C, Li L, Liu X J, et al. Parsing-based view-aware embedding network for vehicle re-identification [C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA: IEEE, 2020: 7101-7110.
- [14] Khorramshahi P, Kumar A, Peri N, et al. A dual-path model with adaptive attention for vehicle re-identification [C]// IEEE/CVF International Conference on Computer Vision, Seoul, Korea: IEEE, 2019: 6131-6140.
- [15] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]// IEEE Conference on Computer Vision and Pattern Identification, Las Vegas, NV, USA: IEEE, 2016: 770-778.
- [16] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.
- [17] Sun Y F, Zheng L, Yang Y, et al. Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline) [C]//European Conference on Computer Vision, Berlin, Germany: Springer, 2018: 480-496.
- [18] Wang Z K, He L H, Gao X B, et al. Multi-scale spatial-temporal network for person re-identification [C]//2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2019). Brighton, UK: IEEE, 2019: 2052-2056.
- [19] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate

- shift [EB/OL]. (2015-03-2) [2021-3-03]. <https://arxiv.org/abs/1502.03167>.
- [20] Zhong Z, Zheng L, Kang G L, et al. Random erasing data augmentation [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34 (7): 13001-13008.
- [21] Zhong Z, Zheng L, Cao D L, et al. Re-ranking person re-identification with k-reciprocal encoding[C]// IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, 2017: 3652-3661.
- [22] Shen Y T, Xiao T, Li H S, et al. Learning deep neural networks for vehicle re-id with visual-spatio-temporal path proposals[C]// IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017: 1918-1927.

(责任编辑:康 锋)