

DOI: 10.13382/j.jemi.B2205756

多视图融合和全局特征增强的车辆重识别网络*

孙伟^{1,2} 徐凡¹ 张小瑞^{2,3,4,5} 胡亚华¹ 赵宇煌¹

(1. 南京信息工程大学自动化学院 南京 210044; 2. 南京信息工程大学江苏省大气环境与装备技术协同创新中心 南京 210044; 3. 南京信息工程大学数字取证教育部工程研究中心 南京 210044; 4. 南京信息工程大学无锡研究院 无锡 214100; 5. 南京信息工程大学计算机与软件学院 南京 210044)

摘要: 车辆重识别是智能交通领域重要应用之一, 现有的车辆识别方法大多集中于预定义的局部区域特征或全局外观特征。然而, 在复杂的交通环境下, 传统的方法难以获取预定义的局部区域, 同时很难捕捉有价值的车辆全局特征信息。因此, 本文提出一种具有多视图融合的混合注意力机制和全局特征增强的端到端双分支网络。该网络旨在通过增强车辆的特征表达能力和特征质量来获得更完整、更多样的车辆特征。本文通过视图解析网络对车辆图片4个视角的视图进行分割, 并通过视图拼接方法缓解分割不准确导致的信息丢失问题。为了更好地突出拼接视图中的显著性局部区域, 本文提出一种由通道注意力机制和自注意力机制组成的混合注意力模块。通过该模块从车辆拼接视图中分别获取关键局部信息和局部信息之间的相关性, 更好地凸显拼接视图中车辆局部的细节信息。除此之外, 还提出了一个全局特征增强模块, 通过池化和卷积获得全局特征的空间和通道关系。该模块不仅能提取到语义增强的车辆特征, 而且还使车辆特征中包含完好的细节信息, 解决获取的车辆图像受视角变化、光照条件变化等因素的影响。在 Veri-776 和 VehicleID 数据集上的大量实验表明, mAP、CMC@1 和 CMC@5 分别达到了 82.41%、98.63% 和 99.23%, 优于现有的方法。

关键词: 车辆重识别; 视图分割; 视图拼接; 注意力机制; 特征增强

中图分类号: TP391; TN919.8 **文献标识码:** A **国家标准学科分类代码:** 510.4

Vehicle re-identification network with multi-view fusion and global feature enhancement

Sun Wei^{1,2} Xu Fan¹ Zhang Xiaorui^{2,3,4,5} Hu Yahua¹ Zhao Yuhuang¹

(1. College of Automation, Nanjing University of Information Science & Technology, Nanjing 210044, China; 2. Jiangsu Collaborative Innovation Center on Atmospheric Environment and Equipment Technology, Nanjing University of Information Science & Technology, Nanjing 210044, China; 3. Engineering Research Center of Digital Forensics, Ministry of Education, Nanjing University of Information Science & Technology, Nanjing 210044, China; 4. Wuxi Research Institute, Nanjing University of Information Science & Technology, Wuxi 214100, China; 5. College of Computer and Software, Nanjing University of Information Science & Technology, Nanjing 210044, China)

Abstract: Vehicle re-identification is one of the important applications in the field of intelligent transportation. Most of the existing vehicle Re-ID methods focus on pre-defined local area features or global appearance features. However, in the complex traffic environment, it is difficult for traditional methods to acquire pre-defined local regions, and it is difficult to capture valuable vehicle global feature information. Therefore, an end-to-end dual-branch network with multi-view fusion hybrid attention mechanism and global feature enhancement is proposed. The network aims to obtain more complete and diverse vehicle features by enhancing the feature representation ability and feature quality of vehicles. This paper uses the view parsing network to segment the four views of the vehicle image, and uses the view stitching method to alleviate the problem of information loss caused by inaccurate segmentation. To better

收稿日期: 2022-08-17 Received Date: 2022-08-17

* 基金项目: 国家自然科学基金(61304205, 62272236)、江苏省自然科学基金(BK20191401, BK20201136)、江苏省研究生科研与实践创新计划项目(SJCX21_0363)、大学生创新创业训练项目(XJDC202110300601, 202010300290, 202010300211, 202010300116E)资助

highlight salient local regions in stitched views, this paper proposes a hybrid attention module consisting of a channel attention mechanism and a self-attention mechanism. Through this module, the correlation between the key local information and the local information is obtained from the stitched view respectively, so as to better highlight the detailed information of the vehicle part in the stitched view. Besides, this paper also proposes a global feature enhancement module to obtain the spatial and channel relationship of global features through pooling and convolution. This module not only extracts the semantically enhanced vehicle features, but also makes the vehicle features contain complete detailed information, and solve the influence of the acquired vehicle images by factors such as changes in viewing angle and lighting conditions. Extensive experiments in this paper on the Veri-776 and VehicleID datasets show that mAP, CMC@1, and CMC@5 reach 82.41%, 98.63%, and 99.23%, respectively, which is better than the existing methods.

Keywords: vehicle re-identification; view segmentation; view stitching; attention mechanism; feature enhancement

0 引言

在给定车辆身份查询图像的情况下,车辆重识别任务的目标是从大型摄像机网络捕获的图像数据库中检索该身份的所有图像。由于其在智能交通、公共安全等领域广泛应用,引起了计算机视觉界的极大关注。

随着深度学习和大数据集的发展,车辆重识别方法取得了显著的成功。然而,在真实的交通场景下车辆面临复杂的视角变化,这使得同一车辆在不同视图下类内差异大,不同车辆由于相同的颜色和车型,在特定视图下类间差异小。在最近的研究中,研究人员发现车辆的局部区域往往包含辨识度更高的特征,如品牌标志、车窗、车灯和内饰等,这就要求 Re-ID 模型能够更多地关注显著性局部区域。例如,Wang 等^[1]标记了车辆的 20 个关键点,以提取车辆的局部特征,基于标记的关键点检测生成了方向不变特征。Zhang 等^[2]则采用相似的目标检测方法从车辆图像中获取车灯、年检标志、装饰等局部区域,利用这些局部部件对车辆进行重识别。然而,由于这些关键特征点和局部区域的位置被事先确定,无法适应车辆外观在复杂交通环境下多变的情况。

为了更好地聚焦这些显著性局部区域,Meng 等^[3]利用车辆部件分割模型将整个车辆图像依据不同的视角分割成不同视图区域,然后通过分割后的每个视图区域进行计算来缓解重识别模型对视角变化的依赖。但是这又带来一个新的问题,如模型分割后的车辆视图区域可能不精确,这会导致车辆关键部件区域分割不完整和特征匹配不准确,严重影响车辆重识别的结果。

进一步的研究发现,车辆的全局特征,如车辆颜色^[3]、车型^[4-5]、方向^[6-7]和时空信息^[8-9]等,能够弥补车辆局部特征的不足。但这些全局特征依赖清晰的拍摄图像。在真实交通环境中,由于摄像头视角变化、光照条件变化等因素的影响,导致采集到的车辆图像模糊失真、可用样本少,使得传统方法很难捕捉到有价值的车辆全局特征信息。例如,Sun 等^[5]提出了一种具有局部感知能力和特征互补学习的三支嵌入网络(TBE-Net)。在 TBE-Net 中,全局分支学习车辆的全局外观特征,局部分

支学习具有细微差别的局部特征,两者相互配合来增强车辆特征表示。Chu 等^[10]提出了一种新的车辆特征匹配算法,利用预先训练好的视点分类器计算车辆全局视点,将车辆全局特征分为两个特征空间,并采用不同的损失函数进行训练。虽然这些方法的识别准确率都有了很大的提高,但车辆全局特征容易受分辨率、光照变化和遮挡等因素的影响,使得在真实交通环境中车辆重识别效果较差。因此,如何提高模型的全局特征表达能力同样是车辆重识别面临的挑战。

综上所述,在复杂交通环境下,车辆视角和光照条件变化剧烈,导致传统的方法难以获取有用的全局和局部信息。因此,本文提出一种多视图融合注意力机制和全局特征增强的车辆重识别方法。为了防止分割不精确导致视图信息丢失和特征匹配不准确的问题,将高频出现的侧面视图与其他 3 个视图拼接,将拼接后的视图看作一个整体进行模型训练。通过一个混合注意力模块在拼接后视图图像中自适应地寻找关键局部区域并为关键局部区域赋予较大的权重,让模型更多地关注具有区别性线索的局部区域。作为局部区域特征的有益补充,通过特征增强模块深度挖掘全局特征在空间和通道上的关系,以此增强车辆全局特征的表达能力。

1 方法

如图 1 展示了本文所提出的网络的结构。该网络由两个分支组成,分别为局部分支和全局分支。在局部分支中,通过 U-Net 获取输入车辆的前、后、顶和侧 4 个视图信息。为了缓解车辆视图分割不精确导致的视图信息丢失的问题,将获取的侧面视图分别与其他 3 个视图进行拼接。将拼接后的视图与增强后的全局特征相乘获得一系列的拼接视图特征,再通过混合注意力模块自适应的为拼接视图特征的关键区域分配较大的权重。这使得提取的车辆特征最大化的保留具有辨别性的信息。在全局分支中,通过车辆特征提取网络获取输入车辆的全局特征,并将全局特征输入到由空间关系分支和通道关系分支组成的全局特征增强模块,通过池化和卷积挖掘全局特征在空间和通道关系,利用特征融合对两个分支的

输出进行融合,以此获得增强后的全局特征。最后,本文采用全局平均池化提取特征,将特征图显著信息汇集到

一起,结合 ID 损失和三元组损失对网络进行训练,得到更加丰富的车辆特征。

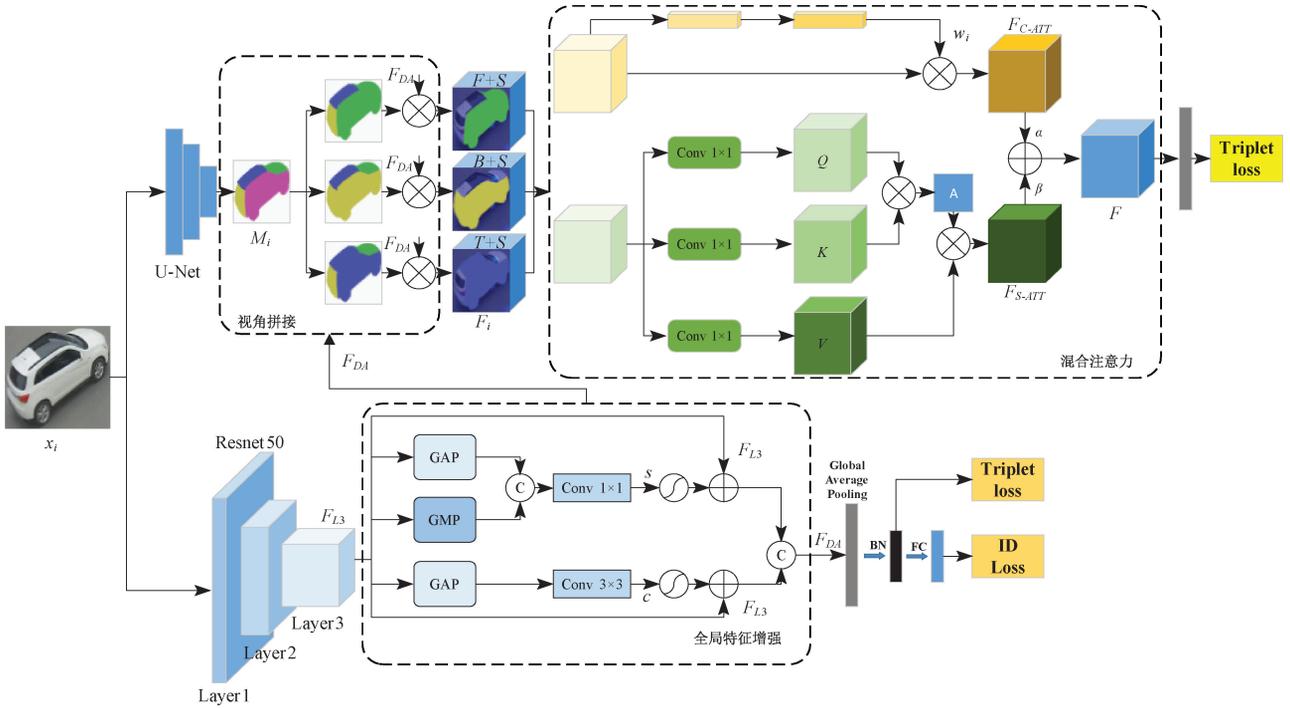


图 1 多视图融合和全局特征增强的车辆重识别网络

Fig. 1 Vehicle Re-ID network with multi-view fusion and global feature augmentation

1.1 视图分割和拼接

受文献[3]的启发,本文根据车辆是刚体、没有物理形变的特点,可以按不同的视图将其划分为不同的区域。由于车辆可以看作左右两侧相同的立方体,拟将车辆分割成前、后、顶和侧 4 个视图面,如图 2 中视角拼接所示。



图 2 可视化注意力图

Fig. 2 Visual attention map

首先,标注拍摄到的部分车辆进行的前(F)、后(B)、顶(T)和侧(S) 4 个视图信息。利用这些数据去训练视图获取网络,并用训练好的网络分割每个车辆的 4 个视图区域 $\{M_i | 0, 1, 2, 3\}$ 。其中,0, 1, 2, 3 分别代表前、后、顶和侧。通过视图拼接得到车辆的前+侧($F+S$)、后+侧($B+S$)和顶+侧($T+S$) 3 个部分,

$$P_k = M_i + M_3 (i, k = 0, 1, 2) \quad (1)$$

其中, $M_i (i = 0, 1, 2)$ 分别表示前视图、后视图和顶

视图, M_3 为侧视图, P_k 则表示拼接后的视图表示。然后,将增强后的全局特征 F_{DA} 送入视图拼接模块中,与拼接后的视图 P_k 进行点乘去获取拼接视图特征 F_i , 获取的拼接视图特征可表示为:

$$F_i = P_k \cdot F_{DA} (i = 0, 1, 2) \quad (2)$$

其中, \cdot 表示基于元素的乘积, 0, 1, 2 分别表示前+侧、后+侧和顶+侧。

1.2 局部特征学习

本文提出一种混合注意力模块,该模块由一个通道注意力机制和一个自注意力机制组成。

Channel-attention^[11]: C-ATT 专注于车辆拼接视图中完整的车辆部件信息,这可以更加全面的获取重要的局部特征。因此,本文利用 C-ATT 模块依靠注意力机制自适应的学习注意力分数 ω_i , 通过给局部区域分配不同的权重来表明他们的重要性。对车辆的拼接视图特征 F_i 进行加权得到注意力图 F_{C-ATT} ,

$$\omega_i = \delta(\text{GAP}(F_i), \vartheta) \quad (3)$$

$$F_{C-ATT} = \omega_i F_i \quad (4)$$

其中, GAP 表示对拼接视图 $F_i \in R^{W \times H \times C}$ 应用全局平均池操作来获得全局感受野,并通过可学习的参数 ϑ 为每个特征通道生成权重 ω_i 。而 $\delta(\cdot)$ 是用于突出显著性局部区域特征的可学习函数。将得到的结果和原拼接视图特征相乘,为每一个通道赋予不同的权重得到 F_{C-ATT} 。

Self-attention^[12]: 自注意力机制通过计算特征中任意两个像素的相关性并根据相关性大小增强特征信息 [10]。给定形状为 (H, W, C) 的输入张量, 自注意力机制首先将其调整为两个形状分别为 $(H \times W, C)$ 和 $(C, H \times W)$ 的矩阵。然后, 利用矩阵乘法得到形状为 (H, W) 的空间相关性矩阵。随后根据空间相关性矩阵对特征内的关键区域进行凸显。因此, 利用自注意力机制捕捉拼接视图特征的内部重要信息并输出注意力图。

将拼接视图 $F_i \in R^{W \times H \times C}$ 输入到自注意力模块中, 分别经过 3 个 1×1 的卷积将特征投影到更深的空间得到 Q (Query)、 K (Key) 和 V (Value) 的投影矩阵 W_Q, W_K, W_V , 对于文献 [12] 中广泛采用的自注意模块, 注意力权重计算如下:

$$A(W_Q, W_K) = \text{softmax}\left(\frac{W_Q^T \cdot W_K}{\sqrt{d}}\right) \quad (5)$$

其中, d 表示 W_Q 和 W_K 的特征维度。然后, 自注意模块输出的注意力图 F_{S-ATT} 被计算为:

$$F_{S-ATT} = A(W_Q, W_K) \cdot W_V \quad (6)$$

最后, 将通道注意力和自注意力两条路径的输出融合得到最终输出的注意力图 F :

$$F = F_{C-ATT} + F_{S-ATT} \quad (7)$$

为了提高模型对相似外观车辆重识别的准确性, 对于局部分支的局部特征, 拟使用三元组损失优化嵌入特征空间中特征之间的距离, 试图减少具有相同身份的样本之间的距离, 同时增加来自不同车辆样本之间的距离, 并且边界参数 m 设置为 0.3。局部分支的损失函数公式如下:

$$L_i^t = \text{Max}(D_{a,p} - D_{a,n} + m, 0) \quad (8)$$

其中, $D_{a,p}$ 表示具有相同身份的样本之间的欧氏距离, $D_{a,n}$ 表示不同车辆样本之间的欧氏距离。

1.3 全局特征增强

为了能够弥补车辆局部特征的不足, 本文还通过车辆的全局特征来提高车辆重识别网络的性能。本文采用在 ImageNet^[13] 上进行预训练的 ResNet50^[14] 作为特征提取网络。输入的车辆图像 x_i , 将 ResNet50 中保留较多空间信息和较大空间尺寸的 Layer3 作为全局特征图 $F_{L3} \in R^{16 \times 16 \times 2048}$, 如图 2 所示。然而, 考虑到真实交通环境中车辆所面临的问题, 提取到的全局特征 F_{L3} 存在语义性较差的问题, 从而导致网络拟合效果不佳。因此, 在本文中, 采用基于图像空间和通道关系的全局特征增强模块对全局特征 F_{L3} 进行增强。该模块包含两个部分: 空间关系分支和通道关系分支。首先, 将提取到的全局特征 F_{L3} 输入到包含空间关系分支和通道关系分支的全局特征增强模块中, 通过深层池化和卷积来获取全局特征上的空间和通道关系, 以此获得具备更强语义性的全局特征。然后,

将输入的全局特征进行连接来提高对细节的感知能力。最后通过融合两个分支的输出, 得到增强后的特征。

空间关系分支: 该分支利用特征间的空间关系, 计算全局特征的空间关系特征图, 然后利用空间关系特征图对输入的全局特征图 F_{L3} 进行重新标定。为了获取空间关系, 首先沿着通道维度分别对全局特征图 F_{L3} 进行全局平均池化和全局最大池化操作以获得多粒度的特征, 并将输出映射连接起来形成空间特征描述符 $s \in R^{W \times H \times C}$ 。然后, 对 s 进行 1×1 的卷积和 sigmoid 激活函数后, 再与输入的 F_{L3} 做信道乘来获得全局特征的空间关系图,

$$F_s = S(\text{concat}(GAP(F_{L3}), GMP(F_{L3}))) \otimes F_{L3} \quad (9)$$

其中, GAP 和 GMP 分别表示对 F_{L3} 进行平均池化和最大池化, S 则表示经过卷积操作和 sigmoid 激活函数后的输出, \otimes 表示信道乘。

通道关系分支: 该分支旨在利用卷积特征的通道间依赖性, 计算全局特征的空间关系特征图。它首先执行压缩操作以对空间全局上下文进行编码, 然后进行激励操作以完全捕获通道关系。通过对全局特征图 F_{L3} 上应用全局平均池化来实现挤压操作, 使得每个二维的特征图变成一个实数来获得全局感受野, 从而产生中间参数 $c \in R^{1 \times 1 \times C}$ 。激励算子使用卷积层和 sigmoid 激活函数重新校准中间参数 c 。最后, 通道关系分支的输出是通过用激活 s 重新缩放的特征图 F_{L3} 来获得的。

$$F_c = S(GAP(F_{L3})) \otimes F_{L3} \quad (10)$$

将两个分支的输出进行特征融合操作后得到增强后的特征图 F_{DA} :

$$F_{DA} = \text{concat}(F_s, F_c) \quad (11)$$

全局特征增强模块的输出需要送入视图拼接模块中, 将 F_{DA} 与拼接后的视图 P_k 进行点乘去获取视图特征。算法 1 总结了视图拼接、特征增强和注意力学习的详细过程。

算法 1 本文的算法流程

```

Input  $x_i$  to U-net and Resnet50
Output  $M_i$  and  $F_{L3}$ 
For  $F_{L3}$  do
 $F_s = S(\text{concat}(GAP(F_{L3}), GMP(F_{L3}))) \otimes F_{L3}$  (空间关系)
 $F_c = S(GAP(F_{L3})) \otimes F_{L3}$  (通道关系)
 $F_{DA} = \text{concat}(F_s, F_c)$  (输出增强后的特征)
end for
For  $M_i$  do
 $P_k = M_i + M_5(i, k = 0, 1, 2)$  (视图拼接)
For  $P_k$  do
 $F_i = P_k \cdot F_{DA}(i = 0, 1, 2)$  (视图特征)
For  $F_i$  do
 $F_{C-ATT} = \omega_i F_i$  (通道注意力)
 $F_{S-ATT} = A(W_Q, W_K) \cdot W_V$  (自注意力)
 $F = F_{C-ATT} + F_{S-ATT}$  (混合注意力输出)
end for
end for
end for

```

另外,增强后的全局特征 F_{DA} 还要被送入平均池化层,由 1×1 组成的卷积缩减块卷积层、批量归一化(BN)层和激活函数(ReLU)以降低特征维数。接下来,将缩减后的 $1 \times 1 \times 256$ 的全局特征向量输入到 softmax 层,用于 ID 损失约束下的车辆身份预测。为了提高识别精度,将全局分支中的 ID 损失函数设置为 softmax 交叉熵损失,以惩罚错误预测。ID 损失函数的公式如下:

$$L_m = \sum_{i=1}^N y_i \log(\hat{y}_i) \quad (12)$$

其中, N 是数据集中车辆标识的数量, y_i 是样本 i 的基本真实性标签, \hat{y}_i 预测概率。在式(12)中,如果样本的预测身份等于真实身份,则 $\hat{y}_i = 1$, 否则 $\hat{y}_i = 0$ 。

除此之外,所有这些全局特征还通过等式定义的基于三元组损失的函数进行优化。优化过程试图减少具有相同身份的车辆样本之间的距离,同时增加来自不同车辆的样本之间的距离,

$$L_i^c = \text{Max}(D_{a,p} - D_{a,n} + m, 0) \quad (13)$$

最后,网络的总损失函数是所有损失函数之和:

$$L = L + L_i^c + L_i^c \quad (14)$$

2 实验

首先,为了评估网络的性能,本研究使用两个指标作为评估指标:平均精度(mAP)和累积匹配特性(CMC)^[4]。然后,在 VeRi-776^[15] 和 VehicleID^[16] 数据集上与最先进的车辆重识别方法进行比较和消融研究。

2.1 数据集和实验设置

在本文的实验中,首先在两个较为广泛使用的数据集 VeRi-776 和 VehicleID 上比较本文的方法与最先进的车辆识别方法。表 1 汇总了这些数据集的详细信息。

表 1 VeRi-776 和 VehicleID 的属性汇总

Table 1 Summary of attributes for VeRi-776 and VehicleID

数据集	#ID	#图片数	#视角
VeRi-776	776	49 357	8
VehicleID	26 328	221 567	2

在实验时,所有图像都调整为 256×256 。在网络训练时,使用 SGD 优化器且动量设置为 0.9。至于权重衰减策略,将基础学习率设置为 3×10^{-3} ,并在第 120 个时期和第 180 个时期分别下降到 3×10^{-4} 和 3×10^{-5} ,以获得更快的收敛。网络训练的批量设置为 64, epoch 总数为 220 次。在训练和测试阶段,都使用欧几里德距离来评估 query 图像和 gallery 图像之间的特征相似性。

2.2 评估指标

平均精度(mAP)、CMC@1 和 CMC@5 作为评估指

标。给定测试图像子集中用于检索其他测试图像的每个查询图像,每个查询 q 的平均精度通过以下公式计算:

$$AP(q) = \frac{\sum_{k=1}^n P(k) \times rel(k)}{N_{gt}} \quad (15)$$

其中, $P(k)$ 表示截止点 k 处的精度,如果 k 处的项目是匹配的车辆图像,则表示函数 $rel(k)$ 等于 1, 否则为 0, n 是用于检索的数量, N_{gt} 表示用于查询的 ground truth 的编号。然后通过以下公式计算所有查询图像的平均精度:

$$mAP = \frac{\sum_{q=1}^Q AP(q)}{Q} \quad (16)$$

其中, Q 是所有查询的数量。 $AP(q)$ 表示每个查询图像 q 的平均精度检索结果。

度量 CMC 表示在前 k 个检索结果中找到正确结果的概率,可表示为如下等式:

$$CMC@K = \frac{\sum_{q=1}^Q rel(q)}{Q} \quad (17)$$

其中,当图像 q 的正确匹配目标在前 k 个检索结果中排名时, $rel(k) = 1$ 。在本文中, k 取 1 和 5。

2.3 与最新方法的比较

1) 在 VeRi-776 上的实验

在 VeRi-776 数据集上对车辆 Reid 方法进行了评估,其中采用了 mAP、CMC@1 和 CMC@5 这 3 个指标。表 2 呈现了本文方法与最先进方法的比较结果。在这些方法中,PVEN 使用解析掩码信息来对齐车辆,将每辆车分割成 4 个不同的部分分别注释。

表 2 mAP、CMC@1 和 CMC@5 在 VeRi-776 上的结果

Table 2 Results for mAP, CMC@1 and CMC@5 on VERI-776

方法	mAP	C@1	C@5
PVEN ^[3]	79.50	95.62	98.43
TBE-Net ^[5]	79.50	96.00	98.50
SAVER ^[11]	79.60	96.40	98.61
TransReid ^[12]	82.00	97.10	-
PRN ^[17]	74.30	94.34	98.91
VPEN ^[18]	81.20	97.30	99.10
DFNet ^[19]	80.97	97.08	99.01
VABPP ^[20]	80.80	96.60	-
DRA ^[21]	78.47	76.03	87.33
SSBVER ^[22]	80.94	97.02	98.45
Ours	82.41	98.63	99.23

而本文方法在 PVEN 的基础上将高频出现的侧面视图特征与其他 3 面拼接,有效地缓解了视图分割不精确的问题,mAP 也提升了 2.91%。而 DFNet 认为车辆的方位是影响车辆特征的关键因素,通过相似方位捕获的车

辆学习详细的差异信息,以此区分相似但不相同的车辆对,但是该方法需要大量的方位标注信息。此外,对于局部特征选取,TBE-Net 采用目标检测算法^[23]来获取具有辨识性的局部区域,如灯光、标志、格栅等,但是不同的局部区域会因为视图的改变导致获取的局部区域不准确,从而影响重识别精度。相比之下,本文通过混合注意力的方法从车辆的拼接视图考虑局部区域的重要性。利用通道注意力机制专注拼接视图中的局部信息,利用自注意力机制捕捉拼接视图的局部信息相关性。再将两个注意力的输出进行融合,更好地凸显了车辆的显著性局部特征,并通过这些局部特征为车辆关键局部区域赋予更大的权重来进行更多的关注。

2) 在 VehicleID 上的实验

本文在 VehicleID 数据集的 3 个测试子集(小、中、大)上与最新的方法比较了 CMC@1 和 CMC@5 分数,表 3 显示了在 3 个不同大小(小、中和大)的测试数据集上的比较结果。观察表中实验结果,本文方法领先于绝大多数最新的方法。首先,在这些最先进的方中,Transreid 设计一个拼图面片模型,利用平移、面片洗牌操作和自注意力来获取扰动不变和稳定的特征表示,但是基于 Transformer 的模型对于训练样本需求量较大,而通过利用 Transformer 中的自注意力与通道注意力结合的方法,不仅降低了模型的计算量,还提高模型的重识别精度。除此之外,TBE-Net 设计一种三支嵌入网络,通过全局和局部分支学习车辆的全局外观特征和具有细微差别的局部特征,通过补充性分支通过合并操作学习了更丰富的车辆特征。VAMI 则应用了车辆的视点信息,但它很难解决同一摄像机视角下不同车辆在外观相似时难以识别的问题。然而,在真实的交通场景下,拍摄到的车辆样本少且容易受视角变化、图像失真和光照条件等外在因素的影响,这些方法很难获取预先定义的局部区域和视点信息,从而导致重识别的准确性较低。相比之

下,本文利用全局特征增强模块使得网络很好地适应了复杂交通环境下存在的问题,并且利用视图拼接的方法有效解决了同一车辆在不同视角下存在较大差异的问题,从而帮助我们获取到语义增强和包含细节的车辆全局特征。

表 3 CMC@1 和 CMC@5 在 VehicleID 数据集的结果

Table 3 Results of CMC@1 and CMC@5 on the VehicleID dataset

设置	小		中		大	
方法	C@1	C@5	C@1	C@5	C@1	C@5
PVEN ^[3]	0.847	0.970	0.806	0.945	0.778	0.920
TBE-Net ^[5]	86.0	98.4	82.3	96.6	80.7	94.9
TransReid ^[12]	85.20	97.50	-	-	-	-
PRN ^[17]	78.92	94.81	74.94	92.02	71.58	88.46
DFNet ^[19]	84.76	96.22	80.61	94.10	79.15	92.86
VABPP ^[20]	82.80	97.90	79.50	95.40	76.20	92.40
DRA ^[21]	-	-	-	-	79.25	91.03
SSBVER ^[22]	85.61	97.73	80.34	94.92	77.26	92.59
EALN ^[24]	75.11	88.09	71.78	83.94	69.30	81.42
VAMI ^[25]	63.12	83.25	52.87	75.12	47.34	70.29
Ours	86.90	98.74	83.84	96.73	81.63	94.47

2.4 消融实验

本文通过消融实验,分别探索了视角拼接、混合注意力模块和全局特征增强模块的性能。在 VeRi-776 数据集上进行了消融实验,并采用了 mAP、CMC@1 和 CMC@5 这 3 个度量指标,以进一步验证模型的有效性。

除此之外,本文还进行了可视化处理。首先,选取 4 种不同的车型,通过模型生成它们的注意力图,如图 2 所示,可识别的局部区域被突出显示,如车灯、品牌标识、装饰和司机,说明本文的模块可以增强重要的局部区域。然后,随机挑选 2 个车辆图片,图 3 为本文方法与仅基于全局特征的方法(Baseline)按照相似度评分后排名前 10 的检索结果,其中,√表示正确匹配;×表示错误匹配。由图 3 可知,本文方法明显优于 Baseline 方法。

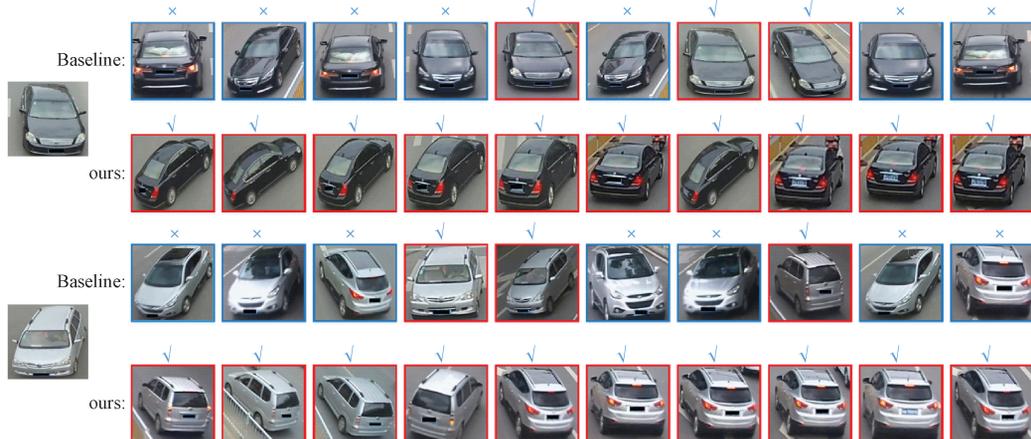


图 3 排名前 10 的检索结果

Fig. 3 Top 10 search results

1) 视图拼接的作用

为了验证视图拼接可以有效地解决分割不准确的问题,对其进行了消融实验,并将结果总结在表 4 中。首先,通过 U-Net 直接获取车辆的前(F)、后(B)、顶(T)和侧(S)4 个视图信息,在未拼接的情况下直接输入混合注意力模块中,根据表中的结果,不采用拼接时的 mAP 值远低于本文采用拼接后的结果。该消融实验表明视图拼接可以获取更丰富的车辆特征。除此之外,为了验证采用侧面拼接的优越性,通过将顶部视图与其他 3 个面拼接来进行比较。消融实验结果表明,采用侧面拼接的方法 mAP 比采用顶部拼接的方法高出了 2.72%。因此表明,采用侧面拼接的方法更有利于获得丰富的特征和提高模型的性能。

表 4 在 Veri-776 上验证视图拼接的结果

Table 4 Results of view stitching on VeRi-776

设置	mAP	CMC@ 1	CMC@ 5
F, B, T, S	78.81	96.86	98.65
$F+T, B+T, S+T$	79.69	97.85	98.83
$F+S, B+S, T+S$	82.41	98.63	99.23

2) 混合注意力模块的作用

为了验证混合注意力模块在局部特征学习的有效性,对其进行了消融实验,并将结果总结在表 5 中。首先在网络模块中删除混合注意力模块提取的特征,即 $F_{C-ATT} + F_{S-ATT}$,通过比较拼接视图特征 F_i 后有无混合注意力模块的影响,实验结果表明,移除混合注意力模块会对车辆重识别的精确性产生负面影响,mAP 最大减少了 5.55%。这表明混合注意力模块可以有效提高车辆识别的准确性。为了进一步验证注意力模块的有效性,还单独验证了通道注意力模块和自注意力模块的作用,即 F_{C-ATT} 和 F_{S-ATT} ,这两个模块单独作用时,在 mAP 上的表现比不加注意力时分别提升了 1.49% 和 2.83%。消融实验结果表明,混合注意力模块对最终的影响较为明显。

表 5 在 Veri-776 上验证混合注意力模块的结果

Table 5 Results of hybrid attention module on VeRi-776

设置	mAP	CMC@ 1	CMC@ 5
Baseline	75.12	90.27	93.45
F_i	76.86	91.81	94.36
F_{C-ATT}	78.35	96.97	98.44
F_{S-ATT}	79.69	97.52	98.87
$F_{C-ATT} + F_{S-ATT}$	82.41	98.63	99.23

3) 全局特征增强模块的作用

为了验证全局分支中全局特征增强模块的有效性,本文对其进行了消融实验。实验将特征增强模块移除,

直接将特征提取网络提取到的特征作为模型的全局特征,实验结果如表 6 所示, F_{L3} 表示增强前的全局特征, F_{DA} 表示经过特征优化后的全局特征。可以明显的发现,加入特征优化后的模型,整体性能提升了 3.29%。除此之外,本文还验证了全局特征增强模块中空间分支和通道分支的影响。单独使用其中的一个,mAP 分别提高了 1.72% 和 2.24%。结果表明,利用全局特征增强模块使得模型获取的特征更优越,与未使用该方法相比,所提出的网络具有更完整和更丰富的粒度。

表 6 在 Veri-776 上验证特征增强模块的结果

Table 6 Results of feature enhancement module on VeRi-776

设置	mAP	CMC@ 1	CMC@ 5
F_{L3}	79.12	90.27	93.45
空间分支	80.84	96.54	98.12
通道分支	81.36	97.63	98.97
F_{DA}	82.41	98.63	99.23

3 结论

针对真实交通场景中存在的各种干扰因素,提出了一种具有多视图融合的混合注意力机制和全局特征增强的双分支网络。从提高特征表达能力和特征质量的角度出发,缓解了车辆图像受视角变化、光照条件变化等真实交通环境的影响。在特征学习中,局部分支通过 U-net 网络对车辆图像进行分割获得车辆 4 个面的视图信息,并且利用视图拼接的方法将车辆出现率较高的侧面视图与其他 3 个视图进行拼接,以此解决分割不准确的问题。再通过混合注意力模块对车辆具有区分线索的关键局部区域赋予较大的权重,以增强提取的局部区域特征。全局分支通过特征增强模块挖掘全局特征的深层语义信息,并保持原有的空间大小和空间信息。特征增强模块保证模型不仅能提取到语义增强的车辆特征,而且还使车辆特征中包含完好的细节信息。实验结果也验证了所提算法的有效性。

对于未来的研究,可以引入少样本学习的方法对车辆重识别模型进行改进,减少模型对大量样本的依赖,使其针对一些车辆图像较少的特殊场景(肇事逃逸、违法犯罪等)依旧保持良好的性能。此外,对所提出的网络进行轻量化研究,以便训练后的网络可以应用于移动设备,如无人机^[21],实现高效的自主、实时车辆重识别。

参考文献

- [1] WANG Z, TANG L M, LIU X H, et al. Orientation invariant feature embedding and spatial temporal regularization for vehicle re-identification [C].

- Proceedings of IEEE International Conference on Computer Vision. Venice, Italy: IEEE Press, 2017: 379-387.
- [2] ZHANG X, ZHANG R, CAO J, et al. Part-guided attention learning for vehicle instance retrieval [C]. IEEE Transactions on Intelligent Transportation Systems, 2020:1-12.
- [3] MENG D, LI L, LIU X J, et al. Parsing-based view-aware embedding network for vehicle re-identification [C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA: IEEE Press, 2020:7101-7110.
- [4] PENG J, ZHANG Z, HE H. A method for detecting abnormality of CAN bus in vehicle[J]. Instrumentation, 2017, 4(2):28-33.
- [5] SUN W, DAI G, ZHANG X, et al. TBE-Net: A three-branch embedding network with part-aware ability and feature complementary learning for vehicle re-identification [C]. IEEE Transactions on Intelligent Transportation Systems, 2021:1-13.
- [6] 郭秋蕊, 李建良, 田垚, 等. 基于改进 KCF 算法和多特征融合的车辆跟踪研究[J]. 电子测量与仪器学报, 2022, 36(4): 231-240.
- GUO Q R, LI J L, TIAN Y, et al. Research on vehicle tracking based on improved KCF algorithm and multi-feature fusion [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36(4): 231-240.
- [7] ZHOU Y, SHAO L. Viewpoint-aware attentive multi-view inference for vehicle re-identification [C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE Press, 2018: 6489-6498.
- [8] LIU X, LIU W, MEI T, et al. PROVID: Progressive and multimodal vehicle re-identification for large-scale urban surveillance [J]. IEEE Transactions on Multimedia, 2018, 20(3):645-658.
- [9] SHEN Y, XIAO T, LI H, et al. Learning deep neural networks for vehicle re-id with visual-spatio-temporal path proposals [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 1900-1909.
- [10] CHU R, SUN Y, LI Y, et al. Vehicle reidentification with viewpoint-aware metric learning[C]. Proceedings of IEEE International Conference on Computer Vision. Salt Lake City, US: IEEE Press, 2019:8281-8290.
- [11] KHORRAMSHAHI P, KUMAR A, PERI N, et al. A dual-path model with adaptive attention for vehicle re-identification [C]. Proceedings of IEEE International Conference on Computer Vision. Seoul, Korea (South): IEEE Press, 2019:6131-6140.
- [12] HE S, LUO H, WANG P, et al. TransReID: Transformer-based object re-identification [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021:15013-15022.
- [13] RUSSAKOVSKY O, DENG J, SU H, et al. Imagenet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3):211-252.
- [14] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016:770-778.
- [15] GUO H, ZHAO C, LIU Z, et al. Learning coarse-to-fine structured feature embedding for vehicle re-identification[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2018.
- [16] LIU H, TIAN Y, YANG Y, et al. Deep relative distance learning: Tell the difference between similar vehicles [C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016:2167-2175.
- [17] KUMA R, WEILL E, AGHDASI F, et al. Vehicle re-identification: An efficient baseline using triplet embedding[C]. International Joint Conference on Neural Networks (IJCNN). IEEE, 2019:1-9.
- [18] 代广昭, 孙伟, 徐凡, 等. 用于车辆重识别的视角感知局部注意力网络[J]. 计算机工程, 2022, 48(10): 288-297, 305.
- DAI G ZH, SUN W, XU F, et al. Visual angle perception local attention network for vehicle recognition [J]. Computer Engineering, 2022, 48(10):288-297,305.
- [19] BAI Y, LIU J, LOU Y, et al. Disentangled feature learning network and a comprehensive benchmark for vehicle re-identification [C]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021:1-17.
- [20] LI H, WANG Y, WEI Y, et al. Discriminative-region attention and orthogonal-view generation model for vehicle re-identification[C]. Applied Intelligence, 2022:1-18.
- [21] KHORRAMSHAHI P, SHENOY V, CHELLAPPA R. Scalable vehicle re-identification via self-supervision[J]. arXiv preprint arXiv: 2205.07613, 2022.
- [22] YAO A, QI J, ZHONG P. Self-aligned spatial feature extraction network for uav vehicle re-identification [J]. arXiv preprint arXiv: 2201.02836, 2022.
- [23] 闫钧华, 张琨, 施天俊, 等. 融合多层级特征的遥感图像地面弱小目标检测[J]. 仪器仪表学报, 2022, 43(3):221-229.
- YAN J H, ZHANG K, SHI T J, et al. Fusion hierarchy

characteristic of remote sensing image more ground weak target detection [J]. Chinese Journal of Scientific Instrument, 2022, 43(3): 221-229.

- [24] LIU X, LIU W, MEI T, et al. Provid: Progressive and multimodal vehicle reidentification for large-scale urban surveillance [J]. IEEE Transactions on Multimedia, 2017, 20(3): 645-658.
- [25] HU Z, XU Y, WEN J, et al. Global-supervised contrastive loss and view-aware-based post-processing for vehicle re-identification [J]. arXiv preprint arXiv: 2204.07943, 2022.

作者简介



孙伟(通信作者),于 2004 年和 2006 年获得河南科技大学机械制造及其自动化专业学士学位和硕士学位。2010 年获得东南大学仪器科学与技术专业博士学位。2014 年~2015 年,普渡大学 NEXTRANS 中心博士后研究员作为大学 NEXTRANS 中心博士后研究员。现为南京信息工程大学自动化系教授,主要研究方向为车辆再识别、计算机视觉、深度学习和智能车辆环境感知。

E-mail: sunw0125@163.com

Sun Wei (Corresponding author) received the B. Sc. and M. Sc. degrees in mechanical manufacture and automation from Henan University of Science and Technology, China, in 2004 and 2006, respectively, and the Ph. D. degree in instrument science and technology from Southeast University, China, in 2010. From 2014 to 2015, he was a post-doctoral researcher with the NEXTRANS Center, Purdue University, USA. He is currently a professor of Automation with Nanjing University of Information Science and Technology. His main research interests include vehicle re-identification, computer vision, deep learning and environment perception for intelligent vehicles.



徐凡,2020 年获南京信息工程大学电气工程及其自动化专业学士学位。现在在南京信息工程大学攻读控制工程硕士学位,主要研究方向为大规模车辆检索和车辆再识别。

E-mail: xufan110896@163.com

Xu Fan received the B. S. degree in electrical engineering and automation from Nanjing University of Information Science and Technology, China, in 2020. He is currently a M. Sc. candidate in the control engineering with Nanjing University of Information Science and Technology. His main research interests include large-scale vehicle retrieval and vehicle re-identification.



张小瑞,2004 年和 2006 年分别获得河南科技大学机械制造及其自动化专业学士学位和硕士学位。2010 年获得东南大学仪器科学与技术专业博士学位。2013 年~2014 年,在国宾夕法尼亚州费城宾夕法尼亚大学 ViDi 中心的博士后研究员。现为南京信息工程大学计算机科学与技术学院教授,主要研究方向为虚拟现实和人机交互、触觉感知和模式识别。

E-mail: zxr365@126.com

Zhang Xiaorui received the B. Sc. and M. Sc. degrees in mechanical manufacture and automation from Henan University of Science and Technology, China, in 2004 and 2006, respectively, and the Ph. D. degree in instrument science and technology from Southeast University, China, in 2010. From 2013 to 2014, she was a post-doc researcher with the ViDi Center, University of Pennsylvania, Philadelphia, PA, USA. She is currently a professor of Computer Science and Technology with Nanjing University of Information Science and Technology. Her main research interests include virtual reality and human - computer interaction, haptic perception, and pattern recognition.



胡亚华,2020 年获得南京信息工程大学电气工程及其自动化专业学士学位。现在在南京信息工程大学攻读控制工程硕士学位,主要研究方向为大规模车辆检索和车辆再识别。

E-mail: m15852866478@163.com

Hu Yahua received the B. Sc. degree in electrical engineering and automation from Nanjing University of Information Science and Technology, China, in 2020. He is currently a M. Sc. candidate in the control engineering with Nanjing University of Information Science and Technology. His main research interests include large-scale vehicle retrieval and vehicle re-identification.



赵宇煌,2021 年获得南京信息工程大学电气工程及其自动化专业学士学位。现在在南京信息工程大学攻读控制工程硕士学位,主要研究方向为车辆再识别。

E-mail: 1402327414@qq.com

Zhao Yuhuang received the B. Sc. degree in electrical engineering and automation from Nanjing University of Information Science and Technology, China, in 2021. He is currently a M. Sc. candidate in the control engineering with Nanjing University of Information Science and Technology. His main research interests include large-scale vehicle retrieval and vehicle re-identification.