

DOI: 10.13382/j.jemi.B2306555

基于弱监督注意力和知识共享的车辆重识别*

孙伟^{1,2} 赵宇煌¹ 张小瑞^{2,3,4,5} 刘轩诚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. 南京工业大学计算机与信息工程学院 南京 211816; 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:为解决弱监督车辆重识别方法中标签不准确和背景干扰下预定义局部区域难以获得的问题,提出一种基于弱监督注意力和知识共享的车辆重识别网络。在弱监督注意力模块(WAM)中使用弱监督的方法生成车辆部件掩模,通过部件通道对齐步骤使得该模块在复杂背景下也能自适应地进行特征对齐。针对弱监督方法中标签准确性不高导致WAM模块生成部件掩模不稳定的问题,在局部分支中构建了知识共享模块。该模块利用迁移学习从WAM模块中提取车辆部件特征,并进行多尺度部件特征提取,防止了不稳定的车辆部件掩模生成。通过实验,mAP、CMC@1和CMC@5分别达到了82.12%、98.50%和99.12%,优于现有的方法,验证该方法的有效性。

关键词: 车辆重识别; 弱监督; 迁移学习; 注意力机制

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

Weakly supervised attention and knowledge sharing for vehicle re-identification

Sun Wei^{1,2} Zhao Yuhuang¹ Zhang Xiaorui^{2,3,4,5} Liu Xuancheng¹

(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. College of Computer and Information Engineering, Nanjing Tech University, Nanjing 211816, 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: In order to solve the problem that the label is not accurate and the background interference makes it difficult to obtain the predefined local area in the weak supervision vehicle re-identification method. A vehicle re-identification network based on weak supervised attention and knowledge sharing is proposed. In the weak-supervised attention module (WAM), the weak-supervised method is used to generate the vehicle component mask, and the component channel alignment step enables the module to perform adaptive feature alignment under complex background. Aiming at the problem that the mask of WAM module is unstable due to the low accuracy of labels in weak supervision method, a knowledge sharing module is constructed in local branches. The module uses migration learning to extract vehicle component features from WAM module, and performs multi-scale component feature extraction to prevent unstable vehicle component mask generation. Through experiments, mAP, CMC@1 and CMC@5 reached 82.12%, 98.50% and 99.12%, respectively, which are better than the existing methods and verify the effectiveness of this method.

Keywords: vehicle re-identification; weak supervision; transfer learning; attention mechanism

收稿日期: 2023-05-23 Received Date: 2023-05-23

* 基金项目: 国家自然科学基金(62376128, 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(SJCX23_0380)、大学生创新创业训练项目(XJDC202110300601, 202010300290, 202010300211, 202010300116E)资助

0 引言

车辆重识别旨在从给定查询集中查询车辆,从不同摄像头拍摄的图像库中找到与查询车辆具有相同身份的所有车辆。在进行车辆重识别任务时,由于现实交通场景背景复杂,拍摄的车辆图像中可能包含有其他的车辆或行人,很难从中挑出目标车辆。因此对于没有图像语义理解的计算机而言,车辆重识别是一个非常具有挑战性的研究课题。

早期的车辆重识别方法^[1-2]主要关注于手工设计的底层特征,诸如颜色和纹理等,并设计一定的计算规则进行车辆重识别任务的研究。此类方法侧重于对车辆图像某方面特征的抽取,对特定任务表现出不错的识别性能。然而这些底层特征容易受到外界环境的干扰,例如光照、遮挡和视角变化等,使这类算法的识别精度受限。随后,一些算法^[3-4]根据车牌是车辆身份唯一标识这一特性,通过检测图像中车牌的位置,将车牌字符从整幅图像中分割出来,并进一步根据车牌号进行车辆重识别。此类方法在一定程度上缓解了早期方法对手工设计底层特征的依赖。然而,由于车牌像素模糊以及车牌被伪造、遮挡和涉及隐私安全等诸多方面因素,给搜索目标车辆带来极大挑战。

受益于卷积神经网络的快速发展,为了进一步提高车辆重识别的可靠性,一些研究聚焦于车辆显著的局部区域,例如车内的挂件装饰、年检标志等,进而提取具有区分性的信息。然而此类方法^[5]在使用目标检测方法进行关键点提取时依赖手工标注,需要大量像素级注释,既消耗大量时间且容易出错,使得重识别任务成本昂贵。并且由于摄像头不同导致视角不同,标注的关键点难以对齐,因此依靠外部工具预测的部件掩模不够可靠。为了减少人工成本,更好的聚焦这些显著性局部区域,基于弱监督定位的方法^[6-7]被采用,该方法通过使用图像框等弱监督标注的方式,使网络能够学习提取车辆关键部位的特征,在避免人工标注的情况下也能很好的实现部件定位。如 Yao 等^[8]通过弱监督学习以用于部件定位,然而真实场景中拍摄图片中往往不止包含目标车辆,出现行人或其他车辆等复杂背景的干扰,导致弱监督注意力机制生成的掩模不准确。张正等^[9]考虑解决背景复杂的问题,使用语义分割方法较大程度上去除背景信息的干扰,得到了有利于车辆重识别的局部特征信息。相较于早期的重识别研究,减少背景干扰的方法一定程度上缓解了特征对齐的问题,但是语义分割的方法在面对背景多目标的情况下无法很好地提取特定目标,同时,这种方法也需要大量的强监督标注信息,不能完全适用于减少复杂背景干扰的弱监督方法。因此,本文对弱监督注意

力机制进行了改进,并加入了通道对齐和部件掩模生成步骤,以改善弱监督注意力所生成的部件掩模并抑制复杂背景的干扰。使用弱监督学习是为了节约收集监督信息的成本,因此在没有监督信息的情况下让计算机标注不精确的标签,这却使得特征信息不准确,影响重识别的准确性^[10]。

为了提高特征标签的准确性,一些学者将迁移学习应用在重识别研究中^[11-13]。Wang 等^[12]对特征统一划分并进行水平细化,但处理车辆部件特征对齐的能力有限。Ding 等^[11]以迁移学习的方式处理部分错位,然而该研究在特征检测中依赖大量的标注信息,增加了人工成本。受其启发,本文的知识共享模块运用迁移学习将部件特征概念提炼转移,使用弱监督掩模减少人工成本并避免了提取特征信息不准确的影响。

综上所述,使用弱监督注意力机制代替传统的手工标注特征,可以减少大量人工成本。然而弱监督注意力机制容易受到图像中复杂背景的影响,且是在没有监督信息的情况下让计算机标注不精确的标签,这使得特征信息不准确,影响重识别的准确性。为此本文提出一种基于弱监督注意力和知识共享的车辆重识别网络。本文设计部件弱监督注意力模块(weakly supervised attention module, WAM)以聚焦局部特征,使用包括像素级标注、包围框等作为弱监督的标注。由于弱监督在背景复杂的图片中很难清楚地分割目标车辆,为避免背景复杂对弱监督方法的影响,本文加入了部件通道对齐以改进弱监督生成的部件掩模。同时考虑到弱监督方法中不精准的标签使提取的特征信息不准确,导致生成部件掩模不稳定从而影响重识别的精度。因此,本文在局部分支中引入迁移学习,设计知识共享模块。该模块包括指导分支、学习分支和多尺度注意力模块(multiscale attention block, MAB)。其中指导分支经过 WAM 模块学到弱监督部件特征,通过在 MAB 模块中实行权重共享,并构建迁移损失对指导分支和学习分支进行一致性约束,使指导分支将经过 WAM 模块学到的部件特征传授给学习分支。使得学习分支不用经过 WAM 模块也能学到车辆部件特征,从而避免了弱监督方法对部件掩模生成稳定性的影响。在 MAB 模块中还将通道注意力和多尺度注意力的结合,可以获得车辆丰富的多尺度特征并增强局部特征表示。

1 方法

本文提出的基于弱监督注意力和知识共享的车辆重识别网络框架如图 1 所示,它包含全局分支和局部分支。在全局分支中,车辆图片通过特征提取网络并经过一个全局平均池化层(global average pooling, GAP)得到全局

特征。在局部分支中, WAM 模块通过部件通道对齐和部件掩模生成来预测车辆部件掩模。车辆图片首先通过视图分割^[14]得到车辆的前、后、侧和顶的 4 个视图, 并通过与全局分支提取的全局特征相乘得到 4 个视图特征, 随后送入知识共享模块。该模块包括指导分支、学习分支和 MAB。经过视图分割的视图特征直接作为学习分支

的输入特征, 而视图特征通过 WAM 模块的部件掩模后作为指导分支的输入特征。通过在 MAB 模块中实行权重共享, 并构建迁移损失对指导分支和学习分支进行一致性约束, 将指导分支的部件特征提取方式传授给学习分支, 使没有经过弱监督指导的学习分支也能获得弱监督车辆部件特征。

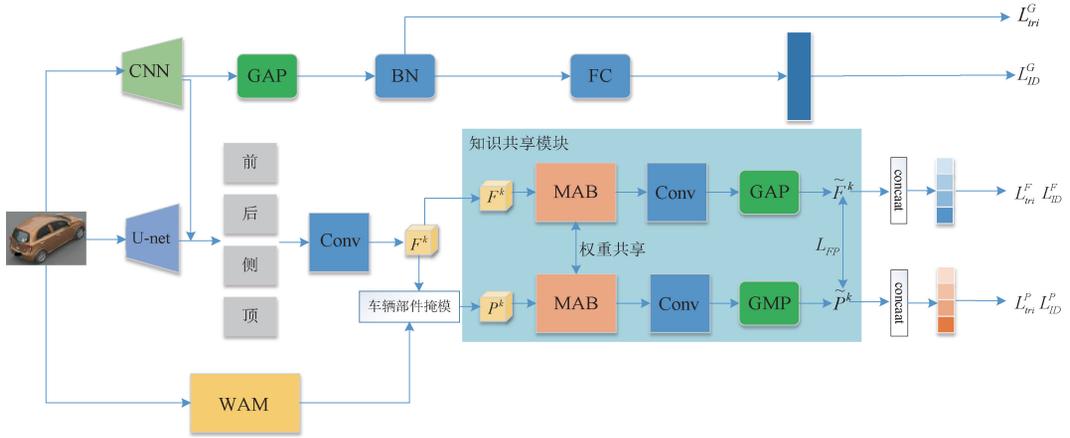


图 1 网络框架

Fig. 1 Network framework

本文使用在 ImageNet^[15] 上进行预训练的 ResNet50^[16] 作为特征提取网络, 考虑在 ResNet50 保留较多空间信息和较大空间尺寸, 将 layer4 作为输出得到大小为 $16 \times 16 \times 2048$ 的全局特征图。之后全局特征将直接被输入进全局平均池化层, 通过归一化层和全连接层后分别进行三元损失^[17] 和 ID 损失的训练。

1.1 弱监督注意力模块 (WAM)

受行人重识别^[8] 的启发, 在卷积神经网络 (convolutional neural networks, CNN) 的顶层上不同通道组描述了人的不同部位。在车辆 Re-ID 以往的研究^[18] 中通道和车辆部件之间也有对应关系, 因此最直观聚焦局部特征的方法是在通道空间中定位车辆部件的突出部分, 然而以往的方法多依赖人工标注, 消耗大量人工成本, 并且容易受复杂背景的干扰。基于此本文提出了 WAM 模块, 其中使用弱监督注意力机制 GrabCut^[19] 算法, 对图片进行检测框标注, 通过迭代步骤逐渐细化目标和背景的分割结果, 使网络能在复杂背景下自适应的提取特征, 节约了人工标注成本。尽管弱监督可以显著减少标注时间, 但其效果容易受到图像中复杂背景的干扰, 从而导致精度不稳定。因此本文提出部件通道对齐和部件掩模生成步骤, 生成一个精细的车辆部件掩模, 以抑制复杂背景对弱监督注意力机制的影响。通过这些步骤, 使 WAM 模块在背景复杂的影响下也能自适应进行显著特征提取。

如图 2 所示, 主干网络为 ResNet50, 将提取到的特征

F 经过 4 个转置卷积层和 sigmoid 函数, 其中前 3 个卷积层后面都含有 BN 层和 ReLU 激活函数。为了在没有手工注释的情况下获得车辆视图的部件掩模, 本文使用 GrabCut^[19] 算法生成每个视图粗略的部件掩模, 将生成的掩模作为伪标签, 使用均方误差进行训练。为了减少背景复杂对弱监督方法造成的影响并学习更鲁棒的零件特征, 本文还设计部件通道对齐的步骤, 其中使用注意力机制执行部件相关的通道校准, 并通过和上述生成的掩模进行部件视图生成, 得到精细的车辆部件掩模作为 WAM 模块的输出。

1) 部件通道对齐

为了突出特征图中车辆部件并抑制背景中不重要的信息, 本文使用注意力机制进行部件通道对齐。如图 2 所示, 对于主干网络提取的特征图 $F \in R^{C \times H \times W}$, 其中通道为 C 高度为 H 宽度为 W , 首先将特征图重塑为矩阵 $\tilde{F} \in R^{C \times HW}$, 同时将特征图 F 输入全局平均池化层, 得到 $F_{sq} \in R^{C \times 1}$, 将 \tilde{F} 和 F_{sq} 相乘并通过 softmax 函数输出, 得到 C 个通道之间的相似性如式 (1) 所示。

$$T(F) = \text{softmax}(\tilde{F} \otimes F_{sq}) \quad (1)$$

式中: \otimes 表示矩阵相乘。

当 $T(F)_{C_1, C_2}$ 的值越大, 说明 C_1 和 C_2 通道越相似, 即通道 C_1 和 C_2 更可能集中在同一车辆部件上。之后将矩阵特征 \tilde{F} 和 $T(F)$ 相乘并通过 softmax 函数输出得到标准化注意力图 \hat{F} , 用于对输入图像中每个位置的重要性

赋予不同的权重,从而使得模型更加集中于输入图像的关键部分,如式(2)所示。

$$\hat{F} = \text{softmax}(T(F) \otimes \tilde{F}) \quad (2)$$

式中: \otimes 表示矩阵相乘。

最后将注意力图重塑成 F 的形状,作为本步骤的输出。

2) 部件掩模生成

首先找出特征图中每个通道的最大连接分量 $\hat{F}_c \in R^{1 \times H \times W} \times c = \{1, 2, 3, C\}$,通过数学形态学中开运算等方法

得到相应的边界框 $\text{Box}(\hat{F}_c)$,之后通过 K 均值聚类将 $\text{Box}(\hat{F}_c)$ 的中心坐标 (x_c, y_c) 分组为 K 个聚类。由于没有监督信息来确定 K 个部分的边界,因此本文没有预定义每个掩模的大小,而是在此过程使用一个随机模型^[20]。在该模型中,从每个聚类中随机选择一个边界框 M_r ,在 WAM 模块每次训练迭代中用该边界框裁剪出细化的车辆部件掩模。随机模型为指导分支提供了车辆部件显著特征,并引导其关注部件特征。

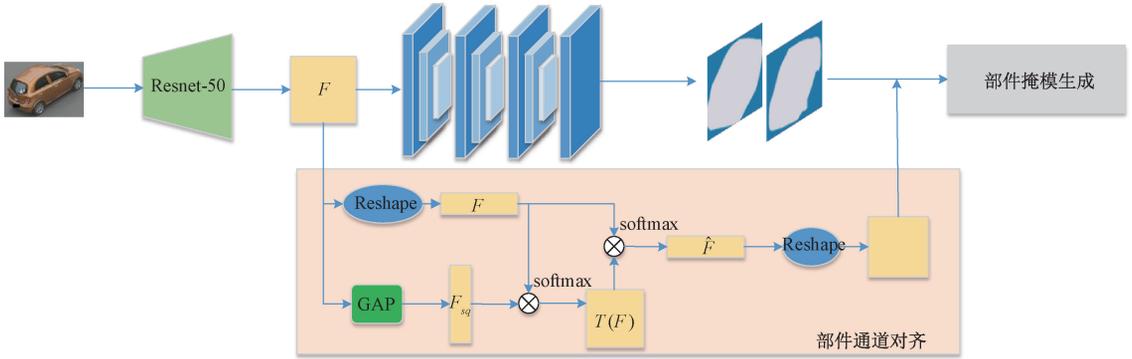


图 2 WAM 模块

Fig. 2 The WAM module

1.2 知识共享模块

本文的目标是在背景复杂的情况下进行车辆重识别任务。由于 WAM 模块中用弱监督方法来学习车辆部件特征,为了减少弱监督方法中不确切的标签可能导致特征信息不准确的情况,本文在局部分支中设计了知识共享模块。该模块包含指导分支、学习分支和 MAB。通过 MAB 模块的权重共享,并提出迁移损失对两个分支的输出进行一致性约束,使学习分支学会指导分支中的弱监督部件特征,从而不需要经过 WAM 模块,避免了弱监督方法中不确切标签的影响。在 MAB 模块中还通过通道注意力和多尺度注意力的结合,可以获得车辆丰富的多尺度特征并增强局部特征表示。

如图 1 所示,知识共享模块的输入为图像分割后的 4 个视图,通过 k 个部件特征学习头,消除了最后的空间下采样操作的两个 Conv5 层以应对空间信息损失。视图特征 F_i^k 直接进入学习分支,而在指导分支中视图特征要先经过 WAM 模块的部件掩模裁剪,并重塑成 F_i^k 的形状作为指导分支的输入特征 P_i^k 。学习分支中输入特征 F_i^k 依次经过多尺度注意力模块,卷积层和 GAP (global average pooling)。GAP 用来学习平移和缩放不变的整体特征。指导分支使用 GMP (global max pooling) 来代替 GAP,其中 GMP 可以实现为最大池化层,池化大小等于部件掩模的面积。相比于 GAP,GMP 有助于在受约束的

部件掩模上挖掘显著的局部特征。指导分支和学习分支在结构上相似但输入不同。学习分支的输入为原始特征图 F_i^k ,指导分支的输入为 P_i^k ,其中 P_i^k 由 WAM 模块生成的部件掩模裁剪。流程如图 3 所示。

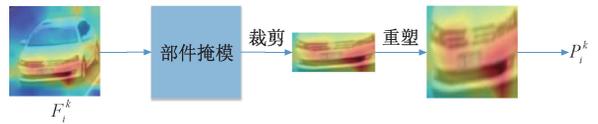


图 3 指导分支输入特征生成

Fig. 3 Input generation for Guidance branch

1) 视图分割

在局部分支中,本文用带注释的数据集训练视角解析网络对待输入进网络的图片进行分割。对于一张输入车辆图片,定义每个视角 i 覆盖的空间区域为 $\{S_i\}_{i=1}^4$ 。在每一个视角下将在视角掩模区域内的元素赋值 1,区域外元素赋值 0,得到一个二元区域矩阵 $M_i \in \{0, 1\}^{H \times W}$,如式(3)所示。

$$M_i(x) = \begin{cases} 0, & x \notin S_i \\ 1, & x \in S_i \end{cases} \quad (3)$$

式中: x 表示视角掩模 M_i 像素位置。按照前、后、侧和顶 4 个视角分割出 4 个视图信息为 $\{V_i \mid 0, 1, 2, 3\}$ 。同时,将全局分支得到的输出特征 F_g 与分割后的 4 个视图点

乘,得到视角特征,如式(4)所示。

$$F_i^k = V_i \otimes F_g(i = 0, 1, 2, 3) \quad (4)$$

式中: \otimes 表示矩阵相乘, k 表示第 k 个部件。随后将得到的4个视角特征输入局部分支。

2) 多尺度注意力模块(MAB)

传统的车辆重识别使用注意力方法多是为了关注局部特征,然而目前的重识别方法容易遗漏小的视觉区域。为了挖掘不同尺度的局部特征及特征的通道间关系,本文设计了MAB模块。该模块中将通道注意力^[21]多尺度注意力共同作用。同时在该模块中进行权重共享,将指导分支的特征提取方式传授给学习分支。

在通道注意力分支中,为了有效地计算通道注意力,采用压缩输入特征映射的空间维度的方法。先将特征图分别通过最大全局池化和全局平均池化操作得到平均池化特征和最大池化特征,将特征维度压缩为 1×1 。随后将两个特征向量通过一个共同的全连接层和隐藏层,将两个输出结果相加并 sigmoid 激活得到通道注意力 $C^k \in R^{1 \times 1 \times c}$ 。

在多尺度注意力分支中,为了让模型更关注多尺度特征并使区域尺度更均匀,将输入特征依次通过卷积核大小分别为3、2、1、2、3的卷积层得到多尺度注意力 $S^k \in R^{1 \times c \times 1}$,其中步长分别设置为4、2、1、2、4。通过不同大小卷积核的级联,提取得到尺度更丰富的特征,与通道注意力互补。本文在指导分支和学习分支都使用MAB模块,整体如式(5)所示。

$$M(F^k, P^k) = (C^k \otimes S^k) \otimes (F^k, P^k) \quad (5)$$

其中, $C^k \in R^{1 \times 1 \times c}$ 表示通道注意力, $S^k \in R^{1 \times c \times 1}$ 表示多尺度注意力, \otimes 表示矩阵相乘。如图4所示。

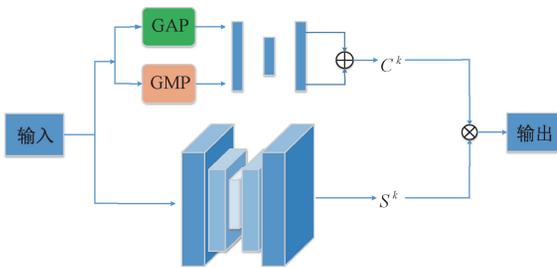


图4 MAB模块

Fig. 4 The MAB module

算法1总结了弱监督学习掩模生成和多尺度注意力输出的详细过程。

算法1 本文算法流程

Input: F to U-net and Resnet50

Output F_i^k

$\bar{F} = \text{reshape}(F)$

for $c = 1$ to C

$T(F) = \text{softmax}(\text{GAP}(F) \otimes \bar{F})$

$\hat{F} = \text{softmax}(T(F) \otimes \bar{F})$ (通道对齐)

for $\hat{F} \leftarrow \text{Box}(F_c)$

Compute the central coordinate (x_c, y_c)

K-means. (x_c, y_c) to $\{C_k\}_{k=1}^K$

end for

end for

for $k = 1$ to K

Mask = $M_f(\text{Box}(\hat{F}_c))$ (部件掩模生成)

end for

$P^k = \text{Mask}(F^k)$

$M(F^k, P^k) = (C^k \otimes S^k) \otimes (F^k, P^k)$ (多尺度注意力输出)

3) 迁移损失

为了实现迁移学习,拉近指导分支和学习分支彼此的距离,以便学习分支能学到指导分支传递的车辆部件特征。本文提出一种批次部分转移的迁移损失来约束指导分支和学习分支之间特征图的相似性。在双分支的每个部件级流中,指导分支通过将车辆零件概念提炼来指导学习分支。尽管某个车辆部件的位置在不同图像之间有所不同,但小批次中的平均位置保持稳定,一个车辆零件的平均位置在不同批次之间是一致的。具体地,首先对指导分支和学习分支的输出特征 \bar{F}^k 和 \bar{P}^k 的批次的通道维度应用全局平均运算,如式(6)所示。

$$(\bar{F}^k, \bar{P}^k) = \frac{1}{C} \sum_1^C \left(\frac{1}{B} \sum_1^B (\tilde{F}_i^k, \tilde{P}_i^k) \right) \quad (6)$$

式中: \tilde{F}_i^k 和 \tilde{P}_i^k 分别表示第 i 个指导分支和学习分支的输出特征,其中 C 和 B 分别表示通道数和批次大小。

之后应用迁移损失来约束指导分支和学习分支的图像之间的相似性,从而进一步优化模型的训练过程。如式(7)所示。

$$L_{FP} = \frac{1}{H} \sum_{i=1}^H \left(\frac{1}{W} \sum_{i=1}^W \| \tilde{F}^k - \tilde{P}^k \|_2 \right) \quad (7)$$

式中:高度为 H ,宽度为 W , $\| \cdot \|_2$ 表示欧氏距离。 \bar{F}^k 和 \bar{P}^k 表示式(6)平均运算后的指导分支和学习分支的输出特征。

1.3 总损失

在本次车辆重识别任务中,总损失包括全局损失 L_G 、指导分支损失 L_P 、学习分支损失 L_F 和迁移损失 L_{FP} 。其中全局损失、指导分支损失和学习分支损失都受到交叉熵损失和三元损失的监督,单独优化其中任何一个损失函数都是低效的。为了增强模型的学习能力,本文针对这些损失使用了多任务学习,对之前提出的损失函数进行适当的组合有利于模型的训练。为了平衡每个损失的权重,使用HUL^[22]来自动学习最优权重。首先将3个FC层分别应用于 L_G 、 L_P 和 L_F 来获得输出softmax概率

\hat{y}^s, L^s 表示全局分支特征、指导分支特征和学习分支特征的交叉熵损失,其中 $s \in \{G, P, F\}$ 。随后 HUL 为每个 \hat{y}^s 引入噪声参数 σ_s ,随着噪声参数的增加,其各自目标的损失权重 σ_s 减小,交叉熵损失 L_{ID} 如式(8)所示。

$$L_{ID} = \sum_{s \in \{G, P, F\}} \frac{1}{\sigma_s^2} L^s(\hat{y}^s, y) + \sum_{s \in \{G, P, F\}} \log \sigma_s \quad (8)$$

其中, y 表示身份标签。同样三元损失 L_{tri} 如式(9)所示。

$$L_{tri} = \frac{1}{3}(L_{tri}^G + L_{tri}^P + L_{tri}^F) \quad (9)$$

最后总损失 L 表示为交叉熵损失、三元损失与迁移损失之和,如式(10)所示。

$$L = L_{ID} + L_{tri} + L_{FP} \quad (10)$$

2 实验

2.1 数据集

本文的模型分别在 VeRi-776^[3] 和 VehicleID^[23] 两个主流数据集上进行评价。

VehicleID 是由 Liu 等^[23] 提出的大型数据集,其中包括了主要从前、后两个视角搜集的约 26 267 辆车的 221 763 张图片集。按数据集的大小,作者将其分为大、中、小 3 个子集。并为数据集中的每张图片都标注了车辆 ID、摄像头位置和车型详细信息等车辆属性。

VeRi-776 也是车辆重识别任务中较为重要的数据集之一。其中包含了 776 辆车的约 50 000 张图片。在标注信息上,不仅有边界框、颜色、车型和厂家信息外,还具有时间戳、地理位置等信息。这些图片分别从 20 个摄像头的不同视角进行拍摄,具有丰富的视角信息。

两个常用的数据集都包含了大量的标注,且拍摄图像考虑到在复杂的交通场景中进行车辆重识别不可避免的伴随着剧烈视角变化,这使得同一车辆在不同视角下要么外观差异巨大产生类内差异性,要么不同车辆由于颜色、车型相同在特定视角下外观却极为相似产生类间相似性,都给车辆重识别带来了巨大的挑战。因此本实验中采用的数据集中样本都包含外观相似但有细微差别的车辆,如图 5(a) 所示,每一行为同一辆车的不同视角图片,每一列为两个不同车辆的相似视角。其中 VeRi-776 数据集中拍摄图片中存在大多包含不止目标车辆一个物体,如图 5(b) 所示,出现背景复杂的情况。更符合本文提出方法的验证,因此在消融实验中使用该数据集以验证本文方法提出模块的作用。两个数据集属性对比如表 1 所示。



(a) 两辆不同车型和颜色相同的车辆示例
(a) Examples of two vehicles with different models and the same color



(b) VeRi-776数据集背景复杂示例
(b) Complex background diagram of the VeRi-776 data set

图 5 数据集车辆示意图

Fig. 5 Vehicle diagram of data set

表 1 VeRi-776 和 VehicleID 数据集属性对比

数据集	#ID	#图片数	#视角
VeRi-776	776	49 357	8
VehicleID (S)	14 042	119 717	2
VehicleID (M)	16 841	131 074	2
VehicleID (L)	17 037	130 293	2

2.2 评价指标

为了评估本网络模型的性能,本文采用两个指标作为评价指标:累积匹配特性(cumulative matching characteristics, CMC)和平均精度(mean average precision, mAP)。CMC 表示查询图像 q 在前 k 个检查结果中匹配到正确图像的概率,其计算如式(11)所示。

$$CMC@K = \frac{\sum_{q=1}^Q gt(q, k)}{Q} \quad (11)$$

其中, Q 表示在查询集中图像总数,当图像 q 的正确匹配图像出现在检索结果序列的第 K 个之前, $gt(q, k) = 1$ 。每个查询 q 的平均精度如式(12)所示。

$$AP(q) = \frac{\sum_{k=1}^n P(k) \times gt(k)}{N} \quad (12)$$

式中: n 为数据集检索的数量, N 为查询集数量。 $P(k)$ 表示匹配序列在点 k 处的精度,若 k 处匹配图像正确,则表示 $gt(k) = 1$, 否则为 0。通过式(13)计算所有查询图像的平均精度。

$$mAP = \frac{\sum_{q=1}^Q AP(q)}{Q} \quad (13)$$

其中, Q 是所有查询集中图像总数, $AP(q)$ 表示每个查询图像 q 的平均精度检索精度。

2.3 与最新方法的比较

1) 在 VeRi-776 上的实验

在 VeRi-776 数据集上对本文车辆重识别方法进行了评估, 其中采用了 mAP、CMC@1 和 CMC@5 这 3 个指标。表 2 呈现了本文的方法与最先进方法的比较结果。

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
PRN ^[24]	74.30	94.34	98.91
PVEN ^[25]	79.50	95.62	98.43
SAVER ^[26]	79.60	96.40	98.61
TBE-Net ^[27]	79.50	96.00	98.50
VPEN ^[28]	81.20	97.30	99.10
DFNet ^[29]	80.97	97.08	99.01
TransReid ^[30]	82.00	97.10	-
VABPP ^[31]	80.80	96.60	-
DRA ^[32]	78.47	76.03	87.33
SSBVER ^[33]	80.94	97.02	98.45
HPGN ^[34]	80.2	96.7	-
本文	82.12	98.50	99.12

其中 PVEN 通过人工标注为图片进行标注, 使用解码信息来对齐车辆。而本文的方法在进行部件对齐时使用了弱监督方法, 生成部件掩模时不需要手工标注, 节约了人工成本, 与 PVEN 相比本文方法在 mAP 上提升 2.6%。在 TBE-Net 中设计了一种三支嵌入模型, 通过全局分支和局部分支学习车辆的全局外观特征和局部显著特征, 使用补充性分支通过合并来学习更丰富的车辆特征。然而在真实的交通场景中, 现场的车辆拍摄数据容易受到视角变化的影响, 导致拍摄车辆尺度大小不同, 使得现有方法很难获得预定义的局部特征及背景信息, 从而影响重识别任务的准确性。相比之下, 本文在网络中设计的通道注意力和多尺度空间注意力结合, 能够获得更多的尺度信息, 丰富车辆特征, 并且 mAP 和 CMC@1 分别提升 2.6% 和 2.5%。HGN 同样是在没有额外手动标签下进行训练, 与之相比本文方法在 mAP 和 CMC@1 分别提升 1.9% 和 1.8%。通过实验证明本文方法在不需要手工标注的情况下, 也能自适应地选取显著的局部特征。

2) 在 VehicleID 上的实验

将本文所提出的方法在 VehicleID 上进行了测试。

表 3 为 VehicleID 数据集上与其他前沿的车辆重识别方法的 CMC@1 和 CMC@5 比较结果。

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 VehicleID

方法	小		中		大	
	C@1	C@5	C@1	C@5	C@1	C@5
Methods						
VAMI ^[35]	63.12	83.25	52.87	75.12	47.34	70.29
PRN ^[24]	78.92	94.81	74.94	92.02	71.58	88.46
PVEN ^[25]	0.847	0.970	0.806	0.945	0.778	0.920
TBE-Net ^[27]	86.0	98.4	82.3	96.6	80.7	94.9
DFNet ^[29]	84.76	96.22	80.61	94.10	79.15	92.86
TransReid ^[30]	85.20	97.50	-	-	-	-
VABPP ^[31]	82.80	97.90	79.50	95.40	76.20	92.40
DRA ^[32]	-	-	-	-	79.25	91.03
SSBVER ^[33]	85.61	97.73	80.34	94.92	77.26	92.59
VPEN ^[28]	0.857	0.980	0.821	0.962	0.801	0.943
本文	86.93	98.7	83.74	96.73	81.02	94.37

在上述最新的方法中, 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, 而本文在网络中设计迁移学习, 减少了部件掩模的训练成本并提高了重识别的精度, 在 VehicleID 的规模最小的子集上 CMC@1 和 CMC@5 上本文方法分别提高 1.7% 和 3.2%。与沿空间维度、通道维度划分的 PRN 相比, 本文方法不仅可以得到每个部件的多尺度特征, 还可以避免背景对局部特征造成的影响, 在 VehicleID 大规模测试子集上的 CMC@1 与 CMC@5 分别提高了 9.5% 和 5.9%。VPEN 通过视角感知下的局部增强模块自适应学习每个视角下有区别性的局部特征, 然而它很难解决同一摄像机视角下不同车辆在背景相似时难以识别的问题。而本文的方法通过在 WAM 模块中弱监督生成车辆部件掩模, 减少了背景带来的影响, 与之相比在 VehicleID 的规模最小的子集上 CMC@1 和 CMC@5 上分别提高 1.4% 和 0.7%。通过实验性能的对比, 证明了本文模型对复杂背景的鲁棒性。

2.4 消融实验

本文进行了消融实验来验证所提出的成分的有效性, 包括 WAM 模块, 以及知识共享模块中的 MAB 模块。在 VeRi-776 数据集上进行消融实验, 并采用了 mAP、CMC@1 和 CMC@5 这 3 个指标, 以进一步验证实验的有效性。

1) WAM 模块的作用

WAM 模块为知识共享模块的指导分支提供弱监督部件掩模。将 WAM 模块替换成简单的语义分割, 它对掩模图像进行局部分块, 分割成 3 个部分, 从而减少背景噪声对局部特征的影响, 方便捕获到特征块的细节信息。

其结果如表 4 所示。其中仅使用 Resnet-50 作为主干网络并视角对齐操作的模型作为 Baseline, 语义分割表示为 R。

表 4 在 VeRi-776 上验证 WAM 模块作用

Table 4 verifies the WAM module role on VeRi-776

设置	mAP	C@ 1	C@ 5
Baseline	74. 30	94. 34	98. 91
Baseline+R	79. 50	95. 62	98. 43
Baseline+WAM	79. 60	96. 40	98. 61

如图 6 所示,使用简单的分割方法是不稳定的,当车身分布不均匀时,使用语义分割均匀分割会存在空间错位,可能遗漏一些关键信息。相比之下,使用 WAM 模块的弱监督掩模能够更关注到突出部件如车顶、车灯及车窗等,并且对背景复杂也有很好的鲁棒性。

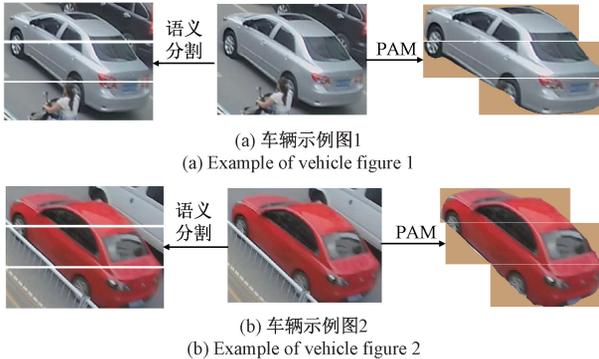


图 6 语义分割和 WAM 掩模图

Fig. 6 Masks for semantic segmentation and WAM

2) 知识共享模块的作用

为了减少弱监督方法生成部件掩模不稳定的影响,在知识共享模块中构建 MAB 模块进行权重共享,以实现迁移学习。并且除了 WAM 模块中的弱监督注意力外,在 MAB 模块中还对车辆部件进行二次注意力学习,以得到像素级的局部特征,减少类间相似性。因此在知识共享模块中分析 MAB 模块的有效性分析。

如图 7 所示,展示了 MAB 模块生成的注意力图。由图 7(b)和图 7(c)对比可以发现,经过 MAB 模块的学习分支能够捕捉到车辆局部部件中更突出的特征,如车灯、车标和年检标志等。

将 MAB 模块与其他效果最好的注意力方法进行比较,以验证其优越性。在现有的车辆重识别方法中,SE-Net^[36]视觉中常用的通道注意力。CBAM^[21]包含了通道注意力模块和空间注意力模块,分别进行通道与空间上的注意力表示。其结果如表 5 所示,其中 Baseline 为一个消融实验中加入 WAM 模块的模型。

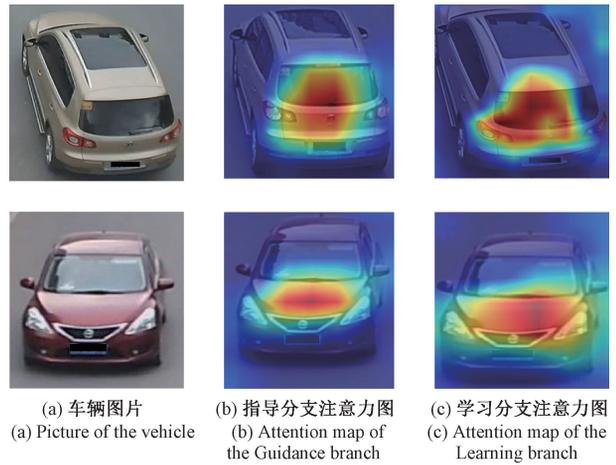


图 7 MAB 模块注意力图

Fig. 7 Attention map of the MAB module

表 5 在 VeRi-776 上验证 MAB 模块作用

Table 5 verifies the role of the MAB module on VeRi-776

设置	mAP	C@ 1	C@ 5
Baseline	79. 60	96. 40	98. 61
Baseline+SE-Net	80. 50	96. 62	98. 73
Baseline+CBAM	81. 10	96. 77	98. 82
Baseline+MAB	81. 60	96. 89	98. 97

3 结 论

为解决现实交通场景中弱监督车辆重识别易受各种复杂背景干扰,并且不准确标签影响重识别精度的问题。本文提出一种基于弱监督注意力和知识共享的车辆重识别网络。相比于以往的手工特征提取算法,本文方法基于弱监督学习,避免了额外的人工成本。通过通道对齐步骤减少了背景复杂对弱监督方法的影响,可以获得更精细的车辆部件掩模。在局部分支中构建知识共享模块,其中设计多尺度注意力模块,通过权重共享和迁移损失来引导迁移学习,避免了弱监督方法的标签准确性不稳定的影响。该模块还通过通道注意力和多尺度卷积层的结合,可以获得车辆更丰富的多尺度特征,同时保留丰富的细节信息。实验表明,本文的方法在车辆重识别领域具有良好的效果。

参考文献

[1] 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.
 [2] LIU X C, LIU W, MA H D, et al. Large-scale vehicle

- reidentification in urban surveillance videos [C]. Proceedings of IEEE International Conference on Multimedia and Expo (ICME). Seattle, USA: IEEE, 2016; 1-6.
- [3] LIU X C, LIU W, MEI T, et al. A deep learning-based approach to progressive vehicle re-identification for urban surveillance [C]. Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands: Springer, 2016; 869-884.
- [4] LIU X B, ZHANG S L, HUANG Q M, et al. RAM: A regionaware deep model for vehicle re-identification[C]. Proceedings of 2018 IEEE International Conference on Multimedia and Expo (ICME). San Diego, USA: IEEE, 2018; 1-6.
- [5] WANG Z, TANG L M, LIU X H. Orientation invariant feature embedding and spatial temporal regularization for vehicle re-identification [C]. Proceedings of IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017;379-387.
- [6] WEI X S, LUO J H, WU J, et al. Selective convolutional descriptor aggregation for fine-grained image retrieval[J]. IEEE Transactions on Image Processing, 2017, 26(6): 2868-2881.
- [7] ZHENG Z, ZHENG L, YANG Y. Pedestrian alignment network for large-scale person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 29(10):3037-3045.
- [8] YAO H, ZHANG S, HONG R, et al. Deep representation learning with part loss for person re-identification [J]. IEEE Transactions on Image Processing, 2019, 28(6):2860-2871.
- [9] 张正, 陈成, 肖迪. 基于图像语义分割的车辆重识别[J]. 计算机工程与设计, 2022, 43(10): 2897-2903.
ZHENG ZH, CHEN CH, XIAO D. Vehicle recognition based on image semantic segmentation [J]. Computer Engineering and Design, 2022, 43(10): 2897-2903.
- [10] ZHI H Z. A brief introduction to weakly supervised learning[J]. National Science Review, 2018, 5(1): 44-53.
- [11] DING C, WANG K, WANG P, et al. Multi-task learning with coarse priors for robust part-aware person re-identification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(3):1474-1488.
- [12] WANG K, WANG P, DING C, et al. Batch coherence-driven network for part-aware person re-identification[J]. IEEE Transactions on Image Process, 2021, 30: 3405-3418.
- [13] WANG K, DING C, MAYBANK S J, et al. CDPM: Convolutional deformable part models for semantically aligned person re-identification[J]. IEEE Transactions on Image Process, 2020, 29:3416-3428.
- [14] OLAF R, PHILIPP F, THOMAS B. U-net: Convolutional networks for biomedical image segmentation [C]. Proceedings of Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015-18th International Conference. Munich, Germany, 2015;234-241.
- [15] OLGA R, DENG J, SU H, et al. Imagenet large scale visual recognition challenge[J]. International Journal of Computer Vision, 2015, 115(3):211-252.
- [16] KAIMING H, XIANGYU Z, SHAOQING R, et al. Deep residual learning for image recognition[C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016; 770-778.
- [17] HERMANS A, BEYER L, LEIBE B. In defense of the triplet loss for person re-identification[J]. arXiv preprint arXiv:1703.07737, 2017.
- [18] 宁轩, 何智敏. 基于深度学习的车辆重识别算法综述[J]. 数据通信, 2023, 212(1):43-47.
NING X, HE ZH M. Summary of vehicle re-identification algorithms based on deep learning [J]. Data Communications, 2023, 212(1):43-47.
- [19] ROTHER C, KOLMOGOROV V, BLAKE A. “Grabcut” interactive foreground extraction using iterated graph cuts[J]. ACM Transactions on Graphics (TOG), 2004, 23(3): 309-314.
- [20] XIN Q, LV T, GAO H. Random part localization model for fine grained image classification[C]. Proceedings of IEEE International Conference on Image Processing (ICIP). Taipei, 2019;420-424.
- [21] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]. Proceedings of the European conference on computer vision (ECCV). Munich, Germany, 2018;3-19.
- [22] KENDALL A, GAL Y, CIPOLLA R. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA, 2018;7482-7491.
- [23] LIU H, TIAN Y, YANG Y. Deep relative distance learning: Tell the difference between similar vehicles [C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, USA, 2016;2167-2175.
- [24] HE B, LI J, ZHAO Y. Part-regularized near-duplicate

- vehicle re-identification [C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2019; 3997-4005.
- [25] MENG D, LI L, LIU X J. Parsing-based view-aware embedding network for vehicle re-identification [C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition. Seattle, WA, USA, 2020; 7101-7110.
- [26] KHORRAMSHAHI P, KUMAR A, PERI N. A dual-path model with adaptive attention for vehicle Re-identification [C]. Proceedings of IEEE International Conference on Computer Vision. Seoul, Korea (South), 2019;6131-6140.
- [27] SUN W, DAI G, ZHANG X. TBE-Net: A three branch embedding network with part-aware ability and feature complementary learning for vehicle Re-identification [C]. Proceedings of the IEEE Transactions on Intelligent Transportation Systems, 2021;1-13.
- [28] 代广昭, 孙伟, 徐凡. 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.
- DAI G ZH, SUN W, XU F. Perspective-perceived local attention network for vehicle recognition [J]. Computer Engineering, 2022, 48 (10): 288-297.
- [29] BAI Y, LIU J, LOU Y. Disentangled feature learning network and a comprehensive benchmark for vehicle re-identification [C]. Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021; 1-17.
- [30] HE S, LUO H, WANG P. TransReID: Transformer-based object re-identification [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021;15013-15022.
- [31] LI H, WANG Y, WEI Y. Discriminative-region attention and orthogonal-view generation model for vehicle re-identification [C]. Applied Intelligence, 2022;1-18.
- [32] KHORRAMSHAHI P, SHENOY V, CHELLAPPA R. Scalable vehicle re-identification via self-supervision [J]. arXiv preprint arXiv: 2205. 07613, 2022.
- [33] YAO A, QI J, ZHONG P. Self-aligned spatial feature extraction network for uav vehicle re-identification [J]. arXiv preprint arXiv: 2201. 02836, 2022.
- [34] SHEN F, ZHU J, ZHU X, et al. Exploring spatial significance via hybrid pyramidal graph network for vehicle re-identification [J]. IEEE Transactions on Intelligent Transportation Systems, 2021;1-12.
- [35] ZHOU Y, SHAO L. Aware attentive multi-view inference for vehicle re-identification [C]. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2018; 6489-6498.
- [36] HU J, SHEN L, SUN G. Squeeze-and-excitation networks [J]. arXiv preprint arXiv: 1709.01507, 2017, 7.

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年为普渡大学 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 now 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.



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 now 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 now 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.



刘轩诚, 于 2021 年获得兰州理工大学自动化专业学士学位, 目前在南京信息工程大学攻读控制科学与工程硕士学位, 主要研究方向为小目标检测、轻量级神经网络和深度学习。

E-mail: 865441214@qq.com

Liu Xuancheng received the B. Sc. degree in automation from Lanzhou University of Technology in 2021. He is now a M. Sc. candidate in control science and engineering at Nanjing University of Information Science and Technology. His main research interests include small target detection, lightweight neural network and deep learning.