



中国科技期刊卓越行动计划项目入选期刊

控制与决策

CONTROL AND DECISION



基于多尺度混合注意力与度量融合的小样本行人重识别

陈贵震, 邹国锋, 刘月, 傅桂霞, 高明亮

引用本文:

陈贵震, 邹国锋, 刘月, 傅桂霞, 高明亮. 基于多尺度混合注意力与度量融合的小样本行人重识别[J]. *控制与决策*, 2024, 39(5): 1441–1449.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2022.1402>

您可能感兴趣的其他文章

Articles you may be interested in

行人重识别中度量学习方法研究进展

A survey on metric learning in person re-identification

控制与决策. 2021, 36(7): 1547–1557 <https://doi.org/10.13195/j.kzyjc.2020.0801>

Anchor-free的尺度自适应行人检测算法

Anchor-free scale adaptive pedestrian detection algorithm

控制与决策. 2021, 36(2): 295–302 <https://doi.org/10.13195/j.kzyjc.2020.0124>

基于多尺度特征表示的行人再识别

Multi-scale feature representation for person re-identification

控制与决策. 2021, 36(12): 3015–3022 <https://doi.org/10.13195/j.kzyjc.2020.0952>

基于度量学习和典型相关分析的亲缘关系识别网络

Kinship relationship recognition network based on metric learning and canonical correlation analysis

控制与决策. 2021, 36(8): 1977–1983 <https://doi.org/10.13195/j.kzyjc.2019.1798>

一种基于多层语义特征的图像理解方法

An image understanding method based on multi-level semantic features

控制与决策. 2021, 36(12): 2881–2890 <https://doi.org/10.13195/j.kzyjc.2020.0927>

基于多尺度混合注意力与度量融合的小样本行人重识别

陈贵震, 邹国锋[†], 刘月, 傅桂霞, 高明亮

(山东理工大学 电气与电子工程学院, 山东 淄博 255049)

摘要: 针对行人重识别中可用行人图像不足导致的小样本问题, 以双相似网络为基础, 提出一种基于多尺度混合注意力与度量融合的小样本行人重识别方法. 首先, 将多尺度混合注意力机制引入特征嵌入模块, 即在不同尺度层内的特征提取中引入空间注意力, 在不同尺度层间的特征融合中引入通道注意力, 实现更具判别力的小样本行人特征提取; 然后, 在度量模块, 提出欧氏距离与余弦距离融合的双重度量方法, 实现行人特征的空间绝对距离和方向差异的综合度量, 提升行人相似性度量的可靠性; 接着, 采用双重度量方式和关系度量方式, 分别获得行人特征的相似度得分; 最后, 通过加权融合获得联合度量得分, 构建联合损失实现网络的整体优化和训练. 在 Market-mini、Duke-mini 和 MSMT17-mini 三个小型数据集上的实验表明, 所提出方法在 5-way 1-shot 和 5-way 5-shot 两种模式下的平均识别准确率分别达到 90.40% 和 95.69%、86.77% 和 94.96%、71.08% 和 82.63%, 与其他小样本学习算法相比, 识别性能有较大提升.

关键词: 行人重识别; 小样本; 双相似网络; 多尺度混合注意力; 度量融合; 双重度量

中图分类号: TP391.41

文献标志码: A

DOI: 10.13195/j.kzyjc.2022.1402

引用格式: 陈贵震, 邹国锋, 刘月, 等. 基于多尺度混合注意力与度量融合的小样本行人重识别[J]. 控制与决策, 2024, 39(5): 1441-1449.

Few-shot for person re-identification based on multi-scale mixed attention and metric fusion

CHEN Gui-zhen, ZOU Guo-feng[†], LIU Yue, FU Gui-xia, GAO Ming-liang

(School of Electrical and Electronic Engineering, Shandong University of Technology, Zibo 255049, China)

Abstract: To solve the few-shot problem caused by insufficient available pedestrian images in person re-identification, based on the Bi-Similarity, a few-shot person re-identification method based on multi-scale mixed attention and metric fusion is proposed. In this work, firstly, a multi-scale mixed attention method is introduced into the feature embedding module. This method introduces spatial attention in different feature extraction layers and introduces channel attention in the feature fusion between different scale layers, which can extract more discriminative pedestrian features. Secondly, a dual metric method combining Euclidean and cosine distance is proposed in the metric module to comprehensively measure the absolute spatial distance and directional difference of pedestrian features. In this way, the reliability of pedestrian similarity measurement is improved. Then, pedestrian feature similarity scores are obtained separately using the dual metric and relation metric methods. Finally, the combined metric score is obtained by weighted fusion, and the combined metric score is used to construct the joint loss to realize the overall optimization and training of the network. Experimental results on three small datasets, Market-mini, Duke-mini, and MSMT 17-mini, show that the proposed method significantly improves recognition performance compared to other few-shot learning algorithms. Specifically, in scenarios 5-way 1-shot and 5-way 5-shot, the average recognition accuracies are 90.40% and 95.69%, 86.77% and 94.96%, and 71.08% and 82.63%, respectively.

Keywords: person re-identification; few-shot learning; Bi-Similarity network; multi-scale mixed attention; metric fusion; dual metric

0 引言

行人重识别(person re-identification)是跨视域行人身份关系的匹配过程,属于图像检索的子问题^[1-2].

近年来,基于深度学习^[3-8]的行人重识别算法越来越受到研究者的青睐,但是,该类算法严重依赖于海量的高质量标注行人图像.然而,受遮挡、光照以及人工

收稿日期: 2022-08-03; 录用日期: 2023-02-10.

基金项目: 山东省自然科学基金项目(ZR2022MF307); 国家自然科学基金项目(61801272); 山东省重点研发计划项目(重大科技创新工程项目)(2019JZZY010119).

责任编辑: 张文安.

[†]通讯作者. E-mail: zgf841122@163.com.

标注成本昂贵等因素影响,在很多应用场景中,仅能够采集到少量的可用行人图像和标记缺失的行人数据.因此,样本数量或类别标记不足导致的小样本问题成为行人重识别领域的一项巨大挑战.

受人类仅通过一个或几个示例即可建立对新事物认知的启发,小样本学习^[9]的概念被提出.其中,小样本学习主要解决两类问题:1)数据集规模有限,各类样本数量有限的学习和识别问题;2)数据集规模较大,但是,数据集中样本类别标记缺失或标记错误的学习和识别问题.小样本行人重识别面临同样的挑战.

针对类别标记不足的小样本行人重识别问题,国内外学者做了大量的研究工作^[10-16]. Wu等^[13]提出了渐进式行人重识别学习技术,放大采样的带伪标记数据子集,为每个未标记数据分配标签; Xu等^[14]提出了特征空间正则化方法的行人重识别方法,利用特征空间正则化损失来平衡特征空间中样本的分布;行人步态特征也可作为鲁棒性的识别特征,贾晔焯等^[15]对行人步态识别的现状进行综述,并对步态识别的各种方法进行了总结和分析; Ge等^[16]提出了双网络协同训练的行人重识别技术,通过无监督框架柔性目标领域的伪标签,提炼“硬”伪标签和“软”伪标签来学习目标领域更好的特征.

当前,解决数据量不足问题的小样本算法大致分为两类:基于元学习和基于度量的方法.基于元学习的方法侧重于学会学习,使得模型根据经验,学习良好的初始化、优化器或参数,如 Finn等^[17]、Ravi等^[18]和 Gidaris等^[19]的研究;基于度量的方法侧重于学习良好的特征嵌入或相似度量或两者兼而有之,如 Vinyals等^[20]、Snell等^[21]以及 Sung等^[22]的研究.2020年, Li等^[23]提出了双相似网络(BSNet)结构,以解决小样本环境下鸟类和汽车识别问题.尽管针对数据不足的小样本学习问题已有较多研究,但是,受到行人重识别中干扰因素的影响,以上研究并未在行人重识别领域得以应用.鉴于数据量不足的行人重识别问题的重要性以及缺乏相关研究的现状,本文将BSNet为基础,结合行人图像的特点,研究适用于小样本行人重识别的深度网络架构.

BSNet由单个特征嵌入模块和包含两个相似度量模块(关系模块和余弦模块)的双相似模块组成.在训练过程中,网络先通过特征嵌入模块提取图像特征;再将特征输入至双相似模块,得到关系和余弦度量得分;最后,根据相似得分计算网络联合损失,进行反向传播调整参数.该网络在处理数据不足的图像

识别问题时,表现出良好的性能.但是,仍然存在两个方面的问题:1)网络中的特征嵌入模块仅采用4层卷积,其卷积层数不足,导致特征缺乏语义信息,含有较多噪声.因此,无法获取少量样本关键的特征表示,从而影响网络的识别性能.2)网络中的余弦模块仅考虑特征的方向差异,而未考虑其绝对距离,这意味着并没有完全考虑特征间的差异,导致在样本量有限的情况下,难以获得可靠的度量得分.

综上所述,为了解决行人图像不足情况下的小样本识别问题,本文以双相似网络为基础,提出一种基于多尺度混合注意力与度量融合的小样本行人重识别算法.首先,在特征嵌入模块中引入多尺度混合注意力机制,增强小样本行人图像特征的判别能力,其中,所提出多尺度混合注意力机制,先由空间注意力模块从不同的尺度中获取样本更为充足的特征信息,再将空间注意力模块获得的特征以Add方式融合,由通道注意力模块将融合后的特征在通道维度重标定,使得提取的行人特征更具判别性,从而弥补行人数据不足的问题;然后,在行人特征的度量过程中,特征经双重度量模块和关系模块得到不同的度量得分,再经加权融合得到联合度量得分,其中,所提出双重度量模块以及欧氏距离作为余弦度量权重,综合考虑样本特征的方向差异和绝对距离,从而获得更加可靠的度量得分,关系模块采用双相似网络中的方法;最后,在元训练过程中,根据两种度量得分计算网络联合损失,用以反向调整网络参数.实验中,在Market-1501、DukeMTMC-reID和MSMT 17数据集上分别随机抽取部分数据,生成3个小型数据集:Market-mini、Duke-mini和MSMT 17-mini,并在小型数据集上进行实验验证,表明所提出算法的有效性.

1 双相似网络(BSNet)

双相似网络是一个通过计算支持集与查询集样本间关系得分,对新类别样本图像进行识别的小样本学习框架,双相似网络结构如图1所示.

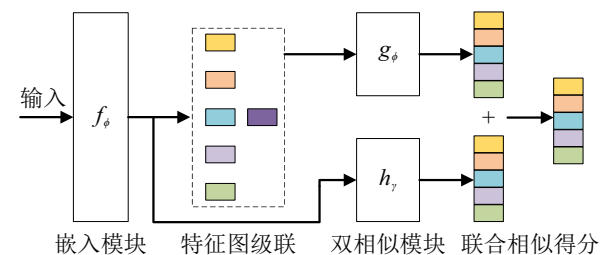


图1 双相似网络结构

特征嵌入模块 f_ϕ 用于提取样本图像抽象特征,生成支持集和查询集样本的特征表示.该模块由4个卷积块构成,前两个卷积块各自包含一个最大池化

层. 其中: 每个卷积块由一个卷积核大小为 3×3 , 通道数为64的卷积层、一个批归一化层、一个Relu激活层以及一个池化层组成. 输入图像经特征嵌入模块提取特征后, 进入双相似模块. 双相似模块由关系模块 g_φ 和余弦模块 h_γ 构成.

关系模块由2个卷积块和2个全连接层组成. 该模块最终产生支持集样本每个类的原型与查询集间的关系相似度得分, 该得分公式如下式所示:

$$S_{i,j}^1 = g_\varphi \left(\frac{1}{K} \sum_{k=1}^K f_\phi(x_{i,k}) || f_\phi(x_j) \right). \quad (1)$$

其中: $S_{i,j}^1$ 为关系模块的相似度得分, K 为支持集中每个类别包含的样本个数, $x_{i,k}$ 为第 i 个类别的第 k 个样本, x_j 为查询集图片, $||$ 为特征级联.

余弦模块由2个卷积块和1个余弦相似层组成. 其中: 第1个卷积块包含最大池化层, 第2个卷积块包含平均池化层. 其查询样本与给定支持集样本的每个类原型间的余弦相似度得分公式如下式所示:

$$S_{i,j}^2 = h_\gamma^{\cos} \left(h_\gamma^{\text{em}} \left(\frac{1}{K} \sum_{k=1}^K f_\phi(x_{i,k}) \right), h_\gamma^{\text{em}} (f_\phi(x_j)) \right). \quad (2)$$

其中: $S_{i,j}^2$ 为余弦模块的相似度得分, h_γ^{\cos} 为余弦相似层, h_γ^{em} 为余弦模块的2个卷积块. 双相似模块的联合得分 $S_{i,j}$ 如下式所示:

$$S_{i,j} = \frac{1}{2} S_{i,j}^1 + \frac{1}{2} S_{i,j}^2. \quad (3)$$

训练中的损失函数采用均方误差MSE^[24], 其联合损失loss如下式所示:

$$\begin{aligned} \text{loss}_1 &= \frac{1}{C \times Q} \sum_{i=0}^{C-1} \sum_{j=1}^Q (S_{i,j}^1(i) - 1(y_i = y_j))^2, \\ \text{loss}_2 &= \frac{1}{C \times Q} \sum_{i=0}^{C-1} \sum_{j=1}^Q (S_{i,j}^2(i) - 1(y_i = y_j))^2, \\ \text{loss} &= \lambda \times \text{loss}_1 + (1 - \lambda) \times \text{loss}_2. \end{aligned} \quad (4)$$

其中: loss_1 , loss_2 分别为关系模块和余弦模块的均方误差损失; C 为支持集中不同行人类别数; Q 为查询集样本数; λ 为联合损失权重系数; y_i , y_j 分别为支持集和查询集样本对应标签, 当 $y_i = y_j$ 时, $1(y_i = y_j) = 1$, 反之则为0.

2 基于多尺度混合注意力与度量融合的小样本行人重识别网络

借鉴双相似网络结构, 本文提出了基于多尺度混合注意力与度量融合的小样本行人重识别网络, 该网络由特征嵌入模块和度量模块组成. 其中: 特征嵌入模块包含特征提取模块和多尺度混合注意力模块, 度量模块包含基于度量融合的双重度量模块和关系模块. 网络的整体架构如图2所示.

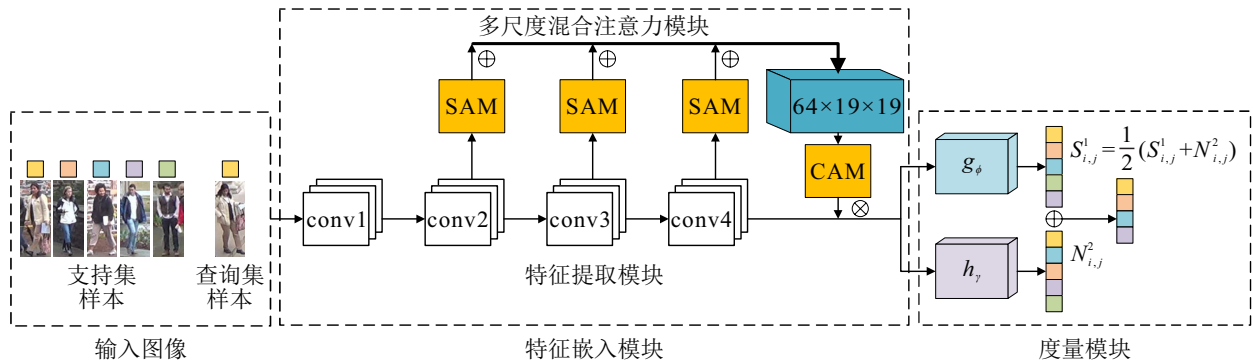


图2 本文网络整体架构

2.1 元框架小样本行人重识别学习

数据量不足的小样本问题, 通常被形式化为 C -way K -shot 问题, 在此问题中, 模型可将未标记的行人, 正确地识别出 C 类中属于该行人的类. 在传统的深度学习中, 训练集、验证集和测试集的标签集是共享的, 而在小样本学习中, 3 者的标签集是不相交的. 具体而言, 给定一个数据集 $D = \{(x_i, y_i), y_i \in L\}_{i=1}^N$, 将其划分为 $D_{\text{train}} = \{(\tilde{x}_i, \tilde{y}_i), \tilde{y}_i \in L_{\text{train}}\}_{i=1}^{N_{\text{train}}}$, $D_{\text{val}} = \{(\hat{x}_i, \hat{y}_i), \hat{y}_i \in L_{\text{val}}\}_{i=1}^{N_{\text{val}}}$ 以及 $D_{\text{test}} = \{(\tilde{x}_i, \tilde{y}_i), \tilde{y}_i \in L_{\text{test}}\}_{i=1}^{N_{\text{test}}}$. 其中: (x_i, y_i) 为第 i 幅图片的原始特征向量和标签信息; 标签集 L_{train} 、 L_{val} 和 L_{test} 是不相交

的, 它们的并集为总标签集.

模型在元学习框架^[25]下完成学习, 即循环生成多个元任务 \tilde{T}_i , 利用元任务完成训练、验证和测试. 以训练集为例, 实验从 L_{train} 中随机选取 C 个不同的行人, 并从 D_{train} 中随机选取每个行人中的 M 张图片. 其中: 每个行人的 M 张图片, 分为 K 和 $M - K$ 张图片两组, 即 $C \times K$ 张图片作为支持集 S_i , $C \times (M - K)$ 张图片作为查询集 Q_i , 以此构成一个元任务. 同样地, 在验证集和测试集上按照相同方法生成元任务. 实验的目标是在训练集的元任务中训练改进的神经网络, 以学习到可转移的行人重识别知识, 通过验证集

的元任务调整超参数,最后由测试集元任务中的模型精度平均值,报告泛化精度.

2.2 多尺度混合注意力机制

如图3所示,多尺度混合注意力模块由SAM (spatial attention module)和CAM(channel attention module)组成.其中: SAM用于探索不同尺度特征空间维度内的关注区域,CAM用于探索融合后特征通道间的相关性.为了保证特征经SAM后满足可加性,对不同尺度特征进行尺度变换操作.

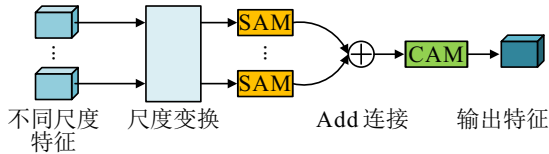


图3 多尺度混合注意力模块

空间注意力模块SAM定义为

$$M_s = \text{sigmoid}(\text{conv}(f_{\text{mp}}(F) \parallel_c f_{\text{ap}}(F))). \quad (5)$$

首先,特征图 F 分别经过最大池化层 f_{mp} 和平均池化层 f_{ap} 后,经过 \parallel_c 按通道维度拼接在一起;然后,拼接后的特征利用卷积层 conv 和 sigmoid 函数生成最终空间注意力图 M_s ;最后, $M_s \times F$ 得到空间加权的新特征.

通道注意力模块CAM定义为

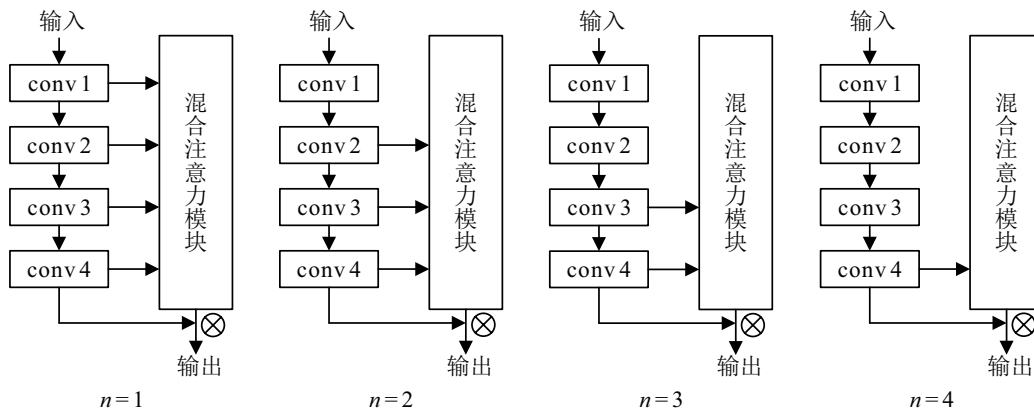


图4 不同n取值对应嵌入模块结构

2.3 基于度量融合的双重度量

本文提出以欧氏距离为权重,修正余弦相似度的度量融合方法,网络结构如图5所示.

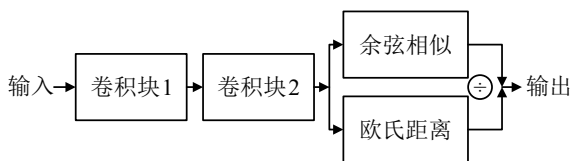


图5 基于度量融合的双重度量模块

首先,支持集样本和查询集样本特征经过两个分别包含最大池化层和平均池化层的卷积块;然后,将

$$M_c = \text{sigmoid}(f_{\text{mlp}}(f_{\text{mp}}(F)) + f_{\text{mlp}}(f_{\text{ap}}(F))). \quad (6)$$

首先,特征图 F 分别经过最大池化层 f_{mp} 和平均池化层 f_{ap} 后,通过共享的多层感知机 f_{mlp} ,该感知机包含两个全连接层FC和一个ReLU激活函数;然后,将得到的特征相加后,输入 sigmoid 函数生成通道注意力图 M_c ;最后, $M_c \times F$ 得到通道加权的新特征.

行人特征包含的位置以及细节等信息,由低层向高层逐层递减,包含的语义信息,由低层向高层逐层递增,因此,所提出多尺度特征融合策略按照由低层特征到高层特征顺序,逐层减少参与融合的低层特征;然后,根据残差网络^[26]构思,经多尺度混合注意力机制处理后的样本特征,引入特征提取模块的输出映射,以保证传入双度量模块的特征包含更多有用的图像信息;最后,基于多尺度混合注意力机制的特征嵌入模块的表达式如下式所示:

$$x = F_4 \times C_n \times M_c \left(\sum_n^4 F_n \times M_s(F_n) \right), \quad n = 1, 2, 3, 4;$$

$$C_n = \sum_n^4 F_n \times M_s(F_n). \quad (7)$$

其中: x 为特征嵌入模块总输出; F_n 为特征提取模块第 n 个卷积块输出. 整体结构如图4所示.

得到的特征同时输入余弦相似层和欧氏距离度量层,分别得到样本特征间的余弦相似得分和欧氏距离得分;最后,根据下式计算双重度量得分:

$$D_{i,j}^2 = h_{\gamma}^{\text{ed}} \left(h_{\gamma}^{\text{em}} \left(\frac{1}{K} \sum_{k=1}^K f_{\phi}(x_{i,k}) \right), h_{\gamma}^{\text{em}}(f_{\phi}(x_j)) \right),$$

$$N_{i,j}^2 = \frac{S_{i,j}^2}{D_{i,j}^2}. \quad (8)$$

其中: $D_{i,j}^2$ 为欧氏距离得分, h_{γ}^{ed} 为欧氏距离度量层, $N_{i,j}^2$ 为重新修正后的度量得分. 支持集样本中不包含查询集图片,因此,二者间的欧氏距离 $D_{i,j}^2$ 不为0,式

(8)中分式成立. 基于此, 所提出网络的最终相似得分和双重度量模块损失 $loss_2$ 分别由式(3)和(4)更新为

$$S_{i,j} = \frac{1}{2}S_{i,j}^1 + \frac{1}{2}N_{i,j}^2, \quad (9)$$

$$loss_2 = \frac{1}{C \times Q} \sum_{i=0}^{C-1} \sum_{j=1}^Q (N_{i,j}^2(i) - 1(y_i = y_j))^2. \quad (10)$$

当样本同属一类时, 余弦相似得分 $S_{i,j}^2$ 较大, 欧氏距离 $D_{i,j}^2$ 较小, 由二者同时约束的双重度量得分 $N_{i,j}^2$ 更大; 反之, 得分更小.

所提出双重度量模块通过上述方法, 利用欧氏距离和余弦距离综合考虑行人特征的绝对距离和方向差异, 以获得更为可靠的行人相似性度量得分.

3 实验和分析

3.1 实验设置

实验对 Market-mini、Duke-mini、MSMT 17-mini 三个数据集采用 5-way 1-shot 和 5-way 5-shot 两种模式进行训练. 训练过程随机取样并设置 epoch 为 200, 每个 epoch 中构建 100 个 episode, 共 20 000 个 episode, 在每个训练的 episode 中, 每个行人设置为 16 张查询图像, 即在一个 episode 中, 5-way 1-shot 模式下训练图像有 $16 \times 5 + 1 \times 5 = 85$ 张, 5-way 5-shot 模式下训练图像有 $16 \times 5 + 5 \times 5 = 105$ 张. 每个 epoch 结束后进行 100 次元验证任务得到验证平均识别准确率, 其中查询图像数量设置与训练任务相同. 测试时, 查询图像数量设置与训练任务相同, episode 数量设置为 1 000, 得到模型泛化精度.

3.1.1 数据集介绍

本文基于 Market-1501、DukeMTMC-reID 以及 MSMT 17 三个数据集, 随机抽取部分样本组成 Market-mini、Duke-mini 和 MSMT 17-mini 三个小型数据集进行实验. 对于每个数据集, 实验以 2 : 1 : 1 的比例将其分为元训练集 D_{train} 、元验证集 D_{val} 以及元测试集 D_{test} . D_{train} 、 D_{val} 、 D_{test} 中所有图片的大小调整为 84×84 . 数据集介绍如下.

Market-1501 数据集包含 1 501 个行人、32 668 个行人矩形框, 本文随机抽取其中 200 个行人, 共 7 228 张图片, 组成 Market-mini 数据集.

DukeMTMC-reID 数据集共包含 1 812 个行人的 36 411 张图片, 本文随机抽取其中 200 个行人, 共 10 775 张图片, 组成 Duke-mini 数据集.

MSMT 17 数据集共包含 4 101 个行人的 126 441 张图片, 本文随机抽取 MSMT 17 数据集中 200 个行人, 共 13 199 张图片, 组成 MSMT 17-mini 数据集.

MSMT 17 数据集共包含 4 101 个行人的 126 441 张图片, 本文随机抽取 MSMT 17 数据集中 200 个行人, 共 13 199 张图片, 组成 MSMT 17-mini 数据集.

3.1.2 评价指标介绍

本文使用平均识别准确率 (avg-acc) 评价指标衡量在 3 个数据集上所提出方法的性能. 在元验证和元测试任务中, 对于每个待查询样本, 均会得到 C 个行人的预测值, 其中预测最大值所对应标签即为预测标签, 再与查询标签对比, 相等则记为识别正确. 由此求得单任务下识别准确率 Acc, 其计算公式如下式所示:

$$Acc = \frac{\sum_{i=1}^Q A_i}{\sum_{i=1}^Q (A_i + N_i)}. \quad (11)$$

其中: Q 为单任务中的查询样本总数, A_i 为正确识别的样本个数, N_i 为错误识别的样本个数. 为了合理规范识别准确度, 以平均精度作为最终模型指标. 在验证和测试时, 以任务数为基准, 将所有识别任务执行完成后, 求得总任务的平均识别准确率.

3.2 实验结果分析

3.2.1 对比实验分析

为了验证所提网络的有效性, 实验在 Market-mini、Duke-mini 和 MAMT 17-mini 三个数据集上与匹配 (matching) 网络^[20]、原型 (prototypical) 网络^[21]、MAML 网络^[17]、关系 (relation) 网络^[22]、CBG + Bsnet

表 1 与其他同类方法的对比实验

method	backbone	Market-mini / %		Duke-mini / %		MSMT 17-mini / %	
		5-way 1-shot	5-way 5-shot	5-way 1-shot	5-way 5-shot	5-way 1-shot	5-way 5-shot
Matching	conv 4	84.62	92.69	79.78	92.21	59.74	74.57
Prototype	conv 4	85.46	94.04	78.66	92.12	53.15	76.59
MAML	conv 4	83.83	91.74	80.21	92.57	62.33	77.61
Relation	conv 4	89.06	94.44	84.47	94.14	65.13	77.70
CBG + Bsnet	conv 4	88.97	94.85	85.07	93.10	68.51	80.11
UARenet	conv 4	89.58	95.09	84.22	94.17	68.57	82.19
FRN	conv 4	85.05	94.41	63.71	83.53	59.65	78.92
FRN + TDM	conv 4	89.72	94.79	75.95	90.48	67.85	82.14
SetFeat	conv 4	89.85	95.06	84.16	93.71	67.70	82.26
DeepBDC	Resnet 12	89.53	95.86	82.65	93.61	70.02	83.52
Prime + DDA	Resnet 50	31.62	—	21.29	—	12.94	—
MML + Circle	Resnet 50	36.19	—	31.92	—	27.55	—
ours	conv 4	90.40	95.69	86.77	94.96	71.08	82.63

网络^[27]、UARNet网络^[28]、FRN网络^[29]、FRN + TDM网络^[30]、SetFeat网络^[31]、DeepBDC网络^[32]、Prime + DDA网络^[33]和MML + Circle网络^[34] 12种相关研究方法进行对比实验,实验结果如表1所示(所提出网络的 λ 和 n 参数设置为0.8和2)。为了表明所提出方法在复杂度方面的优势,分别计算对比了conv 4、Resnet 12与Resnet 50的参数量(params)和计算量(FLOPs),结果如表2所示。

表2 不同网络模型复杂度比较

backbone	image size	params	FLOPs
Resnet 50	84×84	25.56 M	688.86 M
Resnet 12	84×84	12.42 M	3.52 G
conv 4	84×84	0.11 M	97.16 M

表1中数据表明,相较于以conv 4进行特征提取的其他方法,所提出算法在3个数据集上的5-way 1-shot和5-way 5-shot两种实验场景下均实现了最高的识别准确率,其主要原因如下:一方面利用多尺度混合注意力机制,在不同尺度层的特征提取中引入空间注意力,在不同尺度层的特征融合中引入通道注意力,提取到了更具判别力的行人特征;另一方面利用双重量融合方法,综合度量行人特征的空间绝对距离和方向差异,提升了行人相似性度量的可靠性。在两方面共同作用下,所提出方法实现了最优的识别效果。

另外,以Resnet 12为特征提取网络的DeepBDC方法,在Market-mini和MSMT 17-mini两个数据集上的5-way 5-shot实验场景下的识别性能略优于所提出

方法。其原因是,Resnet 12相比于conv 4,拥有更多的参数量和计算量,如表2所示。即DeepBDC方法在获得较高识别准确率的同时,增加了模型的时间复杂度。与其相比,所提出算法在具有较少参数量和计算量的同时,又保持较高的识别准确率,表明了所提出方法具有更高的综合性能。

基于Resnet 50的Prime + DDA和MML + Circle两种方法,在3个数据集上的识别准确率与所提出方法相比,存在明显差距。由于Resnet 50网络过于复杂,在行人数据规模有限的情况下,网络提取的原始特征过多,存在大量嘈杂特征,导致模型过拟合,使得两种方法在3个小型数据集上的识别效果不佳。表明其无法有效解决真实小样本环境下,数据量不足的行人重识别问题。

综上所述,相比于其他方法,所提出多尺度混合注意力机制与度量融合方法,在保证参数量和计算量较少的情况下,可有效地解决小样本环境下的行人重识别问题。

3.2.2 参数调优分析

由式(4)和(7)可知,实验共有两个参数,包含 λ 和 n 。其中: λ 为联合损失权重系数, n 为多尺度混合注意力模块所选不同尺度特征。

首先,针对 λ 对网络平均识别准确率的影响,在Market-mini、Duke-mini和MSMT 17-mini三个数据集上实验,由于参数 λ 与 n 相互独立, n 的取值不影响调整参数 λ 时平均识别准确率的变化趋势。实验将参数 n 的取值设置为2,实验结果如表3所示。

表3 λ 对平均识别准确率的影响

λ	Market-mini / %		Duke-mini / %		MSMT 17-mini / %	
	5-way 1-shot	5-way 5-shot	5-way 1-shot	5-way 5-shot	5-way 1-shot	5-way 5-shot
0.1	89.54	94.14	86.33	93.55	68.84	82.53
0.2	88.70	95.22	84.36	94.81	66.79	81.26
0.3	89.33	95.29	85.96	94.16	66.32	82.51
0.4	88.98	95.56	85.44	94.30	68.15	81.76
0.5	89.75	95.10	85.59	94.61	68.74	81.48
0.6	88.04	95.51	86.65	93.64	63.90	80.60
0.7	89.39	95.45	85.43	94.09	64.59	80.76
0.8	90.40	95.69	86.77	94.96	71.08	82.63
0.9	89.48	95.06	85.77	94.69	68.13	81.08

由表3可见,当 λ 取值为0.8时,所提出方法的平均识别准确率在Market-mini、Duke-mini和MSMT 17-mini三个数据集中的5-way 1-shot和5-way 5-shot两种模式下同时达到最大值,分别为90.40%和95.69%、86.77%和94.96%、71.08%和82.63%。表明联合损失loss以关系模块损失 $loss_1$ 为主,以双重量模块损失 $loss_2$ 为辅,可提升模型性能,当 λ 取值为

0.8时,网络模型性能达到最优。

固定参数 λ 值为0.8,在Market-mini、Duke-mini和MSMT 17-mini三个数据集上继续验证参数 n 对实验结果的影响,实验结果如表4所示。

由表4可见,当 n 取值为2时,在3个数据集的两种模式下,网络的平均识别准确率同时取得最大值,即特征嵌入模块采用图4中 $n=2$ 所示结构时,网络性

表4 n 对平均识别准确率的影响

n	Market-mini / %		Duke-mini / %		MSMT 17-mini / %	
	5-way 1-shot	5-way 5-shot	5-way 1-shot	5-way 5-shot	5-way 1-shot	5-way 5-shot
1	88.13	95.48	86.40	94.53	67.02	81.29
2	90.40	95.69	86.77	94.96	71.08	82.63
3	89.23	95.19	86.38	94.09	57.81	71.45
4	89.48	95.15	85.40	94.60	65.66	77.64

表5 消融实验

method	Market-mini / %		Duke-mini / %		MSMT 17-mini / %	
	5-way 1-shot	5-way 5-shot	5-way 1-shot	5-way 5-shot	5-way 1-shot	5-way 5-shot
BSNet	88.84	94.15	85.77	92.78	68.67	80.34
BSNet + MA	89.69	95.20	86.51	94.09	69.03	80.84
BSNet + DS	89.99	95.19	86.18	94.41	69.46	81.83
BSNet + MA + DS	90.40	95.69	86.77	94.96	71.08	82.63

能达到最优。

3.2.3 消融实验分析

为了验证所提出网络部件的有效性,以BSNet网络为基线,在Marke-mini、Duke-mini和MSMT 17-mini三个数据集上设计多组消融实验,验证多尺度混合注意力模块和基于度量融合的双重度量模块的有效性。其中:MA为多尺度混合注意力模块,DS为双重度量模块,实验结果如表5所示。

由表5可见,在BSNet网络基础上,加入所提出多尺度混合注意力模块后,网络的最终平均识别准确率在Market-mini、Duke-mini和MSMT 17-mini三个数据集的5-way 1-shot和5-way 5-shot两种模式下均得到了提升,分别提升了0.85%和1.05%、0.74%和1.31%、0.36%和0.50%。这表明所设计的多尺度混合注意力模块可提取更具判别性的行人特征,而当采用所提出双重度量模块训练网络时,最终平均识别准确率在3个数据集的两种模式下也均得到了提升,分别提升了1.15%和1.04%、0.41%和1.63%、0.79%和1.49%。表明采用双重度量模块后,使得少量样本间的度量更加可靠,从而提升识别准确率。最后,当多

尺度混合注意力模块和双重度量模块共同加入网络时,平均识别准确率在3个数据集的两种模式下均取得较大提升,分别提升了1.56%和1.54%、1.00%和2.18%、2.41%和2.29%。表明同时采用所提出的两种方法,提升了模型的鲁棒性和有效性,使得模型性能达到最佳效果。

3.2.4 可视化结果分析

为了直观地表现所提出网络的效果,在本节利用Gard-Cam可视化方法生成热力图,与BSNet网络、匹配(matching)网络、原型(prototypical)网络、MAML网络、CBG + Bsnet网络、UARenet网络、关系(relation)网络、FRN网络、FRN + TDM网络、SetFeat网络、DeepBDC网络、Prime + DDA网络和MML + Circle网络进行效果对比。热力图展示了网络特征嵌入模块输出特征,其中颜色越深,表明网络对该区域越关注。

如图6所示,分析可得无论在数据集Market、Duke还是MSMT 17的图像上,所提出网络在特征图中会捕捉更有效的行人信息,表明了该网络对解决样本量不足的行人重识别问题的有效性。

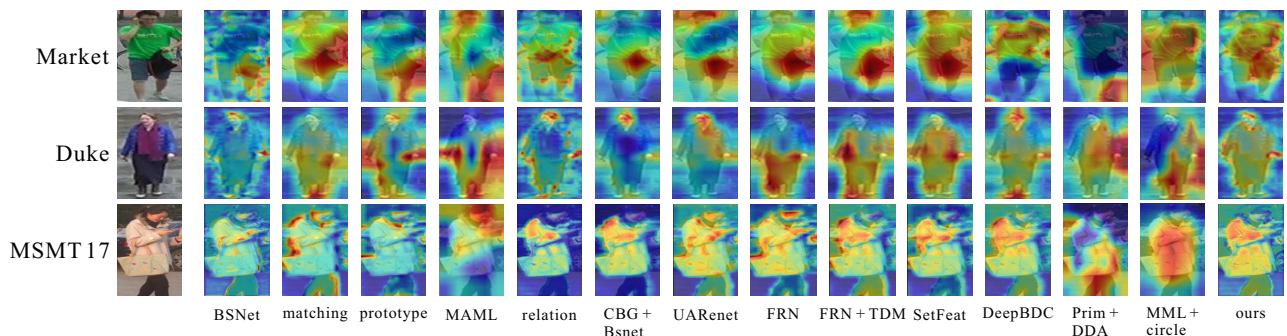


图6 实验效果

4 结论

针对行人重识别中行人数据量不足问题,本文提出了一种基于多尺度混合注意力与度量融合的小样本行人重识别方法.在特征嵌入模块中引入了所提出多尺度混合注意力机制,缓解了因卷积层数不足导致特征含有较多噪声问题,所提取的小样本行人数据更具判别性的特征表示.此外,所提出双重度量模块,综合考虑了行人特征的方向差异和绝对距离,生成比余弦度量更为可靠的度量得分.基于Market-mini、Duke-mini以及MSMT 17-mini三个数据集进行验证,实验结果表明,该网络能够有效克服行人数据量不足问题,实现较高的小样本行人重识别准确率.在未来的工作中,将以所提出算法为基础,继续深入开展行人图像不足情况下的跨域重识别问题研究,以期为解决真实开放环境的行人重识别问题提供解决思路.

参考文献(References)

- [1] Chen Y C, Zhu X T, Zheng W S, et al. Person re-identification by camera correlation aware feature augmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 40(2): 392-408.
- [2] Zheng L, Shen L Y, Tian L, et al. Scalable person re-identification: A benchmark[C]. *IEEE International Conference on Computer Vision*. Santiago, 2016: 1116-1124.
- [3] Wei L H, Zhang S L, Gao W, et al. Person transfer GAN to bridge domain gap for person re-identification[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, 2018: 79-88.
- [4] 王松, 纪鹏, 张云洲, 等. 自适应感受野网络的行人重识别[J]. *控制与决策*, 2022, 37(1): 119-126.
(Wang S, Ji P, Zhang Y Z, et al. Adaptive receptive network for person re-identification[J]. *Control and Decision*, 2022, 37(1): 119-126.)
- [5] 周勇, 王瀚正, 赵佳琦, 等. 基于可解释注意力部件模型的行人重识别方法[J]. *自动化学报*: 2020, 41(x): 1-13.
(Zhou Y, Wang H Z, Zhao J Q, et al. Interpretable attention part model for person re-identification[J]. *Acta Automatica Sinica*, 2020, 41(x): 1-13.)
- [6] 刘玉杰, 周彩云, 李宗民, 等. 基于增强特征融合网络的行人重识别方法[J]. *计算机辅助设计与图形学学报*, 2021, 33(2): 232-240.
(Liu Y J, Zhou C Y, Li Z M, et al. Strong feature fusion networks for person re-identification[J]. *Journal of Computer-Aided Design & Computer Graphics*, 2021, 33(2): 232-240.)
- [7] Wu D, Zheng S J, Zhang X P, et al. Deep learning-based methods for person re-identification: A comprehensive review[J]. *Neurocomputing*, 2019, 337: 354-371.
- [8] 卢健, 王航英, 陈旭, 等. 基于多尺度特征表示的行人再识别[J]. *控制与决策*, 2021, 36(12): 3015-3022.
(Lu J, Wang H Y, Chen X, et al. Multi-scale feature representation for person re-identification[J]. *Control and Decision*, 2021, 36(12): 3015-3022.)
- [9] Li X Y, Long S P, Zhu J. Survey of few-shot learning based on deep neural network[J]. *Application Research of Computers*, 2020, 37(8): 2241-2247.
- [10] Wang Y X, Girshick R, Hebert M, et al. Low-shot learning from imaginary data[C]. *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, 2018: 7278-7286.
- [11] 傅桂霞, 邹国锋, 毛帅, 等. 融合Gabor特征与卷积特征的小样本行人重识别[J]. *山东大学学报: 工学版*, 2021, 51(3): 22-29.
(Fu G X, Zou G F, Mao S, et al. Small sample person re-identification combining Gabor features and convolution features[J]. *Journal of Shandong University: Engineering Science*, 2021, 51(3): 22-29.)
- [12] 刘月, 邹国锋, 陈贵震, 等. 基于分布约束的非对称度量学习无监督行人重识别[J]. *控制与决策*, 2023, 38(6): 1703-1711.
(Liu Y, Zou G F, Chen G Z, et al. Asymmetric metric learning approach based on distribution constraints for unsupervised person re-identification[J]. *Control and Decision*, 2023, 38(6): 1703-1711.)
- [13] Wu Y, Lin Y T, Dong X Y, et al. Progressive learning for person re-identification with one example[J]. *IEEE Transactions on Image Processing*, 2019, 28(6): 2872-2881.
- [14] Xu T, Li J L, Wu H, et al. Feature space regularization for person re-identification with one sample[C]. *IEEE the 31st International Conference on Tools with Artificial Intelligence*. Portland, 2020: 1463-1470.
- [15] 费晔焯, 徐森, 王科俊. 行人步态的特征表达及识别综述[J]. *模式识别与人工智能*, 2012, 25(1): 71-81.
(Ben X Y, Xu S, Wang K J. Review on pedestrian gait feature expression and recognition[J]. *Pattern Recognition and Artificial Intelligence*, 2012, 25(1): 71-81.)
- [16] Ge Y, Chen D, Li H. Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification[J/OL]. 2020, arXiv: 2001.01526.
- [17] Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks[J/OL]. 2017, arXiv: 1703.03400.
- [18] Ravi S, Larochelle H. Optimization as a model for few-shot learning[C]. *International Conference on*

- Learning Representations. San Juan, 2016: 1-11.
- [19] Gidaris S, Komodakis N. Dynamic few-shot visual learning without forgetting[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 4367-4375.
- [20] Vinyals O, Blundell C, Lillicrap T, et al. Matching networks for one shot learning[J]. Advances in Neural Information Processing Systems, 2016, 29(1): 3637-3645.
- [21] Snell J, Swersky K, Zemel R S. Prototypical networks for few-shot learning[C]. Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, 2017: 4080-4090.
- [22] Sung F, Yang Y, Zhang L, et al. Learning to compare: Relation network for few-shot learning[J/OL]. 2017, arXiv: 1711.06025.
- [23] Li X X, Wu J J, Sun Z, et al. BSNet: Bi-Similarity network for few-shot fine-grained image classification[J]. IEEE Transactions on Image Processing, 2020, 30: 1318-1331.
- [24] Wang Z, Bovik A C. Mean squared error: Love it or leave it? A new look at signal fidelity measures[J]. IEEE Signal Processing Magazine, 2009, 26(1): 98-117.
- [25] Vanschoren J. Meta-learning: A survey[J/OL]. 2018, arXiv: 1810.03548.
- [26] 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 Recognition. Las Vegas, 2016: 770-778.
- [27] 陆妍, 王阳萍, 王文润. 基于 Transformer 的小样本细粒度图像分类方法[J]. 计算机工程与应用, <https://kns.cnki.net/KCMS/detail/detail.aspx?filename=JSGG20221025003&dbname=CJFD&dbcode=CJFQ>. (Lu Y, Wang Y P, Wang W R. Transformer-based few-shot and fine-grained image classification method[J]. Computer Engineering and Applications, <https://kns.cnki.net/KCMS/detail/detail.aspx?filename=JSGG20221025003&dbname=CJFD&dbcode=CJFQ>.)
- [28] 孟浩, 田洋, 孙宇婷, 等. 全局注意力关系网络的小样本船舶识别[J]. 仪器仪表学报, 2021, 42(12): 220-227. (Meng H, Tian Y, Sun Y T, et al. Few shot ship recognition based on universal attention relationnet[J]. Chinese Journal of Scientific Instrument, 2021, 42(12): 220-227.)
- [29] Wertheimer D, Tang L M, Hariharan B. Few-shot classification with feature map reconstruction networks[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, 2021: 8008-8017.
- [30] Lee S B, Moon W J, Heo J P. Task discrepancy maximization for fine-grained few-shot classification[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, 2022: 5321-5330.
- [31] Afrasiyabi A, Larochelle H, Lalonde J F, et al. Matching feature sets for few-shot image classification[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, 2022: 9004-9014.
- [32] Xie J T, Long F, Lv J M, et al. Joint distribution matters: Deep Brownian distance covariance for few-shot classification[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, 2022: 7962-7971.
- [33] 张晓艳, 张宝华, 吕晓琪, 等. 深度双重注意力的生成与判别联合学习的行人重识别[J]. 光电工程, 2021, 48(5): 57-65. (Zhang X Y, Zhang B H, Lv X Q, et al. The joint discriminative and generative learning for person re-identification of deep dual attention[J]. Opto-Electronic Engineering, 2021, 48(5): 57-65.)
- [34] 常禾雨, 司念文, 屈丹, 等. 元学习框架下基于圆损失度量的行人重识别模型[J]. 信息工程大学学报, 2022, 23(1): 93-102. (Chang H Y, Si N W, Qu D, et al. Person re-identification model based on circle loss metric under the framework of meta learning[J]. Journal of Information Engineering University, 2022, 23(1): 93-102.)

作者简介

陈贵震(1997—), 男, 硕士生, 从事行人重识别的研究, E-mail: chen17853309228@163.com;

邹国锋(1984—), 男, 副教授, 博士, 从事行人重识别、人脸识别、视频监控技术等研究, E-mail: zgf841122@163.com;

刘月(1998—), 女, 硕士生, 从事行人重识别的研究, E-mail: liuyue_sdut@163.com;

傅桂霞(1985—), 女, 讲师, 博士, 从事行人重识别、智能信息处理技术等研究, E-mail: fgx45101@163.com;

高明亮(1985—), 男, 副教授, 博士, 从事运动目标检测与跟踪、视频分析技术等研究, E-mail: sdut_mlgao@163.com.