

# 控制与决策

Control and Decision

## 融合注意力机制的域泛化行人再识别

于明, 李学博, 郭迎春

引用本文:

于明, 李学博, 郭迎春. 融合注意力机制的域泛化行人再识别[J]. *控制与决策*, 2022, 37(7): 1721–1728.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2020.1844>

---

## 您可能感兴趣的其他文章

Articles you may be interested in

### [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>

### [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>

### [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>

### [结合注意力机制的循环神经网络复述识别模型](#)

Recurrent neural networks based paraphrase identification model combined with attention mechanism

*控制与决策*. 2021, 36(1): 152–158 <https://doi.org/10.13195/j.kzyjc.2019.0638>

### [一种基于深度学习的时间序列预测方法](#)

A time series prediction method based on deep learning

*控制与决策*. 2021, 36(3): 645–652 <https://doi.org/10.13195/j.kzyjc.2019.0809>

# 融合注意力机制的域泛化行人再识别

于明, 李学博, 郭迎春<sup>†</sup>

(河北工业大学 人工智能与数据科学学院, 天津 300400)

**摘要:** 域泛化的行人再识别能够在源数据集进行训练并在目标数据集进行测试, 具有更广泛的实际应用意义. 现有域泛化模型往往由于专注解决光照和色彩变化问题而忽略对细节信息的有效利用, 导致识别率较低. 为了解决上述问题, 提出一种融合注意力机制的域泛化行人再识别模型. 该模型首先通过叠加卷积层的瓶颈层 (bottleneck layer) 设计提取出包含不同视野域的多尺度特征, 并利用特征融合注意力模块对多尺度特征进行加权动态融合; 然后通过多层次注意力模块挖掘细节特征的语义信息; 最后将包含丰富语义信息的细节特征输入到判别器进行行人再识别. 此外, 设计风格正则化模块 (style normalization module) 用于降低不同数据集图像明暗变化对模型泛化能力的影响. 在 Market-1501 和 DukeMTMC-reID 数据集上进行对比和消融实验, 表明了所提出方法的有效性.

**关键词:** 注意力机制; 多尺度; 行人再识别; 风格正则化; 域泛化; 瓶颈层

中图分类号: TP183

文献标志码: A

DOI: 10.13195/j.kzyjc.2020.1844

引用格式: 于明, 李学博, 郭迎春. 融合注意力机制的域泛化行人再识别 [J]. 控制与决策, 2022, 37(7): 1721-1728.

## Domain generalization person re-identification based on attention mechanism

YU Ming, LI Xue-bo, GUO Ying-chun<sup>†</sup>

(School of Artificial Intelligence, Hebei University of Technology, Tianjin 300400, China)

**Abstract:** Domain generalization person re-identification models can be trained in the source dataset and tested in the target dataset, which has wide practical application significance. The existing domain generalization models tend to focus on solving the problem of illumination and color change, and ignoring the effective utilization of detailed information, which leads to the low recognition rate. In this paper, a domain generalized person re-identification model is proposed, which is based on the attention mechanism. Firstly, the model extracts multi-scale features with different visual fields through the design of the bottleneck layer superimposed on the convolutional layer, and a feature fusion attention module is used to perform dynamic fusion of multi-scale features and assign weights. Then the semantic information of refinement features is mined by the multilevel attention module. Finally, the feature containing rich semantic information is input to the discriminator for person re-identification. In addition, a style regularization module is designed to reduce the influence of image light and shade changes on model generalization ability. Abundant comparison and ablation experiments conducted on market-1501 and DukeMTMC-reID datasets demonstrate the effectiveness of the proposed method.

**Keywords:** attention mechanism; multi-scale; person re-identification; style normalization; domain generalization; bottleneck layer

## 0 引言

行人再识别旨在匹配出现在不同场景中的同一个人, 因其在刑侦、智慧城市等领域拥有广阔的应用前景, 吸引了大批研究者. 然而, 行人识别面临着行人姿态、光照条件差别大以及存在遮挡等严峻的挑战, 是一个很有挑战性的课题<sup>[1]</sup>.

自深度卷积网络出现并在多个领域证明其有效后, 越来越多的研究者使用深度卷积网络解决行人再识别问题. 一些研究人员主要聚焦于有监督的行人再识别, 从特征提取角度入手<sup>[2-3]</sup>, 通过学习更有效的特征表达使得模型学习到更有判别力的特征表达, 以此提升模型性能; 另一些方法从度量学习角度入

收稿日期: 2020-12-31; 录用日期: 2021-04-07.

基金项目: 国家自然科学基金项目 (61806071); 河北省自然科学基金面上项目 (F2019202381, F2019202464).

责任编辑: 张国山.

<sup>†</sup>通讯作者. E-mail: guoyingchun@hebut.edu.cn.

手<sup>[4-5]</sup>,通过提出更有效的度量方法监督模型学习到更合理的参数,更准确地找到不同场景下的同一个目标行人.另有一些研究者致力于无监督的行人再识别<sup>[6]</sup>,不使用任何源域的标签信息,通过对源域图像进行聚类生成伪标签,以此指导模型的训练.为了更贴近实际应用条件,还有一些研究者提出了域自适应方法<sup>[7-9]</sup>,将一个模型在源数据集上做预训练,使用目标域的图像对预训练好的模型进行微调.比域适应更具实际意义的是域泛化方法,强调将源数据集上预训练好的模型直接应用于目标数据集,完全不需要任何目标域的图像或者标注<sup>[10]</sup>.

现有的域泛化模型往往由于专注解决光照和色彩变化问题而忽略对细节信息的有效利用.行人再识别的数据集里存在广泛的相似行人,他们拥有相似的衣着和姿态,仅从解决光照及色彩变化问题入手不能有效提升模型针对这些图像的判别能力.然而,这些图像细节信息并不相同,通过这些细节信息可以轻易辨识出这些行人的类别.如果模型可以学习到这些细节特征,获得对细节特征的捕捉能力,则能够提升模型的泛化能力.

本文聚焦于域泛化方法,着重解决现有域泛化行人再识别方法中存在的忽略对细节信息的有效利用问题,提出融合注意力机制的域泛化行人再识别模型.利用卷积层叠加的方式构建多尺度特征提取模块,并设计特征融合注意力模块对提取出的多尺度特征进行动态融合.提出多层次注意力模块,用于提取特征的多层次信息,进一步挖掘行人图像中不同部分更复杂的细微差别,提取到更强力的行人图像特征表达.同时设计风格正则化模块以降低不同数据集图像明暗变化对模型泛化能力的影响.在Market-1501和DukeMTMC-reID数据集上进行对比和消融实验,表明了所提出方法的有效性.

## 1 相关工作

研究人员最初聚焦于有监督的行人再识别,随着越来越多优秀算法的提出,有监督的行人再识别取得了辉煌的成就,然而其需要大量的额外标注,这些费时费力的工作需求始终制约着行人再识别的进一步发展与应用<sup>[11-13]</sup>.因此,无监督行人再识别成为研究的热点,根据实现方式不同可分为基于聚类的无监督行人再识别方法和跨域行人再识别方法,后者又可细分为域自适应方法和域泛化方法.

基于聚类的无监督行人再识别不使用任何目标域的标签信息,利用源域的有标签图像对模型进行预训练后,对目标域中无标签的图像进行聚类生成伪标

签,以此指导模型训练.该算法虽然不需要目标域的标签信息,但是仍然需要类别总数信息指导聚类簇参数的设定,并没有完全摆脱对目标域的依赖<sup>[14]</sup>.

域自适应的跨域行人再识别是将一个模型在源数据集上做预训练,然后使用目标域的图像对预训练好的模型进行微调,使得模型更适应目标域的图像风格,微调过程中只使用图像,不使用行人标签<sup>[15-17]</sup>.域自适应方法的缺点是仍然没有摆脱对目标域图像信息的需求<sup>[18-20]</sup>.

域泛化的跨域行人再识别完全不需要任何目标域的图像或者标注,强调将在源数据集上预训练好的模型直接应用于目标数据集,但是不经过不同域之间微调的操作,如何保证模型的学习效果并增强模型的泛化能力成为需要解决的难点<sup>[21-23]</sup>.Liu等<sup>[18]</sup>专注于利用有判别力的信息提升模型的泛化和适应能力,基于ResNet-50提出了两种注意力机制,一种基于长域依赖的注意力,另一种基于直接生成的注意力,通过简单整合注意力的方式发现注意力的恰当使用能够显著增强高层和中层的语义特征,从而增加模型的泛化能力.Jin等<sup>[24]</sup>认为不同行人图像在照明、色调以及色彩对比度等方面会呈现出很大的风格差异,提出了一个可以实现即插即用的风格归一与恢复模块SNR(style normalization and restitution module),SNR分为风格归一化部分和风格恢复部分,前者利用IN层做归一化减少特征在不同域之间的差异,后者负责从被删除的信息中提取出与身份相关的特征,并将其恢复到网络中.Wang等<sup>[22]</sup>认为已有模型在不同场景下的泛化能力依然较差的一个原因是目前仍缺乏大规模和多样性的训练数据,由此提出了一种随机自动产生大量三维人物模型的方法,并在游戏环境里模拟真实监控进行渲染,得到一个大规模的虚拟行人数据集RandPerson.Chen等<sup>[25]</sup>认为现有很多方法选择利用多个源数据集对模型进行训练,但没有考虑不同数据集之间某些图像的相似性,由此提出了一种双分布对齐网络,通过有选择地对齐多个源域的图像分布将图像映射到一个跨域不变的特征空间,这种具有双重约束的对齐网络实现了一种新颖的域对抗性特征学习和身份相似性增强,有效提高了跨域行人再识别模型的泛化能力.

域泛化方法因为其更广阔的发展前景和应用空间,吸引了越来越多的研究人员的兴趣.然而,现有域泛化模型往往由于专注解决光照和色彩变化问题而忽略对细节信息的有效利用,鉴于此,本文提出融合注意力机制的域泛化行人再识别模型,通过对图像细

节信息的提取利用,进一步提升域泛化行人再识别模型的性能.

## 2 模型提出

模型整体流程如图1所示.由图1可见,模型主要由卷积层和瓶颈层组成,行人图像首先经过卷积层和最大池化层获得图像特征向量,通过两组瓶颈层和

风格正则化模块的组合,提取到图像的多尺度特征并融合,降低特征的风格特殊性;然后依次通过两组卷积层,平均池化层和瓶颈层的组合,细化图像的细节特征;最后经过卷积层、平均池化层,全连接层,将Softmax函数预测的分类结果作为模型的输出.每一个瓶颈层中包含多尺度特征提取、特征融合注意力模块和多层次注意力模块.

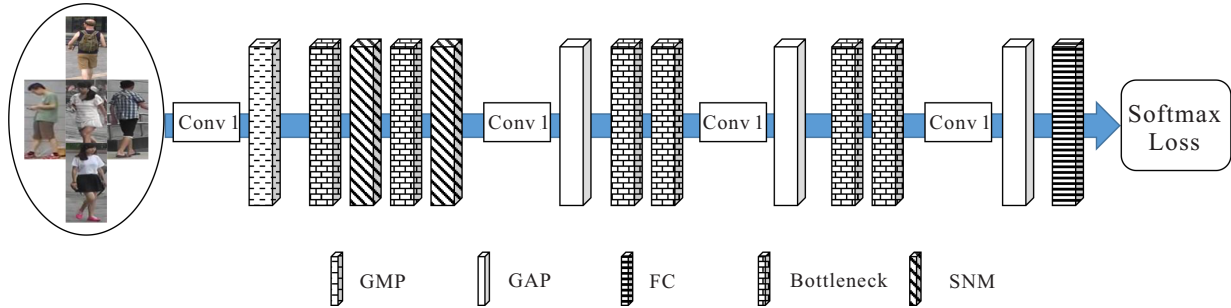


图1 整体流程

### 2.1 问题定义及模型流程

行人再识别旨在检索出跨摄像头条件设定下的同一个行人.

给定一个源域  $S = \{X_s, L_s\}$ , 包含  $P$  个人的  $N_s$  张图像. 其中:  $X_s$  为行人图像,  $L_s$  为图像的标签信息,  $x_s^i$  和  $l_s^i$  分别为源域中第  $i$  幅图像以及该图像的标签信息, 每个  $x_s^i$  对应一个  $l_s^i$ .

给定一个目标域  $T = \{X_t\}$ , 包含与源域  $S$  无重

叠的  $N_t$  张图像. 域泛化行人再识别方法是使用源域  $S = \{X_s, L_s\}$  对初始化参数为  $\theta$  的模型进行训练, 然后将训练完成的参数为  $\theta_t$  的模型应用在  $T = \{X_t\}$  上进行测试.

模型设计使用了瓶颈层, 其具体细节流程如图2所示. 瓶颈层首先通过叠加卷积层的设计提取出包含不同视野域的多尺度特征, 并利用特征融合注意力模块对多尺度特征进行加权动态融合; 然后通过多层次注意力模块挖掘细节特征的语义信息.

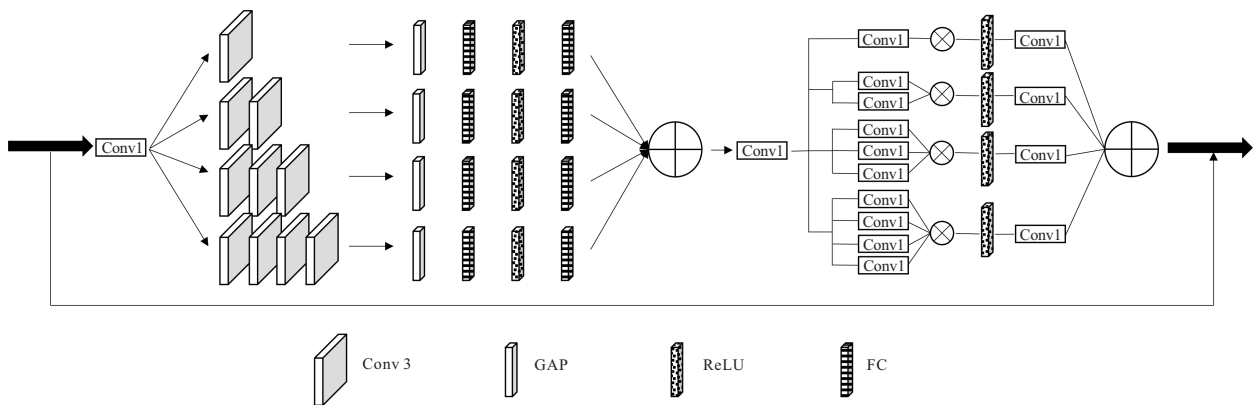


图2 瓶颈层细节

瓶颈层通过卷积层叠加提取出包含不同视野域大小的多尺度特征, 将瓶颈层的输出作为残差与输入相加, 并作为下一层的输入, 有

$$\begin{aligned} \hat{f} &= f + \bar{f}, \\ \text{s.t. } \bar{f} &= \theta(f). \end{aligned} \quad (1)$$

其中:  $f$  为瓶颈层的输入;  $\theta(\cdot)$  为瓶颈层操作;  $\hat{f}$  为瓶颈层输出与输入之和, 作为下一阶段的输入.

### 2.2 多尺度特征提取及融合

域泛化行人再识别模型由有判别力的特征表示和有效的度量学习保障性能<sup>[26-27]</sup>, 不仅要解决光照及色彩变化问题, 而且应对细节信息加以利用. 本文通过提取出包含更多细节信息的特征提升模型的泛化能力, 这便需要提取出不同视野域大小的特征<sup>[28-29]</sup>. 行人图像包含多种多样的特征信息, 不同视

视野域大小的多尺度特征不仅可以提取出图像中的全局特征,也可以提取出局部特征,使模型对图像整体信息进行学习的同时,也能够学习到图像的细节信息,这种对细节信息的提取能力可以有效增加模型的泛化能力。

叠加卷积层可以改变视野域的大小,提取出包含不同视野域大小信息的特征<sup>[30]</sup>。为了减少参数的数量采用轻量级的卷积块<sup>[31-32]</sup>,通过其叠加提取不同视野域大小的多尺度特征。对于第 $m$ 个提取的多尺度特征,采用 $m$ 个轻量级 $3 \times 3$ 卷积块叠加,提取视野域为 $(2m+1) \times (2m+1)$ 大小的特征信息,不同视野域大小的特征融合使用,如下所示:

$$\begin{aligned} \tilde{f} &= \sum_{m=1}^M F_m(f); \\ \text{s.t. } M &\geq 1. \end{aligned} \quad (2)$$

其中: $\tilde{f}$ 为融合后的特征, $M$ 为不同视野域大小的多尺度特征总数, $f$ 为多尺度特征提取的输入, $F_m(\cdot)$ 为对每一尺度特征进行的卷积操作。

为了有效利用不同视野域大小的多尺度特征,利用注意力机制代替多尺度特征直接相加的融合方式。注意力机制可以动态赋予不同特征不同的权重,使得重要的特征权重增大,不重要的特征权重减小<sup>[33-35]</sup>。由此提出了特征融合注意力模块对多尺度特征进行动态融合。特征融合注意力模块由全局平均池化层、全连接层、ReLU激活层和另一个全连接层依次组成,使用残差连接的方式使当前瓶颈层学习到的多尺度特征能够在下一层网络中保留,避免网络深度增加带来的网络退化隐患。每一尺度的特征先经过融合注意力模块,然后将输出与输入进行Hadamard乘积作为这一尺度的输出,不同尺度的特征经过融合注意力模块加权后相加,作为下一阶段的输入,如下所示:

$$\begin{aligned} \tilde{f} &= \sum_{m=1}^M A(f_m) \otimes f_m; \\ \text{s.t. } f_m &= F_m(f). \end{aligned} \quad (3)$$

其中: $A(\cdot)$ 为特征融合注意力模块进行的操作, $\otimes$ 代表Hadamard乘积。为了减少参数,使参数量与 $M$ 无关,对不同尺度特征的融合注意力模块进行了参数共享。多尺度特征提取及融合流程如图3所示。

### 2.3 多层次注意力模块

在第2.2节中本文提取出了多尺度特征,并利用融合注意力模块对不同尺度的特征进行了动态融合。多尺度特征包含了不同视野域大小的信息,具有

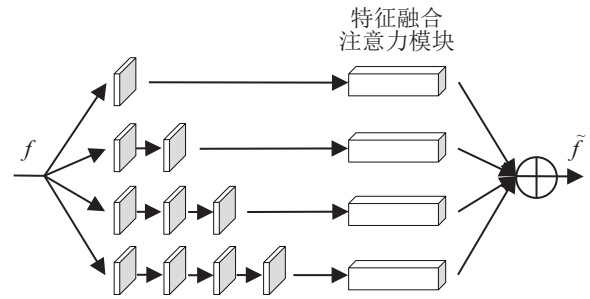


图3 多尺度特征提取及融合

一定的辨别力。然而,域泛化行人再识别问题面临着视角、姿态差异等巨大挑战,单一层次的特征不能很好地捕捉行人图像之间的细微差别,很难提供足够丰富的辨别信息。而图像的多层次信息可以挖掘行人图像不同部分的高维联系,增强特征的丰富性和判别力。

为了充分利用特征的层次信息,进一步挖掘行人图像中不同部分更复杂的细微差别,提取到更强力的行人图像特征表达,设计了多层次注意力模块提取特征的多层次信息。多层次注意力模块由卷积层和ReLU层组成,层数设置为 $K$ ,第 $k$ 层提取部分由 $k+1$ 个卷积层和1个ReLU层组成。首先将输入通过 $1 \times 1$ 卷积降低维度,然后对于第 $k$ 层提取部分,特征将分别经过 $k$ 个 $1 \times 1$ 卷积后进行Hadamard乘积,依次通过一个ReLU层和一个 $1 \times 1$ 卷积层,如图4所示。

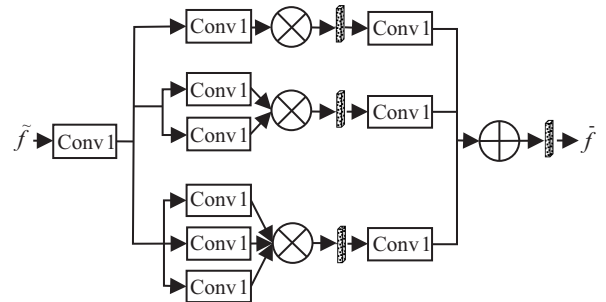


图4 多层次注意力模块

将 $K$ 个层次的特征信息相加作为该模块的输出,有

$$\tilde{f} = \sum_{k=1}^K F_k(\tilde{f}). \quad (4)$$

其中: $F_k(\cdot)$ 为 $k$ 次信息提取的操作, $K$ 为层次总数。

### 2.4 风格正则化模块

在跨域行人再识别方向,尤其在域泛化问题上,因为模型没有任何目标数据集的信息可以学习,所以泛化能力受到了更大的考验。减少从源数据集上学习的特征的风格特殊性,可以有效提升模型泛化能力<sup>[24]</sup>,本文设计了风格正则化模块以降低特征的风格特殊性。

风格正则化模块由一个IN层和一个ReLU层简单构成,首先由IN层计算归一化统计量,然后由ReLU函数增加特征的非线性.不同于BN层跨样本在单通道上计算归一化统计量,也不同于LN层在单样本上跨通道计算归一化统计量,IN层在单通道单样本数据上计算归一化统计量,所以不受通道数和batchsize的影响,过程如下:

$$f_{sn} = \text{ReLU}(\text{IN}(\tilde{f})) = \text{ReLU}\left(\gamma \frac{\hat{f} - \mu(\hat{f})}{\sigma(\hat{f})} + \beta\right). \quad (5)$$

其中: $\hat{f}$ 为风格正则化模块的输入, $\mu(\cdot)$ 为均值, $\sigma(\cdot)$ 为标准差, $\gamma$ 和 $\beta$ 为模型训练过程中学习到的参数.

通过Softmax函数计算输入图像属于每一类别的概率,对应概率值最大的类别即为模型最终预测的结果,有

$$R_j = \underset{c}{\text{argmax}} \left( \frac{\exp(W_j \cdot F)}{\sum_{c=1}^C \exp(W_c \cdot F)} \right). \quad (6)$$

其中: $F$ 为模型最终提取到的行人特征, $W$ 为用作分类的全连接层的权重, $C$ 为目标域 $T$ 中行人类别总数.将预测结果输入到交叉熵损失函数中,计算预测标签与该行人真实标签之间的距离,利用随机梯度下降法反向传播更新模型各层网络的参数,损失函数如下所示:

$$\text{Loss} = - \sum_j^C L_j \log(R_j), \quad (7)$$

其中 $L_j$ 为第 $j$ 个行人的真实标签.

### 3 实验分析

#### 3.1 数据集介绍

**Market-1501:** Market-1501数据集在清华大学校园中夏季拍摄采集,并在2015年构建和公开.它包括由5个高清摄像头和1个低清摄像头拍摄到的1501个行人、32668个检测到的行人矩形框.数据集里每个行人至少由2个摄像头捕获到,并且在一个摄像头中可能具有多张图像.其中训练集有751人,包含12936张图像,平均每个人有17.2张训练数据;测试集有750人,包含19732张图像,平均每个人有26.3张测试数据.该数据集提供的固定数量的训练集和测试集均可在single-shot或multi-shot测试设置下使用.

**DukeMTMC-reid:** DukeMTMC-reid是行人跟踪数据集DukeMTMC的行人再识别子集,并且提供了人工标注的bounding box. DukeMTMC数据集是一个大规模标记的多目标多摄像机行人跟踪数据集. DukeMTMC-reid从DukeMTMC视频中每120帧

采样一张图像,得到36411张图像.共有1404个人出现在大于两个摄像头下,有408个人只出现在一个摄像头下.其中bounding\_box\_test下是用于测试集的702人,包含随机采样的17661张图像; bounding\_box\_train下是用于训练集的702人,包含16522张图像; query为测试集中的702人在每个摄像头中随机选择一张图像,共有2228张图像;剩下的图像加入gallery,并将之前的408人作为干扰项,也加到gallery中.

#### 3.2 实验设置

ResNet-50在越来越多的分类问题中证明了其先进性,所以本文采用在ImageNet上预训练的ResNet-50作为模型的Baseline,并使用一个线性全连接层和Softmax函数作为分类器,利用随机梯度下降算法监督模型的训练直到收敛.选择带有标签平滑的交叉熵损失函数作为全模型的损失,在实验过程中,将所有数据集中的图片设置为 $256 \times 128$ ,初始学习率为0.0015并训练150个epochs,学习率每60个epochs衰减0.1.在前10个epochs中,固定预训练网络的参数并只训练分类器,然后开始对模型和分类器的联合训练.为了提高模型的性能,使用一些常见的数据扩充方法,包括随机裁剪和随机擦除.

#### 3.3 实验结果及消融实验

采用两个行人再识别领域常用的主流评价指标来衡量模型的性能,一个是Rank- $n$ 准确率,另一个是mAP. Rank- $n$ 准确率代表搜索结果中最靠前的 $n$ 张图里有正确结果的概率,一般通过多次实验取平均值. Map(mean average precision)用于衡量算法的搜索能力,是通过精准率和召回率绘制出来的P-R曲线下的面积. AP是某一类P-R曲线下的面积, mAP则是所有类别P-R曲线下面积的平均值.

在Market-1501和DukeMTMC-reid两个常用行人数据集上进行实验,为了验证各个模块的有效性,进行消融实验,结果如表1和表2所示.其中MFEM(multi-scale feature extract module)为多尺度特征提取及融合模块,SNM(style normalization module)为风格正则化模块,Attention为多层次注意力模块.通过消融实验结果可以看到,当多尺度特征提取及融合模块、多层次注意力模块和风格正则化模块共同使用时,模型取得了最佳性能,验证了各模块的有效性.

本文模型有两个重要的超参数,一个是多尺度特征的不同尺度总数 $M$ ;另一个是多层次注意力模块的层次数 $K$ .对于 $M$ ,设置过大会导致参数量过多,设置过小则不能提取到丰富的多尺度特征,经过实验

验证,当  $M$  取值4时模型性能最佳,如图5所示,故本文设置  $M$  为4. 对于  $K$ , 设置过大不仅会导致太高层信息失去实际意义,也会导致参数过多使得模型难以训练,设置过小则不能提取出有效的多层次信息,经过实验验证,随着  $K$  取值增大,模型识别率提升效果逐渐放缓后下降,如图6所示,且随着  $K$  取值增大,模型训练时间显著增长,综合考虑本文设置  $K$  为4.

表1 Duke到Market1501消融实验

模块	Duke → Market1501				
	R1	R5	R10	R20	mAP
MFEM	54.8	72.4	78.2	83.3	25.4
MFEM + SNM	56.2	72.7	78.9	84.4	25.3
MFEM + Attention	55	71.1	77.4	82.4	25.2
MFEM + SNM + Attention	59.4	76.8	82.5	86.9	28.2

表2 Market1501到Duke消融实验

模块	Market1501 → Market1501				
	R1	R5	R10	R20	mAP
MFEM	46.1	62.1	67.6	73.0	26.1
MFEM + SNM	46.6	62.1	67.5	72.5	26.2
MFEM + Attention	46.7	62.5	68.9	73.3	27.0
MFEM + SNM + Attention	49.9	64.6	69.5	74.2	28.4

表3 与主流方法对比

方法	出处	Duke					Market 1501				
		source	R1	R5	R10	mAP	source	R1	R5	R10	mAP
MMFA <sup>[15]</sup>	arXiv'18	M+D(U)	45.3	59.8	66.3	24.7	D+M(U)	56.7	75.0	81.8	27.4
SPGAN <sup>[8]</sup>	CVPR'18	M+D(U)	46.4	62.3	68.0	26.2	D+M(U)	57.7	75.8	82.4	26.7
TJ-AIDL <sup>[16]</sup>	CVPR'18	M+D(U)	44.3	59.6	65.0	23.0	D+M(U)	58.2	74.8	81.1	26.5
ATNet <sup>[17]</sup>	CVPR'19	M+D(U)	45.1	59.5	64.2	24.9	D+M(U)	55.7	73.2	79.4	25.6
CamStyle <sup>[36]</sup>	TIP'19	M+D(U)	48.4	62.5	68.9	25.1	D+M(U)	58.8	78.2	84.3	27.4
HHL <sup>[37]</sup>	ECCV'18	M+D(U)	46.9	61.0	66.7	27.2	D+M(U)	62.2	78.8	84.0	31.4
ECN <sup>[38]</sup>	CVPR'19	M+D(U)	63.3	75.8	80.4	40.4	D+M(U)	75.1	87.6	91.6	43.0
Jia等 <sup>[21]</sup>	arXiv'19	M	39.2	—	—	21.5	D	52.5	—	—	24.6
OSNet <sup>[30]</sup>	ICCV'19	M	48.5	62.3	67.4	26.7	D	57.7	73.7	80.0	26.1
Wang等 <sup>[22]</sup>	ACMMM'20	M	34.6	—	—	18.1	D	44.9	—	—	19.0
Luo等 <sup>[23]</sup>	ECCV'20	M	30.0	44.9	50.9	15.0	D	48.9	65.1	71.8	19.8
本文算法	Proposed	M	49.9	64.6	69.5	28.4	D	59.4	76.8	82.5	28.2

注: M为Market-1501, D为DukeMTMC-reid, U为只使用行人图像.

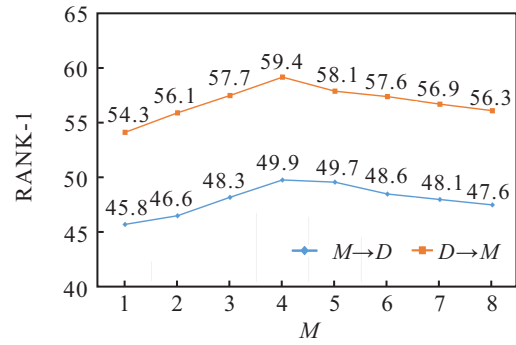


图5 参数M的选取

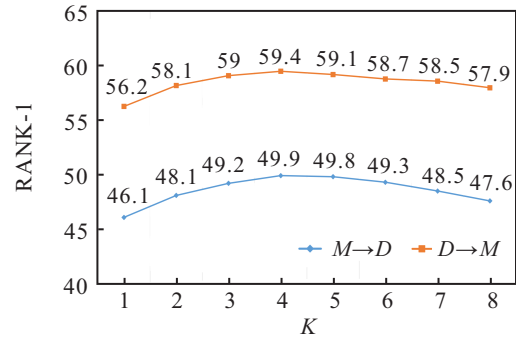


图6 参数K的选取

与主流方法进行对比分析,所对比方法均为近3年较有代表性的方法,实验结果从期刊原文中直接摘录. 以 DukeMTMC-reid 数据集为训练集, Market-1501 数据集为测试集, 本文取得了 59.4% 的 Rank-1 准确率以及 28.2 的 mAP; 以 Market-1501 数据集为训练集, DukeMTMC-reid 数据集为测试集, 本文取得了

49.9%的Rank-1准确率以及28.4的mAP.具体对比结果如表3所示.对比结果表明,所提出算法超过了大多数现有的域泛化行人再识别方法,与使用目标数据集信息的域自适应算法以及使用多个数据集训练的域泛化算法相比,具有一定的竞争力.

## 4 结论

本文提出了一种融合注意力机制的域泛化行人再识别模型,通过多尺度特征提取、特征融合注意力模块、多层次注意力模块以及风格正则化模块,进一步提高了域泛化行人再识别的准确率.在Market-1501和DukeMTMC-reid上进行实验,实验结果验证了所提出方法的有效性.后续将改进风格正则化模块,针对风格正则化模块设计损失函数,进一步增强所提出方法的泛化能力和鲁棒性.

## 参考文献(References)

- [1] Zheng Z D, Zheng L, Yang Y. Pedestrian alignment network for large-scale person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 29(10): 3037-3045.
- [2] 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.
- [3] Zhong Z, Zheng L, Zheng Z D, et al. Camera style adaptation for person Re-identification[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 5157-5166.
- [4] Chen W H, Chen X T, Zhang J G, et al. Beyond triplet loss: A deep quadruplet network for person Re-identification[C]. IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 1320-1329.
- [5] Yao H T, Zhang S L, Hong R C, et al. Deep representation learning with part loss for person re-identification[J]. IEEE Transactions on Image Processing, 2019, 28(6): 2860-2871.
- [6] Lin Y T, Dong X Y, Zheng L, et al. A bottom-up clustering approach to unsupervised person re-identification[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33: 8738-8745.
- [7] Bak S, Carr P, Lalonde J F. Domain adaptation through synthesis for unsupervised person re-identification[C]. European Conference on Computer Vision. Munich: Springer, 2018: 189-205.
- [8] Deng W J, Zheng L, Ye Q X, et al. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person Re-identification[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 994-1003.
- [9] Wang G C, Lai J H, Liang W Q, et al. Smoothing adversarial domain attack and P-memory reconsolidation for cross-domain person Re-identification[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, 2020: 10565-10574.
- [10] Kumar D, Siva P, Marchwica P, et al. Fairest of them all: Establishing a strong baseline for cross-domain person ReID[J/OL]. 2019, arXiv: 1907.12016.
- [11] Fan H H, Zheng L, Yan C G, et al. Unsupervised person re-identification[J]. ACM Transactions on Multimedia Computing, Communications, and Applications, 2018, 14(4): 1-18.
- [12] Yu H X, Zheng W S, Wu A C, et al. Unsupervised person Re-identification by soft multilabel learning[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, 2019: 2143-2152.
- [13] Zheng Z D, Zheng L, Yang Y. Pedestrian alignment network for large-scale person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 29(10): 3037-3045.
- [14] Ge Y, Chen D, Li H. Mutual mean-teaching: Pseudo label re-identification for unsupervised domain adaptation on person re-identification[J/OL]. 2020, arXiv: 2001.01526.
- [15] Lin S, Li H L, Li C T, et al. Multi-task mid-level feature alignment network for unsupervised cross-dataset person re-identification[J/OL]. 2018, arXiv:1807.01440.
- [16] Wang J Y, Zhu X T, Gong S G, et al. Transferable joint attribute-identity deep learning for unsupervised person Re-identification[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 2275-2284.
- [17] Liu J W, Zha Z J, Chen D, et al. Adaptive transfer network for cross-domain person re-identification[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, 2019: 7195-7204.
- [18] Liu H J, Cheng J, Wang S G, et al. Attention: A big surprise for cross-domain person re-identification[J/OL]. 2019, arXiv: 1905.12830.
- [19] Zheng Z D, Yang X D, Yu Z D, et al. Joint discriminative and generative learning for person re-identification[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, 2019: 2133-2142.
- [20] Huang Y R, Peng P X, Jin Y, et al. Domain adaptive attention model for unsupervised cross-domain person Re-identification[J/OL]. 2019, arXiv: 1905.10529.
- [21] Layne R, Hospedales T M, Gong S G. Attributes-based re-identification[M]. London: Springer, 2014: 93-117.
- [22] Wang Y N, Liao S C, Shao L. Surpassing real-world

- source training data: Random 3D characters for generalizable person Re-identification[C]. Proceedings of the 28th ACM International Conference on Multimedia. Seattle, 2020: 3422-3430.
- [23] Luo C C, Song C F, Zhang Z X. Generalizing person Re-identification by camera-aware invariance learning and cross-domain mixup[M]. Cham: Springer International Publishing, 2020: 224-241.
- [24] Jin X, Lan C L, Zeng W J, et al. Style normalization and restitution for generalizable person Re-identification[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, 2020: 3140-3149.
- [25] Chen P X, Dai P Y, Liu J Z, et al. Dual distribution alignment network for generalizable person Re-identification[J/OL]. 2020, arXiv: 2007.13249.
- [26] 李幼蛟, 卓力, 张菁, 等. 行人再识别技术综述[J]. 自动化学报, 2018, 44(9): 1554-1568.  
(Li Y J, Zhuo L, Zhang J, et al. A survey of person Re-identification[J]. Acta Automatica Sinica, 2018, 44(9): 1554-1568.)
- [27] 包宗铭, 龚声蓉, 钟珊, 等. 基于双向KNN排序优化的行人再识别算法[J]. 计算机科学, 2019, 46(11): 267-271.  
(Bao Z M, Gong S R, Zhong S, et al. Person Re-identification algorithm based on bidirectional KNN ranking optimization[J]. Computer Science, 2019, 46(11): 267-271.)
- [28] Chang X B, Hospedales T M, Xiang T. Multi-level factorisation net for person Re-identification[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, 2018: 2109-2118.
- [29] Qian X L, Fu Y W, Jiang Y G, et al. Multi-scale deep learning architectures for person Re-identification[C]. IEEE International Conference on Computer Vision. Venice, 2017: 5409-5418.
- [30] Zhou K Y, Yang Y X, Cavallaro A, et al. Omni-scale feature learning for person Re-identification[C]. IEEE/CVF International Conference on Computer Vision. Seoul, 2019: 3701-3711.
- [31] Chollet F. Xception: Deep learning with depthwise separable convolutions[C]. IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, 2017: 1800-1807.
- [32] Howard A G, Zhu M L, Chen B, et al. MobileNets: Efficient convolutional neural networks for mobile vision applications[J/OL]. 2017, arXiv: 1704.04861.
- [33] Chen B H, Deng W H, Hu J N. Mixed high-order attention network for person re-identification[C]. IEEE/CVF International Conference on Computer Vision. Seoul, 2019: 371-381.
- [34] Zhang R M, Li J Y, Sun H B, et al. SCAN: Self-and-collaborative attention network for video person re-identification[J]. IEEE Transactions on Image Processing, 2019, 28(10): 4870-4882.
- [35] Zhang C R, Wu Y X, Lei T. Unsupervised region attention network for person Re-identification[J]. IEEE Access, 2019, 7: 165520-165530.
- [36] Zhong Z, Zheng L, Zheng Z D, et al. CamStyle: A novel data augmentation method for person Re-identification[J]. IEEE Transactions on Image Processing, 2019, 28(3): 1176-1190.
- [37] Zhong Z, Zheng L, Li S Z, et al. Generalizing a person retrieval model hetero- and homogeneously[C]. European Conference on Computer Vision. Munich: Springer, 2018: 172-188.
- [38] Zhong Z, Zheng L, Luo Z M, et al. Invariance matters: Exemplar memory for domain adaptive person Re-identification[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, 2019: 598-607.

### 作者简介

于明(1964—), 男, 教授, 博士生导师, 从事计算机视觉、人工智能等研究, E-mail: yuming@hebut.edu.cn;

李学博(1996—), 男, 硕士生, 从事深度学习、行人再识别的研究, E-mail: 1412473462@qq.com;

郭迎春(1970—), 女, 副教授, 博士, 从事图像处理、模式识别等研究, E-mail: guoyingchun@hebut.edu.cn.

(责任编辑: 郑晓蕾)