http://bhxb.buaa.edu.cn jbuaa@buaa.edu.cn
DOI: 10.13700/j.bh.1001-5965.2023.0386

基于多标签对抗领域自适应的行人属性识别算法

胡强梁[1,2]，陈琳[1,2,*]，尚明生[1,2]

(1. 中国科学院 重庆绿色智能技术研究院 电子信息所，重庆 400714；2. 中国科学院大学重庆学院 人工智能学院，重庆 400714)

摘 要：针对无监督领域自适应算法通常局限于单标签学习问题，难以适配针对行人属性的多标签分类任务，提出一种多标签对抗领域自适应的行人属性识别算法。为适应行人属性多标签领域迁移任务，基于多标签特征分离模块，利用特定类别语义对主干网络提取的深度特征进行属性分离，有效提取特定属性的表征信息。针对不同领域属性特征分布差异较大的难点，提出基于分类器复用的多标签领域鉴别模块，同时实现多标签领域对齐和多标签分类，有效利用预测的鉴别信息捕获特征分布的多模式结构。实验结果表明：所提算法对比基准模型有明显提升，在平均准确率、准确率、召回率和F1指标上分别提升了4.49%、5.5%、11.44%和5.89%；所提算法为多标签领域自适应学习提供了新思路。

关 键 词：行人属性识别；多标签学习；多标签领域自适应；多属性特征提取；多标签分类
中图分类号：V351; TB391
文献标志码：A **文章编号**：1001-5965(2025)07-2478-10

行人属性识别通过提取给定人物图像中的年龄、性别等结构化属性信息，为视频智能化分析提供了结构化信息。与常规的单标签图像识别任务不同，行人属性识别通常需要对单个行人图像中数十个属性同时进行分析，因此，行人属性识别是更复杂的多标签分类任务。传统的行人属性识别方法通常侧重于从手工特征、属性分类器或属性关系的角度开发鲁棒的特征表示。随着深度学习技术的不断发展，基于深度学习的行人属性识别方法得到了广泛的研究和应用，其中包括基于全局特征、基于视觉注意力、基于序列预测、基于损失函数改进和基于图神经网络等诸多研究方向[1]。但基于深度学习的行人属性方法通常聚焦于有监督的学习任务，需要大量的标注数据实现模型的充分训练。

然而，构建大规模的行人图像数据集需要手动标记行人属性的多标签信息，相比单标签数据集，需要数倍的人力和时间成本。在处理未标记行人属性数据集时，如何充分利用已有标记行人属性数据集进行无监督训练，仍是一个重要且亟待解决的研究课题。

由于不同数据集使用不同图像传感器采集，并且受到光线、气候、角度等因素影响，在成像模态、分辨率和图片比例等方面都会存在差异，这就使得这些数据集之间存在显著的分布差异，这种差异被称为域偏移[2-3]。在存在域偏移的数据上直接进行无监督训练得到的模型通常难以保证其有效的识别精度。

在行人分析任务中，无监督领域自适应(unsupervised domain adaption, UDA)方法是有效解决域偏移的方法之一[2]，现有UDA研究主要通过探索领域不变特征表示方法，将领域知识迁移到目标任务中。例如，文献[4]将已标注的行人辅助身份标签转换为可学习的多标签软标记问题，挖掘潜在的标签信息提高无监督行人重识别任务精度。文献[5]

收稿日期：2023-06-16；**录用日期**：2023-09-23；**网络出版时间**：2023-11-29 09:20
网络出版地址：link.cnki.net/urlid/11.2625.V.20231128.1326.003
基金项目：国家自然科学基金（82030066,61902370）；重庆市教委在渝高校与中科院所合作重点项目（HZ2021008,HZ2021017）
*通信作者. E-mail：chenlin@cigit.ac.cn

引用格式：胡强梁，陈琳，尚明生. 基于多标签对抗领域自适应的行人属性识别算法[J]. 北京航空航天大学学报，2025，51(7)：2478-2487.
HU Q L，CHEN L，SHANG M S. Pedestrian attribute recognition algorithm based on multi-label adversarial domain adaptation[J]. Journal of Beijing University of Aeronautics and Astronautics，2025，51(7)：2478-2487 (in Chinese).

将单标签的无监督行人重识别转换为多标签学习任务, 利用同一身份来约束行人在不同图像中的特征一致性。然而, 直接将现有 UDA 方法直接套用到多标签的行人属性识别领域时, 由于特定属性关注不同的区域特征, 如帽子、眼镜等仅关注头部区域特征, 如果进行统一的领域特征学习, 可能在转换后的领域中混淆特有的属性特征, 导致实际应用效果欠缺。

针对上述问题, 本文提出基于无监督的多标签对抗域自适应行人属性识别算法, 利用对抗学习范式实现源域和目标域之间的多标签特征对齐和多属性预测。该算法由 2 个主要模块组成: 多标签特征分离 (multi-label feature disentanglement, MFD) 模块和多标签领域鉴别 (multi-label domain discriminator, MDD) 模块。其中, MFD 模块利用类别语义特征来指导学习与特定属性类别相关的图像特征, 并分离与该属性无关的图像特征, 使得分离后的属性特征更加关注相应的语义区域, 减少多标签领域迁移中的特征混淆问题。MDD 模块通过统一优化目标函数, 实现多标签领域对齐和分类, 利用预测的鉴别信息捕获多属性特征的复杂分布, 并有效地缓解负迁移现象。

本文工作主要有以下创新点:

1) 提出基于多标签对抗域自适应的行人属性识别算法, 将端到端的领域自适应方法引入多标签行人属性识别问题中, 用于提高无监督行人属性识别准确率。

2) 结合基于多标签语义解耦的多属性特征分离模块, 对主干网络提取的深度特征进行属性分离, 使领域自适应算法适应行人属性的多标签分类任务。同时引入基于多标签分类器的领域鉴别模块, 使对抗性领域自适应的训练过程更加稳定, 有效提高多标签领域迁移效果。

3) 建立了跨领域无监督行人属性识别训练方案, 在多个真实场景的行人属性数据集上进行了对比实验, 结果证明本文算法可以有效提高目标域缺失标注下的属性识别性能。

1　相关工作

1.1　行人属性识别

随着深度学习技术的不断发展, 基于深度学习的属性识别方法得到了广泛的研究和应用, 并取得了显著进展。文献 [6] 提出基于多属性联合学习的 DeepMAR 网络, 首次将神经网络引入到行人属性识别中。在此基础上, 许多学者发现行人属性识别任务与某些任务存在一定相关性, 有助于属性识别。如文献 [7] 提出基于多任务学习的属性识别, 该方法将属性识别与视角预测相结合。 文献 [8] 提出一种结合弱监督和定位方法的属性识别方法。文献 [9] 为减少与属性识别任务无关的额外特征的干扰, 提出结合定位引导任务的行人属性识别方法。文献 [10] 则认为姿态信息有助于属性识别任务, 引入姿态预测和人体部件定位任务。文献 [11] 提出用分割任务进一步提高对属性敏感区域的定位能力。文献 [12] 提出通过利用人体关键点精准捕获人体部件特征的方法。

近年来, 注意力机制也是深度学习领域应用较为广泛的组件之一, 许多学者也提出各种基于注意力机制的行人属性识别方法。文献 [13] 针对行人细粒度特征问题, 引入了注意力特征分支网络, 通过多个分支网络提取不同尺度的特征。文献 [14] 提出基于增广样本注意力一致的方法, 该方法建立在同一个样本图片经过不同的数据增广方法之后注意力区域应保持一致的假设上。文献 [15] 提出基于属性图片互相指导的注意力网络方法, 通过结合图片引导的特征和属性引导的特征, 为不同的属性定制不同综合特征的学习。文献 [16] 结合不同维度的注意力网络, 从多个角度出发捕获有助于属性识别的行人特征。

另外, 许多研究者发现行人属性之间存在着语义和空间关系, 可以利用这些关系提升属性识别效果。文献 [17] 提出可以根据属性之间的关系对所有属性进行分组, 然后使用长短时记忆网络学习组内组间属性的关系用以提升属性识别性能。文献 [18] 提出基于图卷积的属性识别方法, 首次将图卷积应用于该领域, 该方法的输入端除了行人图片额外增加了行人属性的语义词向量。文献 [19] 针对空间-语义关系、空间上下文及语义联系等 3 种关系进行建模, 提出基于序列上下文关系的学习模型。

上述基于深度学习的方法虽然已取得了较大的进展, 然而模型进行充分训练通常依赖于大量已标注数据, 当行人属性识别目标域缺失标注时, 如何基于领域自适应的行人属性识别方法来解决标注缺失的情况, 仍是当前行人属性识别中被忽视的一个难点问题。

1.2　无监督领域自适应

UDA 是迁移学习领域中备受关注的方法之一。主流的 UDA 方法存在几个大的方向, 其中, 基于度量标准和对抗学习的研究成果较为丰硕。在基于度量标准的研究上, 文献 [20] 提出深度适配网络 (deep adaptation network, DAN), 该模型引入了统

计学习中的多核最大均值差异，适配多个高层网络。文献 [21] 进一步提出了联合自适应网络，利用特征和标签的联合概率分布来减少不同域之间的分布差异。

在基于对抗学习的研究上，提出域对抗网络 (domain adversarial neural network, DANN)[22]。该方法关注于如何在不同域之间选择可供迁移的特征，把迁移任务看作域混淆问题，为此引入域判别器。文献 [23] 借鉴了这种思想，但与 DANN 不同的是利用重建目标域图像的方式来学习共享的编码表示。文献 [24] 将特征表示分为共享表示和私有表示，其中，共享表示可以用来进行迁移。 文献 [25] 提出对抗性区分域自适应模型，该模型通过在不同的训练阶段学习到 2 个独立的特征提取网络，但分类器是共享的，从而实现域适配。文献 [26] 认为只用一个判别器不能很好地捕捉到多模式结构，提出用多个判别器来捕捉细粒度特征。文献 [27] 提出循环一致性对抗域适应模型，使用生成性图像空间对齐和潜在表征空间对齐来适应不同领域。文献 [28] 提出的条件对抗域适应 (conditional adversarial domain adaptation, CDAN) 模型，利用分类器预测中所传递的判别信息来辅助对抗性适应。文献 [29] 提出的最大分类器差异 (maximum classifier discrepancy, MCD) 模型，使用 1 个特征提取器和 2 个任务特定的分类器，通过考虑特定任务的决策边界来调整目标域的分布。

以上所述的领域自适应算法通常仅考虑单标签迁移任务，而行人属性识别是典型的多标签多任务学习过程，如何利用多标签数据进行有效的领域知识迁移仍有待进一步的研究。

2 本文工作

本文提出的基于 UDA 的多标签对抗领域自适应 (multi-label adversarial domain adaptation, MLADA) 行人属性识别算法的模型框架如图 1 所示。该算法通过特征提取器 G，从带标注的源域数据和未标注的目标域数据中提取卷积神经网络（convolutional neural network, CNN）特征。这些特征输入到多属性特征分离模块 MFD 中，通过 MFD 将主干特征分离成基于特定属性的多标签特征向量。最终，分离后的特征向量进入多标签分类器产生分类损失。同时，多属性分类器 C 结合核范数 Wasserstein 距离 (nuclear-norm Wasserstein discrepancy, NWD) 差异和梯度反转层实现与特征提取器 G 的对抗学习，使得整个模型在实现属性识别的同时完成了领域对齐。

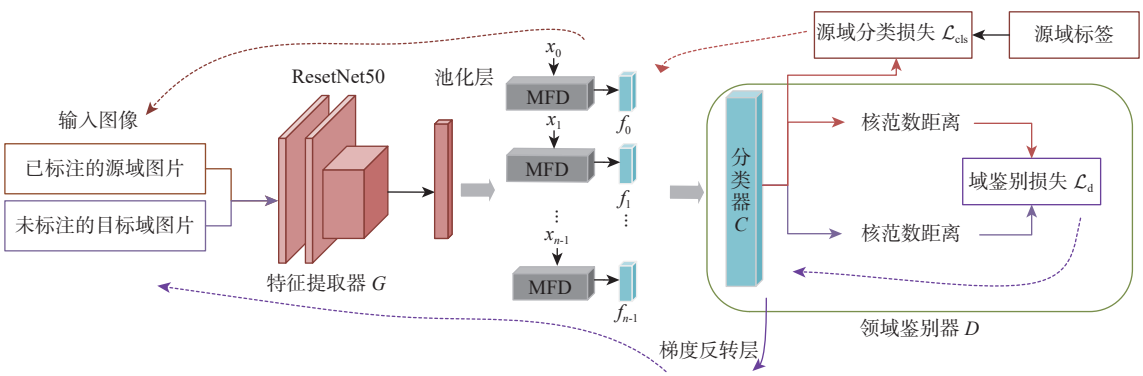

图 1 本文算法框架

Fig. 1 Framework of the proposed algorithm

2.1 符号定义

本文针对 UDA 多标签行人属性识别问题，源域中有标签的行人属性数据集定义如下：$\boldsymbol{D}_s=\{x_i^s,y_i^s\}_{i=1}^{n_s}$，其中，$x_i^s$ 表示源域的第 i 个行人图像样本，y_i^s 表示源域第 i 个图像样本对应的多属性标记，n_s 表示来自源域的样本数量。定义目标域的未标注行人属性数据集 $\boldsymbol{D}_t=\{x_i^t\}_{i=1}^{n_t}$，其中，$x_i^t$ 表示目标域的第 i 个行人图像样本，n_t 表示目标域的样本数量，目标域的样本标签未知。

2.2 模型概述

本文算法框架如图 1 所示，框架中红色部分表示的是特征提取器，使用 ResNet50 作为主干网络。灰色部分表示用于对主干特征进行特征解耦的多属性特征分离模块，蓝色部分表示行人属性识别的多标签分类器，由多个全连接网络构成。本文算法的整体损失函数由 2 个损失组成，分别为源域样本的分类损失 $\mathcal{L}_{cls}$，源域特征和目标域特征输入隐式多标签领域鉴别模块后产生的域鉴别损失 $\mathcal{L}_d$。

2.2.1 多标签特征分离模块

多标签特征分离模块以标签语义的注意机制为指导，学习语义特定的特征表示[30]。给定输入图像 $\boldsymbol{A}$，首先，通过特征提取器提取其特征图 $\boldsymbol{F}\in$

$\mathbf{R}^{W\times H\times N}$, 其中, W、H 和 N 分别为特征图的宽度、高度和通道数, 如式 (1) 所示:

$$\boldsymbol{F}^{\boldsymbol{A}} = f_{\text{cnn}}(\boldsymbol{A}) \tag{1}$$

式中: $f_{\text{cnn}}(\cdot)$为一个特征提取器, 由卷积网络实现, 本文使用的是 ResNet50 作为主干基础网络。对于每个标签 l, 本文算法使用预先训练的 Glove 模型提取 d 维语义嵌入向量, 具体实现时 d 设为 300, 如式 (2) 所示:

$$\boldsymbol{x}_l = f_{\text{g}}(w_l) \tag{2}$$

式中: w_l 为标签 l 的语义词; $f_{\text{g}}(\cdot)$为已训练好的 Glove 模型。得到语义特征向量 $\boldsymbol{x}_l$, 之后引入了一种语义引导的注意机制, 该机制结合了 $\boldsymbol{x}_l$来引导更多地关注语义感知区域, 从而学习与该标签相对应的特征向量。更具体地, 对图像中每个位置 (w,h), 首先, 使用低秩双线性方法融合相应的图像特征 $\boldsymbol{f}_{wh}^{\boldsymbol{A}}$和 $\boldsymbol{x}_l$, 如式 (3) 所示:

$$\tilde{\boldsymbol{f}}_{l,wh}^{\boldsymbol{A}} = \boldsymbol{P}^{\text{T}}\left(\tanh\left(\left(\boldsymbol{U}^{\text{T}}\boldsymbol{f}_{wh}^{\boldsymbol{A}}\right)\odot\left(\boldsymbol{V}^{\text{T}}\boldsymbol{x}_l\right)\right)\right)+\boldsymbol{b} \tag{3}$$

式中: tanh(·)为双曲正切函数; $\boldsymbol{U}\in\mathbf{R}^{N\times d_1}$、$\boldsymbol{V}\in\mathbf{R}^{d_{\text{s}}\times d_1}$、$\boldsymbol{P}\in\mathbf{R}^{d_1\times d_2}$、$\boldsymbol{b}\in\mathbf{R}^{1\times d_2}$为可学习的参数, $\odot$为逐元素乘法运算, d_1 和 d_2 分别为联合嵌入和输出特征的尺寸, d_{s} 为语义向量维度。然后, 在 $\boldsymbol{x}_l$的指导下计算注意力系数$\tilde{a}_{l,wh}$, 该注意力表示位置 (w,h) 的重要程度, 计算过程为

$$\tilde{a}_{l,wh} = f_{\text{att}}\left(\tilde{\boldsymbol{f}}_{l,wh}^{\boldsymbol{A}}\right) \tag{4}$$

式中: $f_{\text{att}}(\cdot)$为注意力函数, 由一个全连接网络实现。对所有位置重复该过程, 为了使不同样本的系数易于比较, 使用 softmax 函数对所有位置进行归一化:

$$a_{l,wh} = \frac{\text{e}^{\left(\tilde{a}_{l,wh}\right)}}{\sum\limits_{w',h'}\text{e}^{\left(\tilde{a}_{l,w'h'}\right)}} \tag{5}$$

最后, 对所有位置执行加权平均池化以获得编码与标签l相关信息的特征向量:

$$\boldsymbol{F}_l = \sum_{w,h} a_{l,wh}\boldsymbol{f}_{wh}^{\boldsymbol{A}} \tag{6}$$

2.2.2　多标签领域鉴别模块

现有的基于对抗训练的无监督域适应方法通常需要额外的鉴别器, 通过优化源域数据分类损失和领域鉴别损失来实现跨领域的知识迁移。然而这些方法忽视了用于源域数据分类的分类器, 其领域对抗性损失是由特征提取器和额外的领域鉴别器对抗产生的, 而分类器只考虑源域数据的分类准确性。事实上, 由于分类器是通过对源域数据进行监督学习得到的, 在源域数据上具有更高的准确性, 而在无标注的目标域上分类器识别率较低, 即: 针对源域中的相似数据, 分类器具有相似的分类结果, 而对于目标域的相似数据, 分类器表现得较为迟钝。

因此, 鉴于多标签分类器具有隐式辨别能力, 可结合核范数用于鉴别源域和目标域, 即基于复用分类器的领域鉴别模块[31], 本文将其扩展到多标签领域自适应问题, 提出 MDD 模块, 以提高多标签对抗学习能力。通过多标签分类器 C 可以得到预测矩阵 $C(\boldsymbol{F})\in\mathbf{R}^{b\times k}$, 其包含 b 个样本的 k 个属性的预测值, $C(\cdot)$表示分类函数。结合能提高预测多样性的核范数, 可以得出 2 个域属性特征分布的 NWD 差异 W, 该值可显示出分类器对 2 个域的鉴别能力, 如式 (7) 所示:

$$W = \sup_{\|\|C\|_*\|_L\leqslant K} E_{\tilde{\boldsymbol{D}}_{\text{s}}}[\|C(\boldsymbol{F})\|_*] - E_{\tilde{\boldsymbol{D}}_{\text{t}}}[\|C(\boldsymbol{F})\|_*] \tag{7}$$

式中: $\|\cdot\|_*$为核范数; $\|\cdot\|_L$为 Lipschitz 半范数; K 为 Lipschitz 常数; E 为期望; $\tilde{\boldsymbol{D}}_{\text{s}}$和 $\tilde{\boldsymbol{D}}_{\text{t}}$分别为 $\boldsymbol{D}_{\text{s}}$和 $\boldsymbol{D}_{\text{t}}$的特征分布。其中, 鉴别器可表述为$D=\|C\|_*$, 即结合了核范数的分类器。当分类器 C 用于分类时, 它可以实现对多标签图片的分类, 与此同时, 当 C 用作鉴别器时, 又能实现领域之间特征级别的对齐域鉴别损失的计算:

$$\mathcal{L}_{\text{d}}\left(\boldsymbol{x}^{\text{s}},\boldsymbol{x}^{\text{t}}\right) = \frac{1}{N_{\text{s}}}\sum_{i=1}^{N_{\text{s}}} D\left(Y\left(G\left(\boldsymbol{x}_i^{\text{s}}\right)\right)\right) - \frac{1}{N_{\text{t}}}\sum_{i=1}^{N_{\text{t}}} D\left(Y\left(G\left(\boldsymbol{x}_i^{\text{t}}\right)\right)\right) \tag{8}$$

式中: $Y(\cdot)$为多属性特征分离模块; $D(\cdot)$为判别器模块。

通过最大化域鉴别损失 $\mathcal{L}_{\text{d}}$使领域鉴别误差的最大化, 即特征提取器 G 生成的特征要混淆源域数据集和目标域数据集:

$$\hat{W} = \max_{D}\mathcal{L}_d\left(\boldsymbol{x}^{\text{s}},\boldsymbol{x}^{\text{t}}\right) \tag{9}$$

通过对抗训练, 当特征提取器 G 生成的特征使 C 无法鉴别时, 此时通过 G 提取的不同领域属性特征实现了一定程度的对齐。

2.2.3　多标签对抗训练损失函数

行人属性识别通过在源域 $\boldsymbol{D}_{\text{s}}$ 中有监督地训练得到多标签分类器 C, 将 C 用于无标签的源域 $\boldsymbol{D}_{\text{t}}$ 中得到预测结果。行人属性多标签分类损失表达式为

$$\mathcal{L}_{\text{cls}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{y}_i\ln(\sigma(C(\boldsymbol{F}))) + (1-y_i)\ln(1-\sigma(C(\boldsymbol{F}))) \tag{10}$$

式中: $\boldsymbol{F}$ 为经特征提取器和多属性特征分离模块处理得到的所有属性图像特征; $\sigma(\cdot)$为 sigmoid 函数。

目前被广泛采用的行人属性公开数据集中，都存在严重的不平衡问题，即在数据集中对应某种属性通常为负样本，只存在少数正样本（如是否光头、戴帽子等）。属性分布不均衡的现象导致行人属性识别模型在进行训练时接收大量负样本学习，模型识别更倾向于负例，导致模型对正样本的识别正确率较低。为缓解这一问题，通过引入样本权重 $\boldsymbol{w}$ 进行平衡[32]，其中，$\boldsymbol{w}=\{w_j\}_{j=1}^{K}$，$K$ 为数据集的属性数，w_j 的定义如下：

$$w_j=\begin{cases}\mathrm{e}^{m_j} & y_{ij}=0\\ \mathrm{e}^{1-m_j} & y_{ij}=1\end{cases} \tag{11}$$

式中：m_j 为第 j 个属性的正样本率。因此，基于样本权重的源域已标注数据的分类损失计算式为

$$\mathcal{L}_{\mathrm{cls}}=\frac{1}{N_{\mathrm{s}}}\sum_{i=1}^{N_{\mathrm{s}}}\boldsymbol{w}(\boldsymbol{y}_i^{\mathrm{s}}\ln(\sigma(C(\boldsymbol{F}_{\mathrm{s}})))+\left(1-\boldsymbol{y}_i^{\mathrm{s}}\right)\ln(1-\sigma(C(\boldsymbol{F}_{\mathrm{s}}))) \tag{12}$$

对抗学习过程如图 1 所示，对抗训练存在于特征提取器和分类器之间，其中，多标签分类器 C 同时完成多属性分类与领域鉴别，特征提取器 G 基于 MFD 模块生成具有区分度的属性特征，同时基于共享的主干网络，使源域与目标域上生成的特征一致来混淆分类器 C，从而实现多标签领域自适应任务。为避免模型的单调交替更新，使用梯度反转层来实现在一次反向传播内完成所有参数的更新，最终利用 min-max 模式完成领域自适应的对抗学习：

$$\mathcal{L}_{\mathrm{d}}=\min_{G}\max_{C}\mathcal{L}_{\mathrm{d}}\left(\boldsymbol{x}^{\mathrm{s}},\boldsymbol{x}^{\mathrm{t}}\right) \tag{13}$$

式中：特征提取器的优化方向为最小化 $\mathcal{L}_{\mathrm{d}}$ 实现不同领域之间的特征对齐，而分类器的优化方向为最大化 $\mathcal{L}_{\mathrm{d}}$ 实现。

综上所述，MLADA 模型的总损失函数 $\mathcal{L}_{\mathrm{total}}$ 由 2 部分组成，一是源域多标签分类损失 $\mathcal{L}_{\mathrm{cls}}$，二是用于领域自适应的鉴别损失 $\mathcal{L}_{\mathrm{d}}$：

$$\mathcal{L}_{\mathrm{total}}=\min_{G}\{\min_{C}\mathcal{L}_{\mathrm{cls}}(\boldsymbol{x}^{\mathrm{s}},\boldsymbol{y}^{\mathrm{s}})+\lambda\max_{C}\mathcal{L}_{\mathrm{d}}(\boldsymbol{x}^{\mathrm{s}},\boldsymbol{x}^{\mathrm{t}}))\} \tag{14}$$

式中：λ 为超参数用于平衡 $\mathcal{L}_{\mathrm{cls}}$ 和 $\mathcal{L}_{\mathrm{d}}$，文中设置为 1。通过对上述损失目标的优化，MLADA 可实现源域到目标域的多标签 UDA，有效提高行人属性识别模型在目标域中的识别准确率。

3 实验设置

3.1 数据集

RAPv1[33] 数据集是一个采集于室内监控场景的数据集，来源于 26 台不同位置的摄像机。包含 41 585 张图像，分辨率范围从 36×92 到 344×554 不等。其中，33 268 张图像被用于训练算法模型，剩余的图像则被用于测试模型的准确性。

RAPv2[33] 在 RAPv1 的基础上进行扩充，包含 84 928 张图像，这些图像被分为 3 个部分：训练集、验证集和测试集，分别包含 50 957 张、16 986 张和 16 985 张图像。

PA100K[13] 数据集是由 598 个真实户外监控摄像头捕获到的图像构建而成，包含 100 000 张行人图像，是目前为止最大的行人属性识别数据集。整个数据集分为训练集、验证集和测试集 3 个部分，比例为 8∶1∶1。

3.2 评估指标

本文使用 2 种评价方法来衡量属性识别算法的性能，分别是基于标签的方法和基于实例的方法。对于基于标签的方法，先计算每个属性的准确率，再取所有属性的平均准确率 (mean accuracy, mA) 作为评价指标，记为 E_{mA}，具体定义如下：

$$E_{\mathrm{mA}}=\frac{1}{2N}\sum_{i=1}^{K}\left(\frac{T_i^{\mathrm{P}}}{P_i}+\frac{T_i^{\mathrm{N}}}{N_i}\right) \tag{15}$$

式中：K 为行人属性个数；T_i^{P} 和 T_i^{N} 分别为第 i 个属性正确预测的正样本和负样本个数；P_i 和 N_i 分别为第 i 个属性的总的正样本和负样本的数量；N 为数据集中所有样本总数。

从基于实例的角度来看，评价指标通常包括准确率（accuracy, Acc）、精准度（precision, Prec）、召回率（recall, Rec）和 F1 值这 4 个常用的指标，分别记为 E_{Acc}、E_{Prec}、E_{Rec} 和 E_{F1}，具体定义如下：

$$E_{\mathrm{Acc}}=\frac{1}{N}\sum_{i=1}^{N}\frac{|y_i\cap\hat{y}_i|}{|y_i\cup\hat{y}_i|} \tag{16}$$

$$E_{\mathrm{Prec}}=\frac{1}{N}\sum_{i=1}^{N}\frac{|y_i\cap\hat{y}_i|}{|\hat{y}_i|} \tag{17}$$

$$E_{\mathrm{Rec}}=\frac{1}{N}\sum_{i=1}^{N}\frac{|y_i\cap\hat{y}_i|}{|y_i|} \tag{18}$$

$$E_{\mathrm{F1}}=\frac{2E_{\mathrm{Prec}}E_{\mathrm{Rec}}}{E_{\mathrm{Prec}}+E_{\mathrm{Rec}}} \tag{19}$$

式中：y_i 为测试样本中第 i 个属性的真实标签；$\hat{y}_i$ 为模型输出的第 i 个属性的预测结果。

3.3 实验设置

本文算法与其他 UDA 最新的先进算法进行了比较，为确保实验结果的准确性，对所有对比算法都进行了官方代码复现实验，并保证了数据集预处理方法和计算评价指标的方法完全相同。在实验中，本文对图片应用了一些数据增强技术（如随机擦除、随机裁剪、水平翻转），以提高数据的多样性

和实验的鲁棒性。基础网络采用 ResNet50 作为主干网络, 优化器采用随机梯度下降(stochastic gradient descent, SGD) 算法, 训练时实验软硬件环境和模型的具体参数如表 1 和表 2 所示。

表 1　实验软硬件环境

Table 1　Experimental hardware and software environment

实验环境	配置
CPU	Intel Xeon E5-2680 v4
GPU	NVIDIA GeForce RTX 1080Ti
操作系统	Ubuntu 18.04.2
深度学习框架	Pytorch 1.9
计算机语言	Python 3.8

表 2　模型参数设置

Table 2　Model parameters setting

参数	数值
图片预处理大小	256×192
批处理大小	64
学习率	0.01
动量大小	0.9
训练轮次	30

为更准确地评估本文算法的性能, 本文在 3 个常用行人属性识别数据集 PA100k、RAPv1 和 RAPv2 上进行实验, 并选取其公共 10 个属性进行 2 组 UDA 实验。第 1 组: RAPv1⇛PA100k。第 2 组: RAPv2⇛PA100k。定义 A⇛B 表示, 以 B 为目标域, A 为源域, 进行 UDA 实验。

4　实验结果与分析

4.1　对比实验分析

为评估 MLADA 在 UDA 行人属性识别任务上的有效性, 将本文算方法与 UDA 中的先进算法进行对比实验, 包括基于域鉴别器对抗的 DANN[22]、基于多核最大均值差异的 DAN[20]、基于最大化双分类器差异的 MCD[29]、基于双调节策略的 CDAN[28] 及基于复用分类器的(Discriminator-free Adversarial Learning Network, DALN)[31]。实验中所有算法的基准模型使用行人属性领域中基于改进损失函数的 StrongBaseline[32], 该基准模型直接在源域上进行训练并在目标域上进行预测得到属性识别结果。

图 2 展示了上述 5 种算法与本文算法的效果对比, 可见本文算法在 RAPv1⇛PA100k 和 RAPv2⇛PA100k 这 2 种实验场景下, 性能都优于其他对比算法。图 2(a)使用 RAPv1 数据集全部 41 585 张图片作为源域已标注数据, 80 000 张 PA100k 训练集数据作为目标域未标注数据, 本文算法在 E_{mA}、

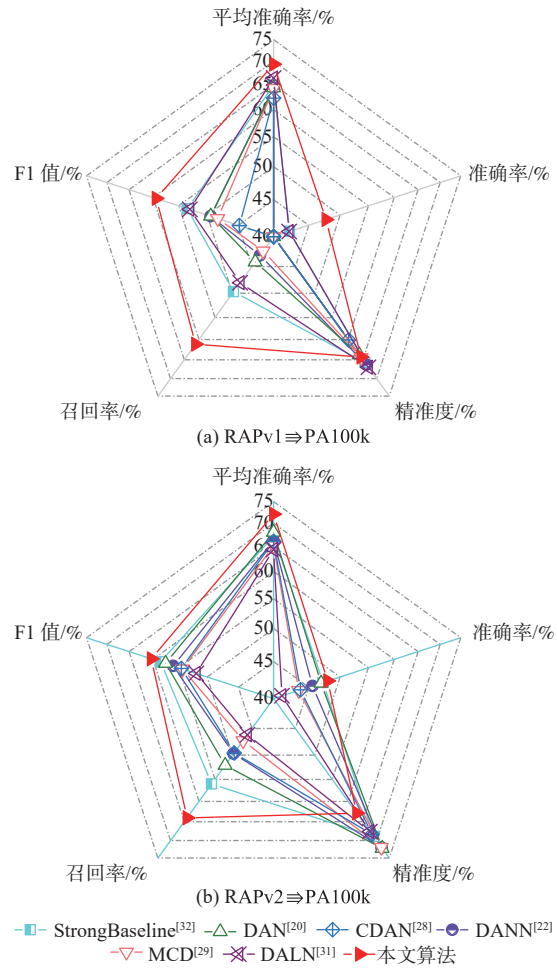

图 2　对比实验效果

Fig. 2　Comparison of experiment results

E_{Acc} 和 E_{Rec} 指标上提升明显, 均超过了基线算法及对比的领域自适应算法, 例如, 对比最近的 UDA 算法 DALN[28], 本文算法在 E_{mA}、E_{Acc} 和 E_{Rec} 提升了 7.7%、6.85% 和 17.86%, 虽然由于召回率大幅提升导致 E_{Prec} 指标下降(−4.71%), 但在综合识别率 E_{F1} 指标上提升了 8.45%。

图 2(b)使用 RAPv2 数据集全部 84 928 张图片作为源域已标注数据, 本文算法在 E_{mA}、E_{Acc}、E_{Rec} 和 E_{F1} 指标上分别比次优效果提升了 4.03%、0.75%、8.01% 和 1.86%。从上述分析可得, 本文算法通过结合 MFD 模块和 MDD 模块, 在应对多标签无监督领域属性识别自适应任务上, 有着明显的优势, 可以有效提升标签缺失情况下的行人属性识别性能。

图 3 展示了第 1 组 RAPv1⇛PA100k 实验中 6 种测试算法的行人属性准确度。在大多数属性中, 本文算法都优于其他对比算法, 只有性别、短袖和靴子不是最优。其中, 本文算法的裤子属性比次优高出了 8.81%。

图 4 展示了第 2 组 RAPv2⇛PA100k 实验中,

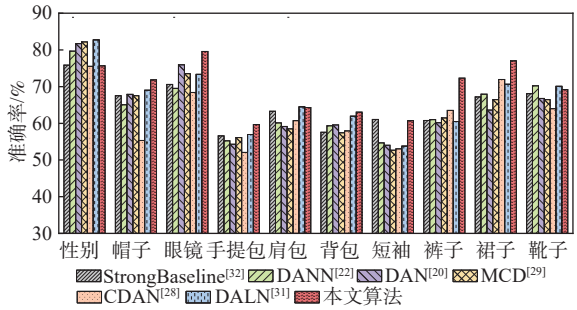

图3 RAPv1⇛PA100k的各属性准确率

Fig. 3 Accuracy of each attribute on RAPv1⇛PA100k

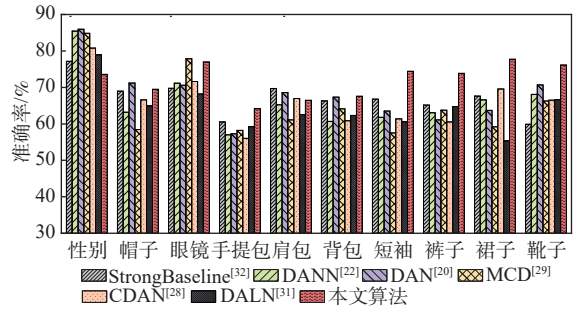

图4 RAPv2⇛PA100k的各属性准确率

Fig. 4 Accuracy of each attribute on RAPv2⇛PA100k

6种算法的各属性准确度对比。在多数属性中，本文算法均优于其他对比算法，尤其是短袖、裤子和裙子属性，比次优算法分别高出7.64%、8.74%和8.17%，其他属性识别结果虽不是最优但也和最优相差无几。

为进一步分析本文算法相较于常规的基线行人属性识别性能差异，使用StrongBaseline[32]和本文算法在PA100k测试集上进行属性识别效果分析，这2种模型都是在RAPv2⇛PA100k实验下使用对应的方法训练而得，如图5所示。相比于未使用领域自适应的基线模型StrongBaseline，本文算法能够更加准确识别多个属性，例如：背包属性，StrongBaseline识别成了手提包或肩包。另外，StrongBaseline更容易造成漏识别，如眼镜、短袖等，上述结果表明本文算法有效降低源域与目标的领域偏移从而得到更好的识别结果。



图5 属性识别效果分析(RAPv2⇛PA100k)

Fig. 5 Performance of attribute recognition on RAPv2⇛PA100k

综上所述，本文算法相较于现有的对比算法，在多标签行人属性识别领域自适应效果上有着更好的表现。

4.2 消融实验分析

为评估不同模块对实验效果的贡献，在不加入对抗损失情况下，以ResNet50为基线模型，仅使用行人属性多标签分类损失$\mathcal{L}_{\text{cls}}$，以及在$\mathcal{L}_{\text{cls}}$中加入多标签分离模块MFD(ResNet50+MFD)进行对比分析。其次，在加入领域自适应鉴别损失$\mathcal{L}_{\text{d}}$中，包括了加入使用MFD及传统的DANN[22]模型进行对抗训练(ResNet50+MFD+DA)，单独加入多标签领域鉴别模块MDD(ResNet50+MDD)，以及加入MFD后(ResNet50+MFD+MDD，即本文算法)的结果进行对比分析。

在RAPv1⇛PA100k和RAPv2⇛PA100k这2种实验场景下进行消融实验，实验结果如表3和表4所示，最优结果用加粗表示。

由表3和表4可知，在加入MFD模块后，在2组实验下E_{mA}值都有所提高，分别提升了0.47%和1.63%，说明多标签特征分离有助于主干网络提取更有利于行人属性识别的特征。在加入对抗损失后，单独加入MDD模块时，第1组实验下E_{mA}值

表 3　在 RAPv1⇒PA100k 上的消融实验

Table 3　Ablation experiments on RAPv1⇒PA100k

算法	E_{mA}/%	E_{Acc}/%	E_{Prec}/%	E_{Rec}/%	E_{F1}/%
ResNet50	64.85	42.22	65.75	49.71	53.44
ResNet50+MFD	65.32	43.04	**67.17**	51.93	54.99
ResNet50+MFD+DA	65.83	39.41	63.43	45.86	50.01
ResNet50+MDD	66.34	42.14	66.95	47.99	52.99
本文算法	**69.34**	**47.72**	64.33	**61.15**	**59.33**

表 4　在 RAPv2⇒PA100k 上的消融实验

Table 4　Ablation experiments on RAPv2⇒PA100k

算法	E_{mA}/%	E_{Acc}/%	E_{Prec}/%	E_{Rec}/%	E_{F1}/%
ResNet50	67.19	47.24	68.93	56.01	58.47
ResNet50+MFD	68.82	47.92	**73.15**	55.28	59.57
ResNet50+MFD+DA	67.74	44.51	66.22	53.26	55.66
ResNet50+MDD	64.34	41.14	67.56	46.15	51.88
本文算法	**72.04**	**47.99**	62.85	**64.01**	**60.33**

提升了 1.49%, 而在第 2 组实验 E_{mA} 值反而下降, 这说明 MDD 模块在没有属性分离时, 对于特定的属性特征提取困难, 导致模型的性能提升并不稳定。值得注意的是, 使用传统对抗训练结合 MFD 模块(ResNet50+MFD+DA)的结果, 对比没有对抗学习的 ResNet50+MFD 模型在 2 个数据集上效果均有所下降, 说明如不使用 MDD 模块复用分类器进行多属性对抗, 导致了一定的领域混淆, 从而导致精度下降。

当同时加入 MFD 模块和 MDD 模块时, 在 2 组实验下 E_{mA}、E_{Acc}、E_{Rec} 和 E_{F1} 值都达到了最高。在第 1 组实验中, 对比基线 ResNet50 模型, 本文算法的 E_{mA}、E_{Acc}、E_{Rec} 和 E_{F1} 值分别提高了 4.49%、5.5%、11.44% 和 5.89%。值得注意的是, 由于召回率的显著性提升(+11.44%), 导致识别精度 E_{Prec} 有所下降(−6.08%), 但综合性能 E_{F1} 值保持了增长(+5.89%)。另外, 考虑实际应用场景中召回率相比识别精度更为重要(如检出背包、特定鞋子等属性), 因此, 本文算法更具有实用价值。由上述分析表明, 在 MFD 和 MDD 模块的共同作用下, 可以有效提升 UDA 多标签属性识别性能。

5　结　论

本文提出一种基于 UDA 的多标签行人属性识别算法, 该算法包括 2 个主要模块: 多属性特征分离模块和基于分类器复用的多标签领域鉴别模块。

1) 多属性特征分离模块通过使用类别的语义特征来指导图像特征的学习, 实现了特定属性特征分离提取。基于分类器复用的多标签领域鉴别模块通过使用对抗范式, 实现了多标签领域对齐和标签分类。

2) 本文在多个公开数据集上进行了实验, 结果表明, 本文算法在各项评价指标上都优于现有算法, 证明了其有效性和可行性。

在下一步工作中, 拟基于半监督自适应的方法, 利用目标域中的少量标签数据对多标签分类器进行增强, 进一步提高行人属性领域自适应学习的精度。

参考文献 (References)

[1] WANG X, ZHENG S F, YANG R, et al. Pedestrian attribute recognition: a survey[J]. Pattern Recognition, 2022, 121: 108220.

[2] WILSON G, COOK D J. A survey of unsupervised deep domain adaptation[J]. ACM Transactions on Intelligent Systems and Technology, 2020, 11(5): 1-46.

[3] STACKE K, EILERTSEN G, UNGER J, et al. Measuring domain shift for deep learning in histopathology[J]. IEEE Journal of Biomedical and Health Informatics, 2021, 25(2): 325-336.

[4] YU H X, ZHENG W S, WU A, et al. Unsupervised person re-identification by soft multilabel learning[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 2148-2157.

[5] WANG D, ZHANG S. Unsupervised person re-identification via multi-label classification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2020: 10978-10987.

[6] LI D W, CHEN X T, HUANG K Q. Multi-attribute learning for pedestrian attribute recognition in surveillance scenarios[C]//Proceedings of 3rd IAPR Asian Conference on Pattern Recognition. Piscataway: IEEE Press, 2015: 111-115.

[7] SARFRAZ M S, SCHUMANN A, WANG Y, et al. Deep view-sensitive pedestrian attribute inference in an end-to-end model[C]//Proceedings of the British Machine Vision Conference. London: BMVA Press, 2017, 134: 1-13.

[8] YU K, LENG B, ZHANG Z, et al. Weakly-supervised learning of mid-level features for pedestrian attribute recognition and localization[C]//Proceedings of the British Machine Vision Conference. London: BMVA Press, 2017: 1-11.

[9] LIU P, LIU X, YAN J, et al. Localization guided learning for pedestrian attribute recognition[C]//Proceedings of the British Machine Vision Conference. London: BMVA Press, 2018: 142-152.

[10] LI D, CHEN X, ZHANG Z, et al. Pose guided deep model for pedestrian attribute recognition in surveillance scenarios[C]//Proceedings of the IEEE International Conference on Multimedia and Expo. Piscataway: IEEE Press, 2018: 1-6.

[11] WU M D, HUANG D, GUO Y F, et al. Distraction-aware feature learning for human attribute recognition via coarse-to-fine attention mechanism[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2020: 12394-12401.

[12] ZHANG J, REN P, LI J. Deep template matching for pedestrian attribute recognition with the auxiliary supervision of attribute-wise keypoints[EB/OL]. (2020-11-13)[2020-05-26]. https://doi.org/10.48550/arXiv.2011.06798.

[13] LIU X, ZHAO H, TIAN M, et al. Hydraplus-net: attentive deep features for pedestrian analysis[C]//Proceedings of the IEEE International Conference on Computer Vision. Piscataway: IEEE Press, 2017: 350-359.

[14] GUO H, ZHENG K, FAN X, et al. Visual attention consistency under image transforms for multi-label image classification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2019: 729-739.

[15] JI Z, HE E, WANG H, et al. Image-attribute reciprocally guided attention network for pedestrian attribute recognition[J]. Pattern Recognition Letters, 2019, 120: 89-95.

[16] TAN Z, YANG Y, WAN J, et al. Attention-based pedestrian attribute analysis[J]. IEEE Transactions on Image Processing, 2019, 28 (12): 6126-6140.

[17] ZHAO H, TIAN M, SUN S, et al. Spindle net: person re-identification with human body region guided feature decomposition and fusion[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 1077-1085.

[18] LI Q, ZHAO X, HE R, et al. Visual-semantic graph reasoning for pedestrian attribute recognition[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2019, 33(1): 8634-8641.

[19] WU J, LIU H, JIANG J, et al. Person attribute recognition by sequence contextual relation learning[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2020, 30(10): 3398-3412.

[20] LONG M, CAO Y, WANG J, et al. Learning transferable features with deep adaptation networks[C]//Proceedings of the International Conference on Machine Learning. Piscataway: IEEE Press, 2015: 97-105.

[21] LONG M, ZHU H, WANG J, et al. Deep transfer learning with joint adaptation networks[C]//Proceedings of the International Conference on Machine Learning. Piscataway: IEEE Press, 2017: 2208-2217.

[22] GANIN Y, LEMPITSKY V. Unsupervised domain adaptation by backpropagation[C]//Proceedings of the International Conference on Machine Learning. Piscataway: IEEE Press, 2015, 9908: 1180-1189.

[23] GHIFARY M, KLEIJN W B, ZHANG M, et al. Deep reconstruction-classification networks for unsupervised domain adaptation[C]// Proceedings of the European Conference on Computer Vision. Berlin: Springer, 2016: 597-613.

[24] BOUSMALIS K, TRIGEORGIS G, SILBERMAN N, et al. Domain separation networks[C]//Proceedings of the Annual Conference on Neural Information Processing Systems. Piscataway: IEEE Press, 2016, 29: 343-351.

[25] TZENG E, HOFFMAN J, SAENKO K, et al. Adversarial discriminative domain adaptation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2017: 7167-7176.

[26] PEI Z, CAO Z, LONG M, et al. Multi-adversarial domain adaptation[C]//Proceedings of the AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2018, 32(1): 3934-3941.

[27] HOFFMAN J, TZENG E, PARK T, et al. Cycada: cycle-consistent adversarial domain adaptation[C]//Proceedings of the International Conference on Machine Learning. Piscataway: IEEE Press, 2018: 1989-1998.

[28] LONG M, CAO Z, WANG J, et al. Conditional adversarial domain adaptation[C]//Proceedings of the Annual Conference on Neural Information Processing Systems. Piscataway: IEEE Press, 2018, 31: 1-11.

[29] SAITO K, WATANABE K, USHIKU Y, et al. Maximum classifier discrepancy for unsupervised domain adaptation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 3723-3732.

[30] CHEN T, XU M, HUI X, et al. Learning semantic-specific graph representation for multi-label image recognition[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. Piscataway: IEEE Press, 2019: 522-531.

[31] CHEN L, CHEN H, WEI Z, et al. Reusing the task-specific classifier as a discriminator: discriminator-free adversarial domain adaptation[C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2022: 7181-7190.

[32] JIA J, HUANG H, YANG W, et al. Rethinking of pedestrian attribute recognition: realistic datasets with efficient method[EB/OL]. (2020-05-25)[2020-05-26]. https://doi.org/10.48550/arXiv.2005.11909.

[33] LI D, ZHANG Z, CHEN X, et al. A richly annotated pedestrian dataset for person retrieval in real surveillance scenarios[J]. IEEE Transactions on Image Processing, 2018, 28(4): 1575-1590.

Pedestrian attribute recognition algorithm based on multi-label adversarial domain adaptation

HU Qiangliang[1, 2], CHEN Lin[1, 2, *], SHANG Mingsheng[1, 2]

(1. Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing 400714, China;
2. School of Artificial Intelligence, Chongqing School of University of Chinese Academy of Sciences, Chongqing 400714, China)

Abstract: Current unsupervised domain adaption algorithms usually consider only single-label learning, failing to adapt to multi-label classification tasks in pedestrian attribute recognition. To address this issue, a multi-label adversarial domain adaptation algorithm was proposed for pedestrian attribute recognition. First, to adapt to the domain transfer task of multiple attribute labels, a multi-label feature disentanglement module was employed to effectively disentangle attributes of deep features extracted from the backbone network with specific attributes based on category-specific semantics. Secondly, to reduce the gap of attribute feature distributions in different domains, a multi-label domain discrimination module based on classifier reuse was proposed to achieve both multi-attribute domain alignment and multi-label classification, which effectively exploited the predicted discrimination information to capture the multi-mode structures of feature distribution. Experimental results show that the proposed algorithm achieves superior results than the baseline model, with improvements of 4.49%, 5.5%, 11.44% and 5.89% on mean accuracy, accuracy, recall and F1 values, respectively. The proposed algorithm provides a new insight for multi-label domain adaptation learning.

Keywords: pedestrian attribute recognition; multi-label learning; multi-label domain adaptation; multi-label feature extraction; multi-label classification

Received: 2023-06-16; **Accepted**: 2023-09-23; **Published Online**: 2023-11-29 09:20

URL: link.cnki.net/urlid/11.2625.V.20231128.1326.003

Foundation items: National Natural Science Foundation of China (82030066,61902370); The Key Project of Chongqing Municipal Education Commission based on Cooperation of Chongqing Local Universities and the Chinese Academy of Sciences (HZ2021008,HZ2021017)

*** Corresponding author**. E-mail: chenlin@cigit.ac.cn