

基于自适应融合网络的跨域行人重识别方法

郭迎春¹ 冯放¹ 阎刚¹ 郝小可¹

摘 要 无监督跨域的行人重识别旨在将从有标签的源域中学习到的知识迁移到无标签的目标域, 具有实用性和有效性而得到广泛关注. 基于聚类的跨域行人重识别可以生成伪标签并对模型进行优化使得其表现较其他方法更优, 然而这类方法由于过于依赖聚类伪标签的准确性, 忽略了对伪标签噪声的处理, 导致噪声随着网络迭代而不断扩大, 影响模型的鲁棒性. 针对这个问题, 提出了基于自适应融合网络的方法, 利用双网络结构共同学习, 并将学习到的知识进行融合得到融合网络; 为了区分两个网络的学习能力, 设计了自适应融合策略; 同时, 利用细粒度风格转换模块对目标域数据集进行处理, 降低行人图像对相机变换的敏感度. 在行人重识别基准数据集 Market1501、DukeMTMC-ReID 和 MSMT17 上, 通过评估指标平均精度均值和 Rank- n 与主流的方法进行了对比实验, 验证了该方法的有效性.

关键词 跨域行人重识别, 自适应融合网络, 细粒度风格转换, 深度学习

引用格式 郭迎春, 冯放, 阎刚, 郝小可. 基于自适应融合网络的跨域行人重识别方法. 自动化学报, 2022, 48(11): 2744–2756

DOI 10.16383/j.aas.c220083

Cross-domain Person Re-identification on Adaptive Fusion Network

GUO Ying-Chun¹ FENG Fang¹ YAN Gang¹ HAO Xiao-Ke¹

Abstract Unsupervised cross-domain person re-identification aims to transfer the knowledge learned from labeled source domain to unlabeled target domain, which has attracted wide attention due to its practicability and effectiveness. Cross-domain person re-identification based on clustering can generate pseudo-labels and optimize the model to make its performance better than other methods. However, these methods rely too much on the accuracy of clustering pseudo labels and ignore to deal with pseudo-label noise, which leads to the continuous expansion of noise with network iteration and affects the robustness of the models. To address this problem, this paper proposes a method based on fine-grained style transfer and adaptive fusion network, which uses dual network structure to learn together and fuse the learned knowledge to obtain a fusion network. To treat the learning ability of the two networks differently, an adaptive fusion strategy is designed based on the different weights of the two networks in each fusion process. At the same time, a fine-grained style transfer module is used to process the target domain dataset, thereby reducing the sensitivity of person images to camera transformation. On the person re-identification benchmark datasets Market1501, DukeMTMC-ReID and MSMT17, the effectiveness of the proposed method was verified by comparing mean average precision and Rank- n with the state-of-the-art methods.

Key words Cross-domain person re-identification, adaptive fusion network, fine-grained style transfer, deep learning

Citation Guo Ying-Chun, Feng Fang, Yan Gang, Hao Xiao-Ke. Cross-domain person re-identification on adaptive fusion network. *Acta Automatica Sinica*, 2022, 48(11): 2744–2756

行人重识别^[1–3]利用计算机视觉技术判断不同摄像头的图像或者视频序列中是否存在特定行人, 广泛应用于智能安防、无人超市、人机交互等多个领域. 随着深度学习的应用普及, 行人重识别获得

了发展条件, 其在有监督领域上的准确率得到了大幅度的提升^[4]. 但是有监督的行人重识别因需要带有真实标签的数据集而脱离实际应用, 无监督的行人重识别弥补了有监督学习的需要真实标签的劣势. 目前无监督的行人重识别所能达到的精度远远不能满足现实生活应用需求, 跨域的行人重识别由于其只需要一个预训练源域数据集带有真实标签而被识别的目标域数据集无需标签, 成为未来研究发展的方向^[5].

由于不同场景存在着摄像头参数、背景和光照等问题, 简单地将在源域上进行训练所得到的模型应用于目标域上进行测试会导致行人重识别性能的

收稿日期 2022-01-28 录用日期 2022-07-30

Manuscript received January 28, 2022; accepted July 30, 2022

国家自然科学基金 (60302018, 61806071, 62102129), 河北省自然科学基金 (F2019202381, F2019202464) 资助

Supported by National Natural Science Foundation of China (60302018, 61806071, 62102129) and Natural Science Foundation of Hebei Province (F2019202381, F2019202464)

本文责任编辑 杨健

Recommended by Associate Editor YANG Jian

1. 河北工业大学人工智能与数据科学学院 天津 300401

1. School of Artificial Intelligence, Hebei University of Technology, Tianjin 300401

下降^[6-7]。同时, 传统的跨域问题常常是基于闭集场景下的, 即假设源域和目标域共享完全相同的类, 但是这种假设并不适用于行人重识别问题。跨域的行人重识别数据集通常是在不同时空下获得的, 所以源域与目标域的图像通常具有不同的身份信息, 应该将其看作是开集问题, 这比闭集问题更具有挑战性。

本文研究跨域方法, 着重解决现有基于聚类的跨域行人重识别方法中忽略对伪标签噪声处理的问题, 提出了基于自适应融合网络的行人重识别模型, 主要创新和贡献如下:

1) 本文提出的融合网络模型, 采用双网络结构共同学习并进行网络融合, 利用融合后的网络反过来监督双网络结构的训练, 这种知识共享的模型结构可以有效地防止聚类伪标签噪声的扩大。

2) 本文提出了自适应融合策略, 使得模型在每一次融合过程中, 可以针对两个网络的学习情况分配不同权重自适应地对两个网络知识进行融合。

3) 本文设计了细粒度风格转换模块来降低不同相机间视角下的风格差异性, 提高模型对时空变化的鲁棒性。

本文的安排如下: 第1节介绍跨域行人重识别主要分类及相关工作; 第2节介绍本文提出的基于自适应融合网络的跨域行人重识别方法; 第3节给出实验设置并对实验结果进行分析; 第4节总结本文工作并对未来工作进行展望。

1 相关工作

无监督跨域的行人重识别研究是在无监督的行人重识别研究上发展而来。相较于无监督的方法, 无监督跨域的方法额外利用一个具有标签的源域数据集, 因此会有一定的先验知识作为指导, 识别效果也会更优。最近的无监督行人重识别通过挖掘特征间的内在联系解决图像由于时空变换引起的特征分布的差异性, 从而提高了识别精度, 其中具有代表性的是文献[8-10], 分别利用相机感知代理处理相机变化的问题、元信息构建超图挖掘潜在的价值信息、成对相似性处理不同实例在增强视图间的匹配, 对未来跨域无监督的行人重识别发展具有一定的参考意义。

目前跨域行人重识别方法主要可以分为四类:

1) 域分布对齐的跨域行人重识别^[11-12], 通过对齐源域与目标域的数据分布来减小域间隙。域分布对齐意在将目标域的数据分布尽可能地与源域数据分布一致, 这样可以减少域间隙, 从而提高实验的准确率。Wang 等^[11]利用额外标注的行人属性信息,

通过身份标签分支和属性分支结合训练, 实现网络间信息的交融, 最终学习到行人更本质的特征。Djebril 等^[12]认为解决跨域问题应该将重点放在对齐域之间的成对差异而不是特征表示, 为此提出一个基于非相似最大平均差异 (Dissimilarity-based maximum mean discrepancy, D-MMD) 损失来最小化成对差异。域分布对齐应用在跨域的行人重识别方向往往需要考虑更多的因素而导致模型较为复杂。

2) 域不变性的跨域行人重识别^[13-16], 该方法利用生成对抗网络 (Generative adversarial networks, GAN) 来处理数据图像以获得相似的数据集间图像风格。Liu 等^[13]将跨域转换分解为光照、相机角度、分辨率三个因子转换, 每个因子作为一个子风格, 为每个子风格训练一个生成器并提出一种可以权衡各种因子影响程度从而进行融合的网络 Adaptive transfer network (ATNet), 该网络通过最小化中间层的子任务达到细粒度级别风格迁移。文献[14]提出的相似性保持生成对抗网络 (Similarity preserving GAN, SPGAN) 和 Wei 等^[15]提出的行人迁移生成对抗网络 (Person transfer GAN, PTGAN) 都是利用生成对抗网络将源域的图像风格转换为目标域的图像风格, 同时保留源域图像的标签信息, 使得源域和目标域间的知识转移更具有鲁棒性。Zhong 等^[16]通过挖掘相机配置之间的关系, 基于生成对抗网络通过生成域内不同相机之间的图像来解决域内相机间隙问题。但是这些方法过度依赖于生成图像的质量, 导致准确性普遍低于有监督的行人重识别。

3) 计算特征相似性产生的软标签来优化网络^[17-19], 通常是利用保存所有图像的平均特征或是利用辅助数据集构建的特征空间来表示软标签。计算特征相似性是利用已经训练过的图像或是辅助数据集图像等构建特征空间来表示将要训练的图像。Zhong 等^[17]提出的范例相机邻域不变性 (Exemplar camera neighborhood invariance, ECN) 利用存储器结构来保存平均特征进而为训练图像来分配软标签, 利用样本不变性、相机不变性、邻域不变性这三个不变性监督优化网络。Yu 等^[18]提出的深度软多标签参考学习 (Deep soft multi-label reference learning, MAR) 利用 MSMT17 数据集^[15]作为辅助数据集构建特征空间, 并在特征空间中表示目标域数据集图像。这些方法的不足就是特征或参考数据集不具有足够的代表性, 使得软标签不够准确, 降低模型性能。

4) 基于聚类的方法^[20-26], 为了充分利用无标签的目标域数据, 利用聚类算法产生的伪标签来当作

目标域的标签,这类方法在大量的实验中被证明在当前具有最好的效果. Fu 等^[20]提出的自相似性分组 (Self-similarity grouping, SSG) 模型通过将特征图垂直平均划分为六个局部特征,利用局部特征来分配多尺度的聚类伪标签. Zhai 等^[21]提出一种新的判别聚类的方法增广判别聚类 (Augmented discriminative clustering, AD-Cluster), 通过基于密度的聚类算法, 自适应地扩充样本和判别特征学习来解决无监督跨域的行人重识别问题. Yang 等^[22]提出一种非对称协同框架, 充分利用了在聚类算法中通常被舍弃的离群点来监督网络的训练. Ge 等^[23]专注于聚类伪标签噪声的影响提出同步平均教学网络模型 (Mutual mean teaching, MMT), 利用两个相同的网络相互监督训练. Wang 等^[24]在 MMT 的基础上提出将注意力波动模块 (Attentive waveblock, AWB) 集成到两个网络中, 使得两个网络学习到具有差异性的特征, 增强网络的互补性. 基于聚类的方法往往与聚类算法的准确率有大的关系, 且明显缺陷是模型的准确性过度依赖于聚类产生的伪标签的质量, 而且随着网络的迭代, 伪标签的噪声会越来越大.

受 MMT 模型的启发, 本文在基于聚类算法的跨域行人重识别基础上, 针对聚类算法生成的伪标签带有噪声的问题, 设计自适应融合网络模型. 该模型采用双网络结构共同学习并对双网络结构进行网络融合, 利用融合后的网络监督双网络结构训练, 有效解决了单网络情况下伪标签噪声不断扩大的问题. 同时, 为了降低模型对相机变换的敏感性, 本文设计了细粒度的风格转换模块. 细粒度的风格转换模块区别于以往的利用 GAN 进行数据集扩充的方法, 是在细粒度级别上针对于在同一个数据集上的不同相机风格进行图像生成, 而不是在源域和目标域之间作图像生成, 通过 starGAN 给每张行人图像都生成所有相机风格下的生成图像, 在扩充数据集的同时, 增强了模型的鲁棒性. 相比于联合判别生成学习网络 (Joint discriminative and generative learning network, DGNet)^[27]以身份标签为基础的风格转换模式, 本文方法专注于相机域间的风格转换, 对每个相机下的图像都生成其他相机风格的生成图像, 增强了模型对相机的鲁棒性.

2 基于自适应融合网络的跨域行人重识别方法

在众多的跨域行人重识别方法中, 基于聚类的方法有着强大的竞争力, 由于忽略了对聚类伪标签噪声的处理而导致噪声随着网络迭代而不断扩大.

本文针对于此, 主要研究伪标签噪声在网络训练中扩大的问题, 并提出基于自适应融合网络的行人重识别, 并利用细粒度风格转换降低不同相机间视角下的风格差异性, 提高模型的适应性. 本节将详细介绍提出的方法.

2.1 问题定义

给定一个带有真实标签的源域数据集 $S = \{X_s, Y_s\}$, 其中 X_s 代表源域的图像, Y_s 代表图像 X_s 对应的真实标签, 源域数据集 S 中包含 N_s 张图像, N_s 张图像共具有 I_s 个身份标签, 源域中每个行人图像 $x_{s,i} \in X_s (i = 1, 2, \dots, N_s)$ 具有唯一的身份标签 $y_{s,i} \in Y_s$.

给定一个无标签的目标域数据集 $T = \{X_t\}$, 其中 X_t 代表目标域的图像, 数量为 N_t , 每张行人图像 $x_{t,i} \in X_t$ 都不具有标签信息. 本文利用自适应融合网络模块来学习跨域的行人重识别, 以便将有标签的源域学习到的知识迁移到无标签的目标域上利用. 本文模型见图 1. 首先, 对目标域训练集下的图像进行全相机细粒度风格转换得到新的目标域; 然后, 对模型采用双网络结构进行训练, 并通过自适应的融合策略将两个网络学习到的知识进行结合得到融合网络, 融合网络的分类结果作为双网络结构的监督信号指导模型训练.

2.2 源域预训练

为将源域学习到的知识更好地迁移到目标域上, 本文利用源域数据集对模型进行预训练, 并将预训练的模型参数作为训练目标域数据集的初始化参数.

对于参数为 θ 的深度神经网络模型 M_s , 首先在源域数据集上对模型进行有监督地预训练. 源域中的每张行人图像 $x_{s,i} \in X_s$ 通过模型 M_s 提取出特征 $f(x_{s,i} | \theta)$, 并最终输出对图像的身份预测 $p(x_{s,i} | \theta)$. 本文采用交叉熵损失和三元组损失^[28]来优化源域预训练模型. 交叉熵损失定义为:

$$L_{s,id} = -\frac{1}{N_s} \sum_{i=1}^{N_s} p(y_{s,i} | x_{s,i}^\theta) \quad (1)$$

式中, $p(y_{s,i} | x_{s,i}^\theta)$ 是图像 $x_{s,i}$ 在模型参数为 θ 的情况下预测属于 $y_{s,i}$ 的概率. 三元组损失定义为:

$$L_{s,tri} = \frac{1}{N_s} \times \sum_{i=1}^{N_s} \ln \left(m + \|f_s^+(x_{s,i}^\theta)\|_2 - \|f_s^-(x_{s,i}^\theta)\|_2 \right) \quad (2)$$

式中, $f_s^+(x_{s,i}^\theta) = f(x_{s,i}^\theta) - f(x_{s,i+}^\theta)$, $f_s^-(x_{s,i}^\theta) =$

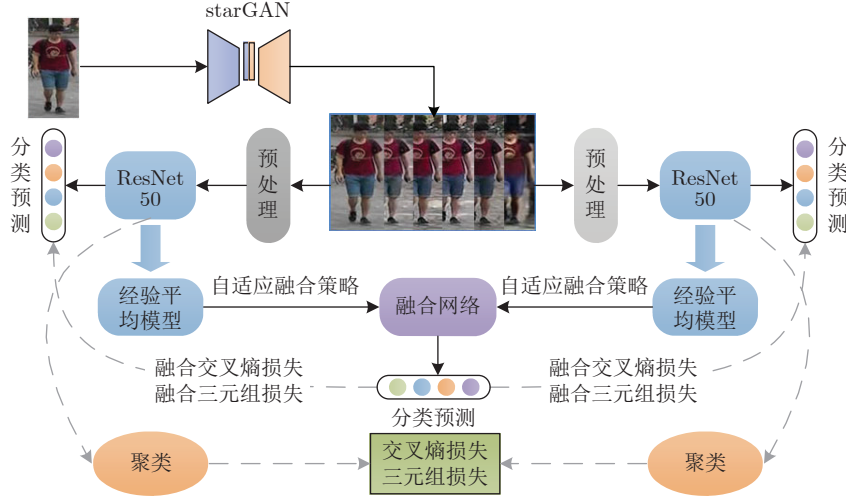


图1 自适应融合网络模型

Fig.1 Adaptive fusion network model

$f(x_{s,i}^\theta) - f(x_{s,i-}^\theta)$, $x_{s,i+}^\theta$ 和 $x_{s,i-}^\theta$ 分别是基准样本 $x_{s,i}^\theta$ 的正样本和负样本, $\|\cdot\|_2$ 代表欧氏距离, m 是边距参数.

2.3 目标域聚类

为了解决目标域数据集缺少真实标签的问题, 本文利用 Mini-Batch k -means 聚类算法产生伪标签用于训练. 首先, 目标域图像经过网络提取得到特征 $f(x_{t,i} | \theta)$. 其次, 对特征 $f(x_{t,i} | \theta)$ 进行 Mini-Batch k -means 聚类算法得到 I_t 个类别, 同一个类别内的图像具有相同的伪标签 y_t .

为了充分利用聚类得到的伪标签进行模型的提高, 本文使用了交叉熵损失和三元组损失. 模型的交叉熵损失定义为:

$$L_{id} = -\frac{1}{N_t} \sum_{k=1}^2 \sum_{i=1}^{N_t} p(y_{t,i} | x_{t,i}^{\theta_k}) \quad (3)$$

式中, k 表示双网络结构中的两个网络, 取值为 1 或 2. $p(y_{t,i} | x_{t,i}^{\theta_k})$ 是图像 $x_{t,i}$ 在模型参数为 θ_k 的情况下预测属于 $y_{t,i}$ 的概率. 模型的三元组损失定义为:

$$L_{tri} = \frac{1}{N_t} \times \sum_{i=1}^{N_t} \ln \left(m + \|f_t^+(x_{t,i}^{\theta_k})\|_2 - \|f_t^-(x_{t,i}^{\theta_k})\|_2 \right) \quad (4)$$

式中, $f_t^+(x_{t,i}^{\theta_k}) = f(x_{t,i}^{\theta_k}) - f(x_{t,i+}^{\theta_k})$, $f_t^-(x_{t,i}^{\theta_k}) = f(x_{t,i}^{\theta_k}) - f(x_{t,i-}^{\theta_k})$.

2.4 双网络结构

基于聚类的跨域行人重识别方法的效果往往与

聚类算法的精度成正相关, 为了增强网络对聚类伪标签的抗噪能力, 本文模型采用双网络结构进行表征学习. 双网络结构采用两个相同的 ResNet50 网络作为主干网络进行目标域的训练, 但对两个网络采用不同的预训练初始化参数. 同时, 为了使两个网络能够学习到区别于彼此的特征, 对输入到两个网络中的目标域图像进行不同的预处理, 包括随机翻转、擦除、裁剪处理.

双网络结构的设计专注于对抗伪标签噪声问题, 防止出现单网络结构下噪声随着迭代不断扩大的情况. 但是, 随着训练的进行, 两个网络学习能力会逐渐靠近, 可能会出现两个网络收敛到彼此相等的位置, 这违背了双网络结构的设计初衷. 对此, 为了保有双网络每一次知识学习的经验, 本文采用经验平均模型代替两个 ResNet50 网络进行融合. 经验平均模型的参数是对应的双网络结构中网络参数的加权平均, 而不是通过反向传播得到的, 所以不会明显增大网络的计算量. 对于在第 T 次迭代下网络 k 的经验平均模型定义为 $H_T(\theta_k)$, 经验平均模型对网络每一次迭代的学习能力进行保留, 并通过式 (5) 进行更新:

$$H_T(\theta_k) = \alpha H_{T-1}(\theta_k) + (1 - \alpha)\theta_k, \quad k = 1, 2 \quad (5)$$

式中, $\alpha \in [0, 1]$ 是一个动量更新因子, $H_{T-1}(\theta_k)$ 是在 $(T-1)$ 次迭代时网络的经验平均模型. 当 $T = 0$ 时, $H_0(\theta_k) = \theta_k$.

2.5 自适应融合网络

为了使双网络结构的训练获得更有效的监督, 本文选择对双网络进行知识融合. 融合后的网络包含两个网络学习到的知识, 相比于单网络结构的学

习, 双网络在学习过程中进行了相互补充, 并且利用融合网络的输出去监督两个网络的训练, 可以有效避免噪声随着迭代的进行而扩大. 虽然两个网络具有相同的结构, 但是学习知识的能力是不同的, 所以在每一次融合的时候不能单纯地平等对待两个网络. 对此, 提出一种自适应融合策略用于解决两个网络的融合权重问题. 自适应融合策略是基于类内离散度和类间离散度计算得到的, 类内离散度表示每个图像样本特征和所属类别平均特征的差距, 类间离散度表示的是每个类别的平均特征和所有样本的平均特征的差距.

具体地, 通过聚类算法将所有目标域样本聚类到 I_t 个类别中, 并用 C 来表示每个聚类, 那么第 k 个网络中的第 i 个聚类 C_i 的类内离散度被定义为:

$$S_{\text{intra},k}^i = \sum_{x \in C_i} \|f(x_t | \theta_k) - \mu_{i,k}\|_2^2, \quad k = 1, 2 \quad (6)$$

式中, $\mu_{i,k}$ 是第 i 个聚类 C_i 中所有图像在第 k 个网络上的平均特征. 第 k 个网络中的第 i 个聚类 C_i 的类间离散度被定义为:

$$S_{\text{inter},k} = \sum_{i=1}^{I_t} n_{t,i} \|\mu_{i,k} - \mu_k\|_2^2, \quad k = 1, 2 \quad (7)$$

式中, μ_k 是目标域中所有训练样本在第 k 个网络上的平均特征, $n_{t,i}$ 是目标域中所有训练样本的数量. 利用类内离散度和类间离散度得出第 k 个网络的自适应平衡因子:

$$J_k = \frac{S_{\text{inter},k}}{\sum_{i=1}^{I_t} S_{\text{intra},k}^i}, \quad k = 1, 2 \quad (8)$$

自适应平衡因子 J_k 量化了网络的学习能力, J_k 越大代表网络的学习能力越强. 当类间离散度变大或者类内离散度变小的时候, J_k 也会相应地变大. 通过自适应平衡因子可以计算双网络结构的自适应融合权重:

$$w_k = \frac{J_k}{J_1 + J_2}, \quad k = 1, 2 \quad (9)$$

式中, w_k 代表第 k 个网络的融合权重.

在每次迭代获得两个网络的经验平均模型之后, 通过融合权重比例对双网络进行融合. 由于融合网络只用于双网络结构的监督, 所以不会影响两个网络独立地更新. 参数为 θ_f 自适应融合网络定义为:

$$H_T(\theta_f) = \sum_{k=1}^2 w_k H_T(\theta_k) \quad (10)$$

由于融合网络采用的是网络对行人身份的分类预测, 而不是对聚类算法产生的伪标签进行训练,

所以设计了融合交叉熵损失和融合三元组损失来优化双网络结构模型. 融合交叉熵损失利用了融合网络的分类预测以及每个网络的分类预测, 其定义为:

$$L_{\text{uid}} = -\frac{1}{N_t} \times \sum_{k=1}^2 \sum_{i=1}^{N_t} p_j(x_{t,i} | H_T(\theta_f)) \ln p_j(x_{t,i} | H_T(\theta_k)) \quad (11)$$

融合三元组损失是在三元组损失的基础上结合分类预测得到:

$$L_{\text{utri}} = -\frac{1}{N_t} \sum_{k=1}^2 \sum_{i=1}^{N_t} (F_i(\theta_f) \ln F_i(\theta_k) + (1 - F_i(\theta_f)) \ln (1 - F_i(\theta_k))) \quad (12)$$

式中

$$F_i(\theta_k) = m + \left\| f(x_{s,i}^{\theta_k}) - f(x_{s,i+}^{\theta_k}) \right\|_2 - \left\| f(x_{s,i}^{\theta_k}) - f(x_{s,i-}^{\theta_k}) \right\|_2, \quad k = 1, 2 \quad (13)$$

$$F_i(\theta_f) = m + \left\| f(x_{s,i}^{\theta_f}) - f(x_{s,i+}^{\theta_f}) \right\|_2 - \left\| f(x_{s,i}^{\theta_f}) - f(x_{s,i-}^{\theta_f}) \right\|_2, \quad k = 1, 2 \quad (14)$$

最后结合基于伪标签的交叉熵损失和三元组损失以及基于融合网络的融合交叉熵损失和融合三元组损失定义模型的整体损失:

$$L = (1 - \lambda_{id}) L_{\text{uid}} + \lambda_{id} L_{id} + (1 - \lambda_{tri}) L_{\text{utri}} + \lambda_{tri} L_{tri} \quad (15)$$

式中, λ_{id} 和 λ_{tri} 是权重参数, 用来平衡不同损失之间的影响.

2.6 细粒度风格转换

虽然自适应融合网络可以有效减少聚类伪标签噪声的影响, 但是由于行人重识别数据集的行人图像由多个不同视角的相机捕捉而成, 相机风格的变换使得即使是同一身份的行人也难以被分辨, 造成网络对相机具有一定的敏感性. 受 PTGAN^[12] 启发, 本文利用生成对抗网络降低模型对相机的敏感度. 不同于 PTGAN 在源域和目标域间进行风格转换, 本文的风格转换模块是在目标域相机间进行的细粒度级别的风格转换.

如图 2 所示, 细粒度风格转换模块将行人图像在广义上分割成行人因子和风格因子, 行人因子包含图像中行人部分, 风格因子包含除行人部分外的其他部分. 其中每个相机下的风格因子假定是相同的, 利用 starGAN 网络为每个相机训练一个生成因子的生成器. 最后, 通过保留每张图像的行人因

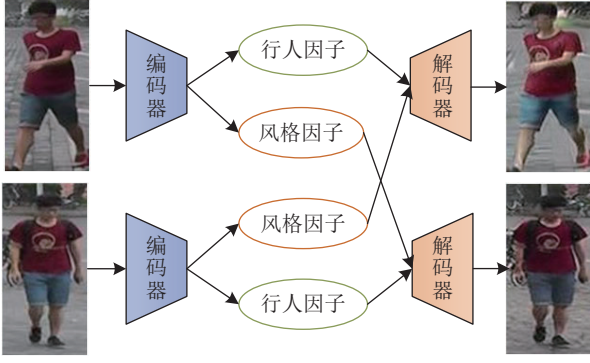


图2 细粒度风格转换模块

Fig.2 Fine-grained style conversion module

子, 而替换图像的风格因子达到行人图像风格转换的目的. 对于存在 E 个不同视角的相机, 图像 $x_{t,i} \in X_t$ 属于相机 A , 那么需要生成相机 $(E - A)$ 风格的转换图像. 最终, 将生成后的图像加入到目标域数据集一起训练.

2.7 总的算法流程

本文提出的自适应融合网络的跨域行人重识别方法总的算法流程见算法 1.

算法 1. 自适应融合网络

输入. 源域数据集 $S = \{X_s, Y_s\}$, 目标域数据集 $T = \{X_t\}$; 动量更新因子 α , 权重参数 λ_{id} 和 λ_{tri} ; 聚类数量 K , 学习率 lr ; 原始模型 $\varphi(\theta_o)$.

输出. 自适应融合网络模型 $\varphi(\theta_t)$.

- 1) 通过源域数据集 S 对原始模型进行预训练得到初始化参数 θ_s .
- 2) 利用 starGAN 对目标域数据集进行细粒度风格转换得到新的目标域数据集.
- 3) 利用两个网络分别对新的目标域数据集进行特征提取得到图像特征 $f(x_{t,i}^{\theta_1})$ 和 $f(x_{t,i}^{\theta_2})$.
- 4) 利用 Mini-Batch k -means 聚类算法为特征 $f(x_{t,i}^{\theta_1})$ 和 $f(x_{t,i}^{\theta_2})$ 生成伪标签 y_1 和 y_2 .
- 5) 基于伪标签, 利用式 (3) 和式 (4) 分别计算交叉熵损失和三元组损失.
- 6) 利用式 (9) 计算两个网络的自适应融合权重 w_1 和 w_2 .
- 7) 基于融合权重, 利用式 (10) 对两个网络的经验平均模型进行融合.
- 8) 基于融合网络, 利用式 (11) 和式 (12) 分别计算融合交叉熵损失和融合三元组损失.
- 9) 利用式 (15) 对模型进行优化.

3 实验结果域分析

为了验证方法的有效性, 本文在 Market1501^[29]、

DukeMTMC-ReID^[30] 和 MSMT17^[15] 三个行人重识别基准数据集上对本文方法进行评估, 包括与主流方法的对比、消融实验和参数分析.

3.1 数据集与评估指标

Market1501^[29] 数据集包含取自 6 个不同相机视角的 1501 个身份的总共 32 668 张行人图像, 其中包含 751 个身份的 12 936 张图像用于训练集, 另外包含 750 个身份的 19 732 张图像用于测试集. 在测试集中又分为包含 19 732 张的 Gallery 集和 3 368 张的 Query 集. 这些图像通过可变形部分模型^[31] 进行身份检测.

DukeMTMC-ReID^[30] 数据集是 DukeMTMC^[32] 的子集, 包含取自 8 个摄像机视角的 1 812 个身份图像. 其中 16 552 用作训练集, 17 661 用作 Gallery 集, 2 228 用作 Query 集. 在 1 812 个身份中, 1 404 个出现在至少 2 个摄像头, 其余的出现在 1 个摄像头中.

MSMT17^[15] 数据集是目前行人重识别领域最大的数据集. 由 12 台室外摄像机和 3 台室内摄像机拍摄的 126 441 张照片组成. 这些图像代表 4 101 个身份, 并按照 1:3 的比例随机分为训练集和测试集. 训练集包含 1 041 个身份, 共 32 621 张图像, 而测试集包含 3 060 个身份, 共 93 820 张图像. 对于测试集, 随机选择 11 659 张图像作为 Query 集, 而其他 82 161 张图像作为 Gallery 集.

本文实验使用平均精度均值 (Mean average precision, mAP) 和 Rank- n 准确率对本文中涉及到的行人重识别模型性能进行量化评价. 其中, mAP 将所有类别的平均精度进行综合加权平均而得到的; Rank- n 是检索结果中前 n 位候选的准确率, 本文主要选择 Rank-1、Rank-5 和 Rank-10 进行评估.

3.2 实验设置

本文模型的训练包含源域的预训练和目标域的跨域自适应两部分. 在图像输入到网络之前, 将图像的大小调整为 256×128 . 根据经验将边距参数 m 设置为 0.5.

本文实验基于 Pytorch 框架, 使用 Pytorch1.1 版本, 使用 2 个 GTX-2080TI GPU 进行训练, 1 个 GTX-2080TI GPU 进行测试. 采用自适应矩估计优化器对网络进行优化, 权值衰减为 0.0005.

在源域的预训练阶段, 本文使用在 ImageNet^[33] 上预训练过的 ResNet-50 作为主干网络. 初始学习率设置为 0.00035, 总共进行 80 次迭代, 并在第 40 次和 70 次的迭代时, 将初始学习率减少到了原来的 1/10.

在目标域的跨域自适应阶段, 利用在源域预训练中得到的预训练权重作为网络的初始权重. 总共进行 80 次迭代, 学习率固定设置为 0.0001, 并将式 (5) 中的动量更新因子 α 设置为 0.999, 式 (15) 中的权重参数 λ_{id} 设置为 0.6, λ_{tri} 设置为 0.8. 分别将作为目标域的 Market1501、DukeMTMC-ReID 和 MSMT17 实验中 Mini-Batch k -means 聚类算法的聚类个数设置为 500、700 和 1500.

本文模型的参数量如表 1 所示. 在计算单网络参数量的情况下乘 2 得到模型整体的参数量.

表 1 本文的自适应融合网络模型参数量表
Table 1 The model parameter number of the proposed adaptive fusion network

参数	取值
总参数	23512128×2
可训练参数	23512128×2
参数大小 (MB)	89.69×2
估计总大小 (MB)	1199.45×2

3.3 实验结果

3.3.1 与主流行人重识别方法比较

本节将本文提出的算法与当前主流的行人重识别算法进行比较. 比较方法包括: 1) 无监督方法. 自底向上聚类 (Bottom-up clustering, BUC)^[34] 和软化的相似性学习 (Softened similarity learning, SSL)^[35]; 2) 无监督跨域的方法. 多任务中层特征对齐网络 (Multi-task mid-level feature alignment, MMFA)^[36]、可迁移联合属性-身份深度学习 (Transferable joint attribute-identity deep learning, TJ-AIDL)^[11]、基于差异的最大平均差异损失 (Dissimilarity-based maximum mean discrepancy loss, D-MMD)^[12]、三重对抗学习和多视角想象推理网络 (Triple adversarial learning and multi-view imaginative reasoning network, TAL-MIRN)^[37] (基于域分布对齐的方法); 自适应迁移网络 (Adaptive transfer network, ATNet)^[13]、相似性保持生成对抗网络 + 局部最大池化 (Similarity preserving generative adversarial network + local max pooling, SPGAN + LMP)^[14]、异构-同构学习 (Hetero-homogeneous learning, HHL)^[16] (基于 GAN 的方法); 范例相机近邻不变性 (Exemplar-invariance, camera-invariance and neighborhood-invariance, ECN)^[17]、多标签参考学习 (Multilabel reference learning, MAR)^[18] (基于特征相似性计算的方法); 无监督领域自适应行人重识别 (Unsupervised domain

adaptive person re-identification, UDAP)^[38]、带有渐进式增强框架的基于部分的卷积基线 (Part-based convolutional baseline-progressive augmentation framework, PCB-PAST)^[39]、自相似性分组 (Self-similarity grouping, SSG)^[20]、增广判别聚类 (Augmented discriminative clustering, AD-Cluster)^[21]、同步平均教学框架 (Mutual mean-teaching framework, MMT)^[23]、多专家头脑风暴网络 (Multiple expert brainstorming network, MEB-Net)^[40]、软迭代标签聚类 (Soft iterative label clustering, SILC)^[41]、双流互反解纠缠学习 (Dual-stream reciprocal disentanglement learning, DR-DL)^[42]、基于渐进式表征增强的自训练 (Self-training with progressive representation enhancement, PREST)^[43]、具有混合记忆的自步对比学习框架 (Self-paced contrastive learning with hybrid memory, SpCL)^[44]、多损失优化学习 (Multi-loss optimization learning, MLOL)^[45]、不确定性引导的噪声回弹网络 (Uncertainty-guided noise resilient network, UNRN)^[46] (基于聚类的方法). “本文方法 + 不确定性”是在文献 [46] 和文献 [47] 的启发下, 利用 Kullback-Leibler (KL) 散度计算双网络结构中两个网络的不确定性, 然后利用不确定性来约束损失函数的计算. 所有对比的方法结果是从源论文中获得的, 无监督方法没有用到源域数据集, 仅在目标域数据集上进行训练.

表 2 展示了在 Market1501 和 DukeMTMC-ReID 数据集上的实验结果. 由表 2 可以看出, 在实验准确率上无监督跨域方法普遍优于无监督方法, 其中采用聚类算法的无监督跨域行人重识别相比于其他 3 个无监督跨域行人重识别算法在整体上可达到最佳效果. 如表 1 所示, 当以 DukeMTMC-ReID 作源域, Market1501 作目标域时, 本文方法的 mAP 达到了 79.1%, Rank-1 达到了 91.8%. 当以 Market1501 作源域, DukeMTMC-ReID 作目标域时, 本文方法的 mAP 达到了 68.5%, Rank-1 达到了 81.7%. 这是由于本文的融合网络模型很好地抑制了伪标签噪声, 所以可以更有效地利用伪标签去训练. 同时, 采用基于细粒度的风格转换模块可以在克服相机敏感性的问题上扩充数据集, 提升了模型的识别能力. 当模型在双网络结构知识互补的基础上对损失加入不确定性的差异约束使得模型准确率得到了进一步的提升, 在以 Market1501 作目标域和 DukeMTMC-ReID 作目标域时, “本文方法 + 不确定性”的 mAP 分别达到了 79.9% 和 69.8%, Rank-1 分别达到了 92.3% 和 82.1%.

与 SpCL 方法相比, 本文方法在 Market-to-

表 2 在 Market1501 和 DukeMTMC-ReID 上与主流方法比较 (%)
Table 2 Comparison with the state-of-the-art methods on Market1501 and DukeMTMC-ReID (%)

方法	Duke-to-Market				Market-to-Duke			
	mAP	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10
BUC ^[34]	38.3	66.2	79.6	84.5	27.5	47.4	62.6	68.4
SSL ^[35]	37.8	71.7	87.4	37.8	28.6	52.5	63.5	68.9
MMFA ^[36]	27.4	56.7	75.0	81.8	24.7	45.3	59.8	66.3
TJ-AIDL ^[11]	26.5	58.2	74.8	81.1	23.0	44.3	59.6	65.0
D-MMD ^[12]	75.1	89.5	95.6	97.1	62.7	79.3	89.3	92.0
TAL-MIRN ^[37]	40.0	73.1	86.3	—	41.3	63.5	76.7	—
ATNet ^[13]	25.6	55.7	73.2	79.4	24.9	45.1	59.5	64.2
SPGAN + LMP ^[14]	26.7	57.7	75.8	82.4	26.2	46.4	62.3	68.0
HHL ^[16]	31.4	62.2	78.8	84.0	27.2	46.9	61.0	66.7
ECN ^[17]	43.0	75.1	87.6	91.6	40.4	63.3	75.8	80.4
MAR ^[18]	67.7	81.9	87.3	40.0	67.1	79.8	84.2	48.0
UDAP ^[38]	53.7	75.8	89.5	93.2	49.0	68.4	80.1	83.5
PCB-PAST ^[39]	54.6	78.4	—	—	54.3	72.4	—	—
SSG ^[20]	58.3	80.0	90.0	92.4	53.4	73.0	80.6	83.2
AD-Cluster ^[21]	68.3	86.7	94.4	96.5	54.1	72.6	82.5	85.5
MMT-500 ^[23]	71.2	87.7	94.9	96.9	63.1	76.8	88.0	92.2
MEB-Net ^[40]	76.0	89.9	96.0	97.5	66.1	79.6	88.3	92.2
SILC ^[41]	61.8	80.7	90.1	93.0	50.3	68.5	80.2	85.4
DRDL ^[42]	42.7	76.8	88.5	91.6	43.2	65.3	76.9	82.2
PREST ^[43]	62.4	82.5	92.1	94.9	56.1	74.4	83.7	85.9
SpCL ^[44]	76.7	90.3	96.2	97.7	68.8	82.9	90.1	92.5
MLOL ^[45]	70.9	86.6	93.1	95.1	69.8	83.1	90.8	93.0
UNRN ^[46]	78.1	91.9	96.1	97.8	69.1	82.0	90.7	93.5
本文方法	79.1	91.8	97.1	98.2	68.5	80.7	90.1	92.6
本文方法 + 不确定性	79.9	92.3	97.4	98.3	69.8	82.1	90.5	93.1

Duke 上的准确率略低, 这是由于 SpCL 对源域和目标域上的所有可用信息进行编码以学习特征, 但是这同样会使其域适应性能力降低. 在双网络结构的基础上利用不确定性约束损失函数的“本文方法 + 不确定性”中, 除了 Rank-1 之外都有所提升, 且 mAP 比 SpCL 高 1%, 说明了利用不确定性对损失函数进行约束可以有效降低伪标签噪声.

MLOL 方法在 Market-to-Duke 上实验表现较好, 然而在其他实验中的识别准确率都远低于本文方法. 如表 3 所示, SpCL 和 MLOL 在 MSMT17 数据集上的低准确率侧面印证了两者的局限性.

为进一步验证本文方法的有效性, 在更大更接近现实场景的 MSMT17 数据集上进行了实验. 表 3 为在 MSMT17 数据集上的实验结果. 当 DukeMTMC-ReID 数据集作为源域时, 本文方法的 mAP 达到 30.2%, Rank-1 达到 60.4%; 当 Market1501 数据集作为源域时, mAP 达到 29.4%, Rank-1 达到

59.6%. 在具有挑战性的大型数据集 MSMT17 上的高性能表现进一步证明了本文方法的有效性, 而且通过不确定性对损失函数进行约束同样在 MSMT17 的数据集上也显示了它的有效性, Duke-to-MSMT17 和 Market-to-MSMT17 的各项指标都有所提升.

3.3.2 消融实验

为了验证双网络结构、经验平均模型、自适应融合策略以及风格转换 4 个模块的有效性, 在 Market1501 和 DukeMTMC-reID 数据集上进行了消融实验, 实验结果如表 4 所示. 直接转换表示目标域数据集直接在源域的预训练模型上进行测试; 基线表示仅使用聚类方法的伪标签部分进行训练. F 为双网络结构且没有采用自适应融合策略, T 为经验平均模型, A 为自适应融合策略, S 为细粒度风格转换模块, w 为添加上述模块.

为了促进网络的融合, 先利用平均融合代替自

表 3 在 MSMT17 上与主流方法比较 (%)
Table 3 Comparison with the state-of-the-art methods on MSMT17 (%)

方法	Duke-to-MSMT17				Market-to-MSMT17			
	mAP	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10
ECN ^[17]	10.2	30.2	41.5	46.8	8.5	25.3	36.3	42.1
SSG ^[20]	13.3	32.2	—	51.2	13.2	31.6	—	49.6
MMT-1500 ^[23]	23.3	50.1	63.9	69.8	22.9	49.2	63.1	68.8
SILC ^[41]	12.6	33.1	45.2	48.0	10.9	27.8	38.1	45.8
TAL-MIRN ^[37]	14.2	39.0	51.5	—	11.2	30.9	43.5	—
DRDL ^[42]	14.9	42.0	53.7	59.1	14.7	38.6	51.4	57.1
PREST ^[43]	18.5	43.8	57.5	63.6	15.9	37.8	51.8	57.8
SpCL ^[44]	26.5	53.1	65.8	70.5	25.4	51.6	64.3	69.7
MLOL ^[45]	22.4	48.3	60.7	66.1	21.7	46.9	59.4	64.7
UNRN ^[46]	26.2	54.9	67.3	70.6	25.3	52.4	64.7	69.7
本文方法	30.2	60.4	73.3	77.9	29.4	59.6	72.8	77.5
本文方法 + 不确定	30.8	61.0	73.9	78.3	30.6	61.0	73.7	78.0

表 4 在 Market1501 和 DukeMTMC-ReID 上的消融实验 (%)
Table 4 Ablation experiments on Market1501 and DukeMTMC-ReID (%)

方法	Duke-to-Market				Market-to-Duke			
	mAP	Rank-1	Rank-5	Rank-10	mAP	Rank-1	Rank-5	Rank-10
直接转换	31.8	61.9	76.4	82.2	29.9	46.2	61.9	68.0
基线	53.5	76.0	88.1	91.9	48.2	66.4	79.8	84.0
本文方法 w/F	74.3	90.2	95.8	97.6	62.9	77.1	87.9	91.5
本文方法 w/(F + T)	77.6	91.5	96.8	98.1	66.3	79.0	89.6	92.3
本文方法 w/(F + T + A)	78.2	91.7	96.9	98.1	66.9	79.9	89.7	92.2
本文方法 w/(F + T + S)	78.9	91.2	96.8	98.0	67.5	80.3	89.9	92.4
本文方法 w/(F + T + A + S)	79.1	91.8	97.1	98.2	68.5	80.7	90.1	92.6

适应融合, 进行“F + A”的实验时再采用自适应融合策略. 另外, 细粒度风格转换模块只应用在目标域数据集上, 而不适用于源域数据集的预训练. 消融实验证明了提出的双网络结构、经验平均模型、自适应融合策略以及风格转换每一个模块单独的有效性及其它们之间相互组合的有效性. 采用双网络结构使模型准确率相比于“基线”得到大幅度的提升, 之后以双网络结构为基础分别加入自适应融合模块、经验平均模型和细粒度风格转换模块都使模型准确率进一步得到提高. 最终, 将 4 个模块一起使用的模型达到了最优的效果, mAP 为 79.1%, Rank-1 为 91.8%. 值得注意的是, 本文的风格转换模块可以作为一个即插即用的模块, 用于增强模型对相机风格变换的鲁棒性, 对于同一数据集只需要进行一次风格转换即可在后续的实验中多次应用.

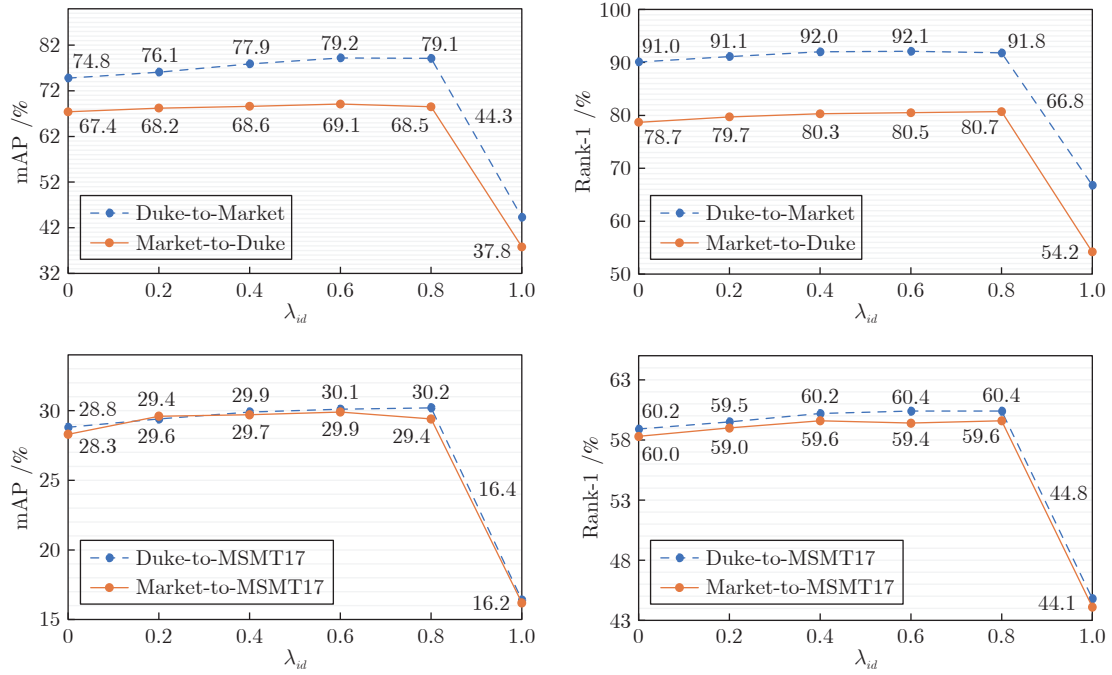
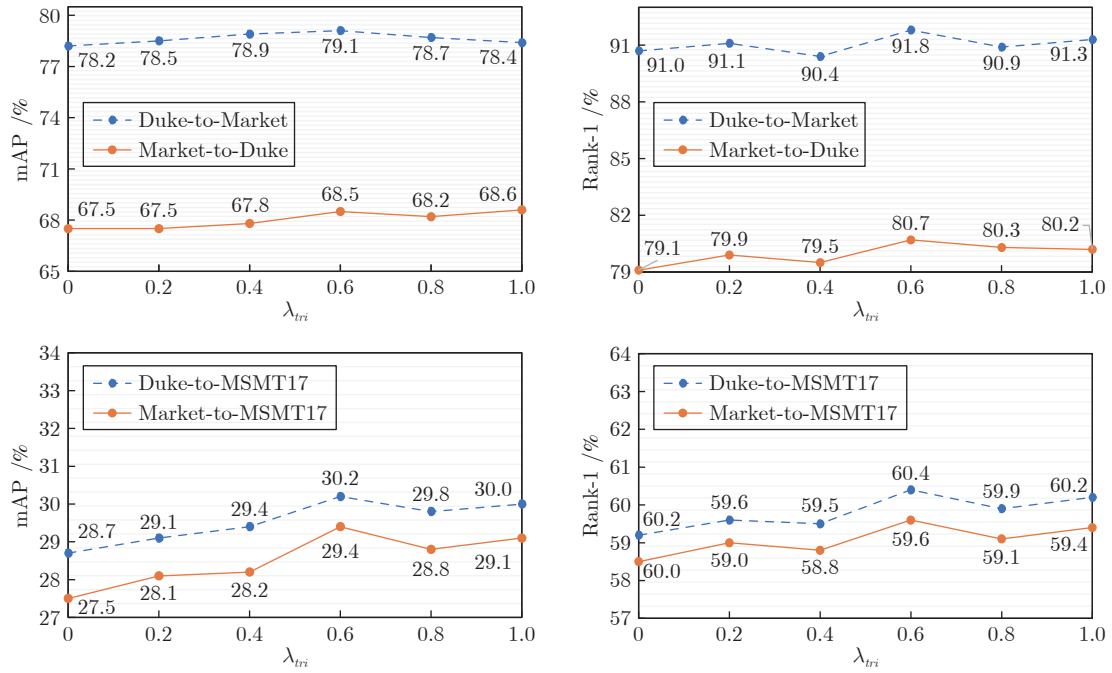
3.3.3 参数分析

本节分析了损失函数中平衡交叉熵损失和融合交叉熵损失的超参数 λ_{id} 、平衡三元组损失和融合三

元组损失的超参数 λ_{tri} 、Mini-Batch k -means 聚类算法应用在不同数据集的超参数聚类数量. 默认情况下, 改变一个超参数的同时另外两个超参数固定不变.

图 3 比较了式 (15) 中不同的 λ_{id} 取值对实验结果的影响. 当 $\lambda_{id} = 0$ 时, 表示本文方法只应用融合交叉熵损失; 当 $\lambda_{id} = 1$ 时, 表示本文方法只应用交叉熵损失. 可以看出, λ_{id} 取值为 0 或 1 的实验结果都没有取中间值好, 这表明同时交叉熵损失和融合交叉熵损失的必要性, 也证明融合网络模型的有效性. 当 $\lambda_{id} = 0.8$ 时, 本文模型达到了最高准确率.

图 4 比较了式 (15) 中不同的 λ_{tri} 取值对实验结果的影响. 当 $\lambda_{tri} = 0$ 时, 表示只应用融合三元组损失; 当 $\lambda_{tri} = 1$ 时, 表示只应用三元组损失. 当 $\lambda_{tri} = 0.6$ 时, 本文模型达到了最高的准确率. 在图 5 中, 比较了不同的聚类数量取值对实验结果的影响. 当进行 Duke-to-Market 和 Market-to-Duke 的实验时, 聚类数量为 500 和 700 时模型效果最好. 当进行 Duketo-MSMT17 和 Market-to-MSMT17 的实

图 3 λ_{id} 取值评估实验Fig.3 Evaluation of different values of λ_{id} 图 4 λ_{tri} 取值评估实验Fig.4 Evaluation of different values of λ_{tri}

验时, 聚类数量为 1500 时模型效果最好.

3.3.4 聚类算法对比

在基于聚类的跨域行人重识别方法中, 聚类算法的效果与实验结果的准确率成正比关系, 即聚类

算法效果越好, 实验结果准确率越高. 聚类算法效果越好, 聚类伪标签所带有的噪声就越少, 模型训练就越不容易产生偏差. 为了探究不同聚类算法的应用对本文模型的影响, 在本节对 Mini-Batch k -

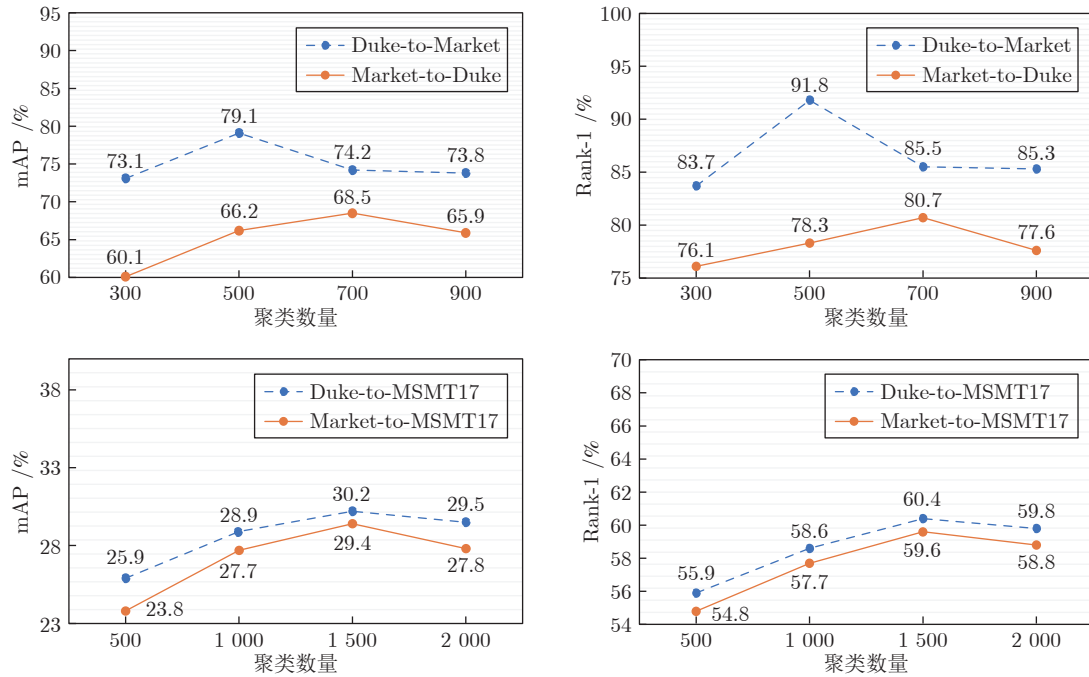


图 5 聚类数量取值评估实验
Fig.5 Evaluation of different numbers of clustering

means、 k -means 和噪声环境下基于密度的空间聚类研究 (Density-based spatial clustering of applications with noise, DBSCAN) 聚类算法进行了对比, 实验结果如表 5 所示. 为了对比模型应用不同聚类算法时在运算效率上的差异, 表 5 展示了模型采用 Mini-Batch k -means、 k -means 和 DBSCAN 聚类算法时每次迭代的运算时间 (s). 当使用 Mini-Batch k -means 或 k -means 聚类算法时, 两者的实验准确率相差不大, k -means 效果略好于 Mini-Batch k -means, 但 k -means 所需运行时间高于 Mini-Batch k -means. 当使用 DBSCAN 聚类算法时, 在 Duke-to-Market 实验时 mAP 达到了 80.1%, Rank-1 达到了 92.3%; 在 Market-to-Duke 实验时 mAP 达到了 69.9%, Rank-1 达到了 82.1%. 相比于使用 Mini-Batch k -means 或 k -means, 实验准确率有着明显的提升, 但是 DBSCAN 所需的运算时间也是远高于 Mini-Batch k -means 和 k -means 的. 出于对运算时间的考虑和准确率的综合

考虑, 本文模型采用 Mini-Batch k -means 进行实验.

4 结束语

目前基于聚类的跨域行人重识别方法忽略了聚类伪标签的噪声问题, 导致模型效果无法到达有监督的行人重识别方法水平. 本文提出基于细粒度风格转换的自适应融合网络方法, 采用双网络结构共同学习, 并对双网络进行自适应融合, 利用融合后的网络监督双网络训练. 同时, 为了解决数据集图像对相机敏感性的问题, 对目标域图像进行细粒度的风格转换, 将转换后的扩充目标域数据集应用于训练. 与现有的基于聚类的跨域方法相比, 本文模型取得了更高的准确率. 在三个行人重识别基准数据集 Market1501、DukeMTMC-ReID、MSMT17 上的实验结果充分验证了本文方法的有效性. 未来研究工作将考虑进行多个不同网络结构的融合, 以便使不同网络可以学习到彼此更具有差异性的知识, 引导模型走出局部最优, 减少伪标签噪声的影响.

表 5 聚类算法对比
Table 5 Comparison of clustering algorithms

方法	Duke-to-Market					Market-to-Duke				
	mAP (%)	R-1 (%)	R-5 (%)	R-10 (%)	运行时间 (s)	mAP (%)	R-1 (%)	R-5 (%)	R-10 (%)	运行时间 (s)
Mini-Batch k -means	79.1	91.8	97.1	98.2	811	68.5	80.7	90.1	92.6	908
k -means	79.3	91.8	97.2	98.1	1472	68.8	80.9	90.1	92.6	1669
DBSCAN	80.1	92.3	97.4	98.4	3224	69.9	82.1	90.7	92.9	3643

References

- 1 Ye Yu, Wang Zheng, Liang Chao, Han Zhen, Chen Jun, Hu Rui-Min. A survey on multi-source person re-identification. *Acta Automatica Sinica*, 2020, **46**(9): 1869–1884 (叶钰, 王正, 梁超, 韩镇, 陈军, 胡瑞敏. 多源数据行人重识别研究综述. *自动化学报*, 2020, **46**(9): 1869–1884)
- 2 Ye M, Shen J B, Lin G J, Xiang T, Shao L, Hoi S C H. Deep learning for person re-identification: A survey and outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, **44**(6): 2872–2893
- 3 Li You-Jiao, Zhuo Li, Zhang Jing, Li Jia-Feng, Zhang Hui. A survey of person re-identification. *Acta Automatica Sinica*, 2018, **44**(9): 1554–1568 (李幼蛟, 卓力, 张菁, 李嘉峰, 张辉. 行人再识别技术综述. *自动化学报*, 2018, **44**(9): 1554–1568)
- 4 Bai S, Bai X, Tian Q. Scalable person re-identification on supervised manifold. In: Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (ICCV). Honolulu, USA: IEEE, 2017. 3356–3365
- 5 Luo Hao, Jiang Wei, Fan Xing, Zhang Si-Peng. A survey on deep learning based person re-identification. *Acta Automatica Sinica*, 2019, **45**(11): 2032–2049 (罗浩, 姜伟, 范星, 张思朋. 基于深度学习的行人重识别研究进展. *自动化学报*, 2019, **45**(11): 2032–2049)
- 6 Zhang Yun-Peng, Wang Hong-Yuan, Zhang Ji, Chen Li, Wu Lin-Yu, Gu Jia-Hui, et al. One-shot video-based person re-identification based on neighborhood center iteration strategy. *Journal of Software*, 2021, **32**(12): 4025–4035 (张云鹏, 王洪元, 张继, 陈莉, 吴琳钰, 顾嘉晖, 等. 近邻中心迭代策略的单标注视频行人重识别. *软件学报*, 2021, **32**(12): 4025–4035)
- 7 Liu Yi-Min, Jiang Jian-Guo, Qi Mei-Bin, Liu Hao, Zhou Hua-Jie. Video-based person re-identification method based on GAN and pose estimation. *Acta Automatica Sinica*, 2020, **46**(3): 576–584 (刘一敏, 蒋建国, 齐美彬, 刘皓, 周华捷. 融合生成对抗网络和姿态估计的视频行人再识别方法. *自动化学报*, 2020, **46**(3): 576–584)
- 8 Wang M L, Lai B S, Huang J Q, Gong X J, Hua X S. Camera-aware proxies for unsupervised person re-identification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, **35**(4): 2764–2772
- 9 Wu Y M, Wu X T, Li X, Tian J. MGH: Metadata guided hypergraph modeling for unsupervised person re-identification. In: Proceedings of the 29th ACM International Conference on Multimedia. Virtual Event China: 2021. 1571–1580
- 10 Chen H, Lagadec B, Bremond F. ICE: Inter-instance contrastive encoding for unsupervised person re-identification. In: Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE, 2021. 14940–14949
- 11 Wang J Y, Zhu X T, Gong S G, Li W. Transferable joint attribute-identity deep learning for unsupervised person re-identification. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 2275–2284
- 12 Mekhazni D, Bhuiyan A, Ekladious G, Granger E. Unsupervised domain adaptation in the dissimilarity space for person re-identification. In: Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: 2020. 159–174
- 13 Liu J W, Zha Z J, Chen D, Hong R C, Wang M. Adaptive transfer network for cross-domain person re-identification. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019. 7195–7204
- 14 Deng W J, Zheng L, Ye Q X, Kang Q L, Yi Y, Jiao J B. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 994–1003
- 15 Wei L H, Zhang S L, Wen G, Tian Q. Person transfer GAN to bridge domain gap for person re-identification. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 79–88
- 16 Zhong Z, Zheng L, Li S Z, Yang Y. Generalizing a person retrieval model hetero- and homogeneously. In: Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer, 2018. 172–188
- 17 Zhong Z, Zheng L, Luo Z M, Li S Z, Yang Y. Invariance matters: Exemplar memory for domain adaptive person re-identification. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019. 598–607
- 18 Yu H X, Zheng W S, Wu A C, Guo X W, Gong S G, Lai J H. Unsupervised person re-identification by soft multilabel learning. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019. 2143–2152
- 19 Saito K, Watanabe K, Ushiku Y, Harada T. Maximum classifier discrepancy for unsupervised domain adaptation. In: Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE, 2018. 3723–3732
- 20 Fu Y, Wei Y C, Wang G S, Zhou Y Q, Shi H H, Uic U, et al. Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification. In: Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea (South): IEEE, 2019. 6111–6120
- 21 Zhai Y P, Lu S J, Ye Q X, Shan X B, Chen J, Ji R R, et al. AD-Cluster: Augmented discriminative clustering for domain adaptive person re-identification. In: Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE, 2020. 9018–9027
- 22 Yang F X, Li K, Zhong Z, Luo Z M, Sun X, Cheng H, et al. Asymmetric co-teaching for unsupervised cross-domain person re-identification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, **34**(7): 12597–12604
- 23 Ge Y X, Chen D P, Li H S. Mutual mean-teaching: Pseudo label refinery for unsupervised domain adaptation on person re-identification. arXiv: 2001.01526, 2020
- 24 Wang W H, Zhao F, Liao S C, Shao L. Attentive WaveBlock: Complementarity-enhanced mutual networks for unsupervised domain adaptation in person re-identification and beyond. *IEEE Transactions on Image Processing*, 2022, **31**: 1532–1544
- 25 Bertocco G C, Andaló F, Rocha A. Unsupervised and self-adaptive techniques for cross-domain person re-identification. *IEEE Transactions on Information Forensics and Security*, 2021, **16**: 4419–4434
- 26 Sheng K K, Li K, Zheng X W, Liang J, Dong W M, Huang F Y, et al. On evolving attention towards domain adaptation. arXiv: 2103.13561, 2021
- 27 Zheng Z D, Yang X D, Yu Z D, Zheng L, Yang Y, Kautz J. Joint discriminative and generative learning for person re-identification. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE, 2019. 2133–2142
- 28 Hermans A, Beyer L, Leibe B. In defense of the triplet loss for person re-identification. arXiv: 1703.07737, 2017
- 29 Zheng L, Shen L Y, Tian L, Wang S J, Wang J D, Tian Q. Scalable person re-identification: A benchmark. In: Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015. 1116–1124
- 30 Zheng Z D, Zheng L, Yang Y. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro. In: Proceedings of the 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017. 3774–3782
- 31 Felzenszwalb P F, Girshick R B, McAllester D, Ramanan D.

- Object detection with discriminatively trained part-based models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010, **32**(9): 1627–1645
- 32 Ristani E, Solera F, Zou R, Cucchiara R, Tomasi C. Performance measures and a data set for multi-target, multi-camera tracking. In: *Proceedings of the European Conference on Computer Vision*. Amsterdam, Netherlands: Springer, 2016. 17–35
- 33 Russakovsky O, Deng J, Su H, Krause J, Satheesh S, Ma S A, et al. ImageNet large scale visual recognition challenge. *International Journal of Computer Vision*, 2015, **115**(3): 211–252
- 34 Lin Y T, Dong X Y, Zheng L, Yan Y, Yang Y. A bottom-up clustering approach to unsupervised person re-identification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2019, **33**(1): 8738–8745
- 35 Lin Y T, Xie L X, Wu Y, Yan C G, Tian Q. Unsupervised person re-identification via softened similarity learning. In: *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, USA: IEEE, 2020. 3387–3396
- 36 Lin S, Li H L, Li C T, Kot A C. Multi-task mid-level feature alignment network for unsupervised cross-dataset person re-identification. In: *Proceedings of the 29th British Machine Vision Conference*. Newcastle, UK: 2018.
- 37 Li H F, Dong N, Yu Z T, Tao D P, Qi G Q. Triple adversarial learning and multi-view imaginative reasoning for unsupervised domain adaptation person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, 2022, **32**(5): 2814–2830
- 38 Song L C, Wang C, Zhang L F, Du B, Zhang Q, Huang C, et al. Unsupervised domain adaptive re-identification: Theory and practice. *Pattern Recognition*, 2020, **102**: 107173
- 39 Zhang X Y, Cao J W, Shen C H, You M Y. Self-training with progressive augmentation for unsupervised cross-domain person re-identification. In: *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision*. Seoul, Korea (South): IEEE, 2019. 8221–8230
- 40 Zhai Y P, Ye Q X, Lu S J, Jia M X, Ji R R, Tian Y H. Multiple expert brainstorming for domain adaptive person re-identification. In: *Proceedings of the 16th European Conference on Computer Vision*. Glasgow, UK: 2020. 594–611
- 41 Ainam J P, Qin K, Owusu J W, Lu G M. Unsupervised domain adaptation for person re-identification with iterative soft clustering. *Knowledge-Based Systems*, 2021, **212**: Article No. 106644
- 42 Li H F, Xu K X, Li J X, Lu G M, Xu Y, Yu Z T, et al. Dual-stream reciprocal disentanglement learning for domain adaptation person re-identification. arXiv: 2106.13929, 2021
- 43 Zhang H, Cao H H, Yang X, Deng C, Tao D C. Self-training with progressive representation enhancement for unsupervised cross-domain person re-identification. *IEEE Transactions on Image Processing*, 2021, **30**: 5287–5298
- 44 Ge Y X, Zhu F, Chen D P, Zhao R, Li H S. Self-paced contrastive learning with hybrid memory for domain adaptive object re-ID. In: *Proceedings of the 34th International Conference on Neural Information Processing Systems*. Vancouver, Canada: Curran Associates Inc., 2020. 11309–11321
- 45 Sun J, Li Y F, Chen H J, Peng Y H, Zhu J L. Unsupervised cross domain person re-identification by multi-loss optimization learning. *IEEE Transactions on Image Processing*, 2021, **30**: 2935–2946
- 46 Zheng K C, Lan C L, Zeng W J, Zhan Z Z, Zha Z J. Exploiting sample uncertainty for domain adaptive person re-identification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021, **35**(4): 3538–3546
- 47 Zheng Z D, Yang Y. Rectifying pseudo label learning via uncertainty estimation for domain adaptive semantic segmentation. *International Journal of Computer Vision*, 2021, **129**(4): 1106–1120



郭迎春 河北工业大学人工智能与数据科学学院副教授。2006 年获得天津大学信号与信息处理专业博士学位。主要研究方向为数字图像处理和计算机视觉。

E-mail: gyc@scse.hebut.edu.cn

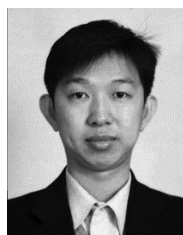
(**GUO Ying-Chun** Associate professor at the School of Artificial Intelligence, Hebei University of Technology. She received her Ph.D. degree in signal and information processing from Tianjin University in 2006. Her research interest covers digital image processing and computer vision.)



冯 放 河北工业大学人工智能与数据科学学院硕士研究生。主要研究方向为行人重识别。

E-mail: fengfang0901@163.com

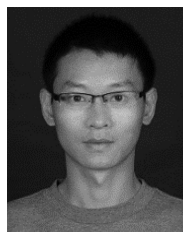
(**FENG Fang** Master student at the School of Artificial Intelligence, Hebei University of Technology. His main research interest is person re-identification.)



阎 刚 河北工业大学人工智能与数据科学学院副教授。2019 年获得河北工业大学微电子学与固体电子学专业博士学位。主要研究方向为图像处理和智能监控。本文通信作者。

E-mail: yangang@scse.hebut.edu.cn

(**YAN Gang** Associate professor at the School of Artificial Intelligence, Hebei University of Technology. He received his Ph.D. degree in microelectronics and solid state electronics from Hebei University of Technology in 2019. His research interest covers image processing and intelligent surveillance. Corresponding author of this paper.)



郝小可 河北工业大学人工智能与数据科学学院副教授。2017 年获得南京信息工程大学计算机与技术专业博士学位。主要研究方向为机器学习, 医学图像分析和生物信息学。E-mail: haoxiaoke@scse.hebut.edu.cn

(**HAO Xiao-Ke** Associate professor at the School of Artificial Intelligence, Hebei University of Technology. He received his Ph.D. degree in computer science and technology from Nanjing University of Information Science and Technology in 2017. His research interest covers machine learning, medical image analysis and bioinformatics.)