计算机科学
COMPUTER SCIENCE

# 一种单阶段无监督可见光-红外跨模态行人重识别方法

娄刃, 和任强, 赵三元, 郝昕, 周跃琪, 汪心渊, 李方芳

引用本文

娄刃, 和任强, 赵三元, 郝昕, 周跃琪, 汪心渊, 李方芳. 一种单阶段无监督可见光-红外跨模态行人重识别方法[J]. 计算机科学, 2024, 51(6A): 230600138-7.

LOU Ren, HE Renqiang, ZHAO Sanyuan, HAO Xin, ZHOU Yueqi, WANG Xinyuan, LI Fangfang. Single Stage Unsupervised Visible-infrared Person Re-identification [J]. Computer Science, 2024, 51(6A): 230600138-7.

---

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.230600138

# 一种单阶段无监督可见光-红外跨模态行人重识别方法

娄 刃[1] 和任强[2] 赵三元[2,3] 郝 昕[2] 周跃琪[1] 汪心渊[1] 李方芳[4]
1 浙江省交通运输科学研究院 杭州 310000
2 北京理工大学计算机学院 北京 100081
3 北京理工大学长三角研究院(嘉兴) 浙江 嘉兴 314011
4 浙江交投高速公路运营管理有限公司企业研究院 杭州 310000
(179787711@qq.com)

**摘 要** 无监督"可见光-红外"跨模态行人重识别任务能够缓解智能监控场景中需要大量人工标注的问题。常见多阶段模型用于处理不同模态数据。文中提出了一种有效的单阶段无监督跨模态行人重识别的方法,设计了基于置信因子的聚类算法和图嵌入的跨模态特征处理方法,分别用于解决无标签问题和跨模态问题。实验结果表明,相较于现有算法,所提方法在 $r=1$ 时精度至少取得了7%的提高。

**关键词**:跨模态学习;无监督行人重识别;可见光-红外行人重识别;无监督学习;跨模态特征处理
**中图分类号** TP391

## Single Stage Unsupervised Visible-infrared Person Re-identification

LOU Ren[1],HE Renqiang[2],ZHAO Sanyuan[2,3],HAO Xin[2],ZHOU Yueqi[1],WANG Xinyuan[1] and LI Fangfang[4]
1 Zhejiang Academy of Transportation Sciences,Hangzhou 310000,China
2 School of Computer Science,Beijing Institute of Technology,Beijing 100081,China
3 Yangtze River Delta Research Institute(Jiaxing),Beijing Institute of Technology,Jiaxing,Zhejiang 314011,China
4 Enterprise Institute of Zhejiang Communications Investment Expressway Operation Management Co. Ltd,Hangzhou 310000,China

**Abstract** The unsupervised visible-infrared multi-modal person re-identification can alleviate the problem that a lot of manual labeling is required in the intelligent monitoring scene. Common multi-stage models are used to process different modal data separately. This paper proposes an effective single-stage unsupervised cross-modal pedestrian recognition method,and designs a clustering algorithm based on confidence factor and a cross-modal feature processing method based on graph embedding to solve the unlabeled problem and cross-modal problem respectively. Experimental results show that compared with the existing algorithms, the proposed algorithm has achieved an improvement of at least 7% in the case of $r=1$.

**Keywords** Cross-modal learning,Unsupervised person re-identification,Visible-infrared person re-identification,Unsupervised learning,Cross-modal feature processing

## 1 引言

行人重识别是计算机视觉领域的一项重要研究任务,根据查询人员图像,查询到此人出现过的其他场景下的图像,对智能安防、城市安全起到重大作用。为满足全天候监控需求,监控相机在夜间切换为红外模式。对于现实世界海量的监控数据,人工标注将带来巨大成本。本工作在无标签的情况下建立跨模态行人重识别模型,是一项具有社会价值的研究任务。

目前基于聚类的伪标签生成方法是无监督行人重识别任务中常用的方法。其缺点是伪标签的准确程度将直接影响后续监督训练的好坏,错误的伪标签将严重影响模型的性能,且不恰当的伪标签更新方式会使其准确率随着迭代次数的增加逐渐降低。另一方面,现有的无监督行人重识别工作大多围绕单模态数据展开,对于多模态数据的研究经验较少。由于模态间较大的差异,聚类算法通常会将两种模态的数据判定为不同身份,因此难以得到理想的聚类结果。

针对上述问题,本工作的贡献点主要包括以下几个方面:

1)提出了基于置信因子的跨模态聚类算法,通过分析聚类簇的稳定性,选择可靠度更高的伪标签,有效提高无监督训练时的伪标签可信度。

2)设计了基于图嵌入的跨模态特征处理方法,鼓励学习

基金项目:浙江省交通运输厅科技计划项目:交通流雷视融合感知系统评测技术研究(202209);浙江省科学技术厅公益性项目:基于雷视一体设备的车辆轨迹数据质量评测技术研究(LGC22E080003)
This work was supported by the Research on Evaluation Technology of Traffic Flow based on Vision-Radar Fusion Perception System(202209) and Research on Vehicle Trajectory Data Quality Evaluation Technology Based on Radar-vision Integrated Equipment(LGC22E080003).
通信作者:赵三元(zhaosanyuan@bit.edu.cn)

相同身份的不同模态样本，惩罚不同身份的相同模态样本，为模态间和模态内样本关系提供了一个表达框架。

3)针对无监督跨模态行人重识别问题设计了一个有效的网络框架，这是对无监督跨模态行人重识别领域的新的尝试。实验证明，相比现有算法，本文提出的方法在 SYSU-MM01 和 RegDB 公开数据集上取得了更好的结果，并显著提高了模型的精度。

## 2 相关工作

### 2.1 无监督行人重识别相关工作

对于无监督单模态行人重识别，很多方法关注于生成伪标签来转换成监督学习。例如，基于聚类的模型使用 $k$-means、DBSCAN 和层次聚类来预测伪标签进行自训练。伪标签生成的正确与否会在很大程度上影响后续的训练学习。Lin 等提出的 BUC 模型[1]将每个图像视为一个聚类簇，然后逐步合并簇。Lin 等[2]将基于聚类的伪标签替换为基于相似度的软标签。为了提高伪标签的质量，Zeng 等[3]提出了分层聚类算法。此外，由于每一个身份标签可能包含在多个正实例中，Yang 等[4]提出的 MMCL 模型在无监督行人重识别中引入了基于内存的多标签分类损失。然而，这些方法假设输入图像是来自于同一模态的，聚类方法能很容易应用于这些模型中。

对于无监督的跨模态行人重识别任务，由于模态间存在巨大的模态差异，类内特征差异过大，因此聚类假设不再成立，难以生成可靠的跨模态标签。因此，如何弥合模态间的鸿沟，成为无监督跨模态行人重识别的首要问题。Liang 等提出的 H2H[5]通过从同质到异质两个阶段的学习，先对每个模态形成行人身份的伪标签，再通过模态的一致性学习来对齐两个模态。该方法为无监督跨模态行人重识别问题提供了一个解决方案，但其没有设计针对跨模态数据的聚类方法，并且训练是两阶段的。Yang 等的 ADCA 模型[4]提出了一个双流对比学习框架，利用通道增强的方法学习到与颜色无关的特征，以减小模态差异。目前无监督跨模态行人重识别的研究相对较少，尚未形成成熟的体系。

### 2.2 无监督跨模态探索

早先的方法常使用典型相关分析[6-8]或自编码器[9-10]来学习模态不变特征。近年来一些研究尝试使用对抗学习的思想生成模态不可分特征来进行跨模态匹配。He 等[11]引入了一种模态分类器，学习可能混淆模态分类器的特征表示；Chung 等[12]通过对比学习在特征空间中对齐语音和文本模态；Li 等[13]使用两种对抗网络来最大化不同模态的语义相关性和一致性；Wang 等[14]提出一种耦合的循环生成对抗网络，在外循环中学习通用特征表示，在内循环中生成相应的哈希码。然而，由于不同模态的先验知识存在较大差异，这些无监督跨模态检索方法具有较大的局限性。文本和图像模态在低级语义上并不共享特征，而可见光和红外模态的图像在低级语义上共享纹理和形状特征。因此，有必要为无监督的跨模态行人重识别任务设计适合的算法。本工作提出了一个有效的单阶段无监督跨模态行人重识别模型，在降低模型复杂度的同时，提升了模型的预测精度。

## 3 图嵌入无监督行人重识别网络

本工作提出了图嵌入无监督行人重识别网络，其网络结构如图 1 所示，主要包括两个部分：置信因子聚类算法和图嵌入的跨模态特征处理方法。根据输入的可见光和红外的两种模态的图像，使用相同的特征编码器对其提取特征，再使用基于置信因子的聚类算法对无标签数据进行伪标签赋值。并提出图嵌入的跨模态特征处理方法，对红外、可见光样本建立图关系模型，通过关系图鼓励相同身份的不同模态数据在特征空间中相互靠近，使用惩罚图对不同身份的特征进行疏远。基于图嵌入的跨模态特征处理方法能方便地应用多种约束策略，优化聚类效果，使模型随着网络训练不断提高识别精度。

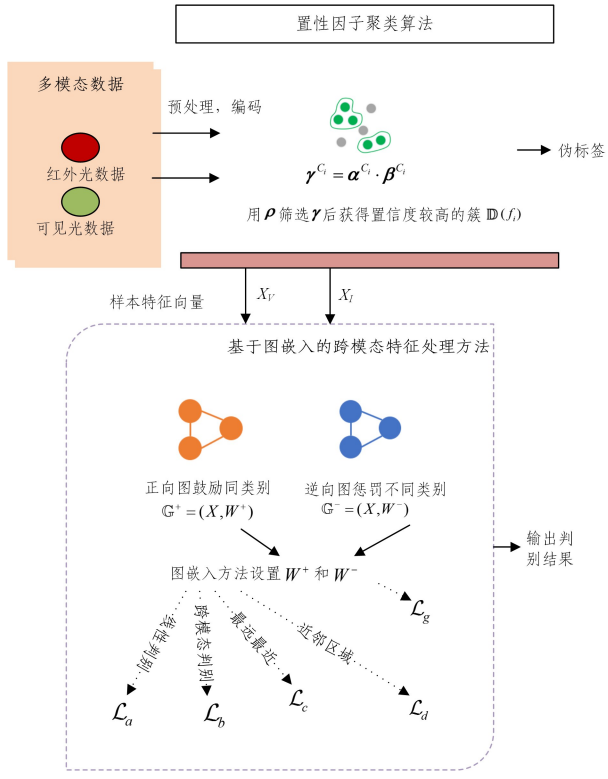


图 1 置信因子聚类和图嵌入的无监督跨模态特征处理方法总览

Fig. 1 Overview of unsupervised cross-modal feature processing methods for confidence factor clustering and graph embedding

### 3.1 基于置信因子的聚类

无监督任务的常见做法是先生成伪标签，再转换为监督问题求解。伪标签的质量将直接影响最终的结果。因此，有必要设计一种合理的评判标准，选取可信度高的簇作为训练样本以保证训练质量。可靠的簇应当拥有良好的簇间分离性和簇内紧密性。

假设输入样本经过特征编码器后得到的特征为 $\boldsymbol{f}=\{\boldsymbol{f}_1,\boldsymbol{f}_2,\cdots,\boldsymbol{f}_n\}$，特征 $\boldsymbol{f}_i$ 经过聚类算法 $\mathbb{D}$ 后得到其所属簇为 $\mathbb{D}(f_i)$。聚类算法受人工设置的参数影响，更宽松的参数设置通常产生较少的离群值，每个簇的样本可能会增加，不同的簇可能被合并。若增加簇的距离阈值，某个簇依然不变，那么可认为该簇表现出良好的分离性；若减少簇的距离阈值，某个簇未被分解或无样本丢失，那么可认为该簇表现出良好的紧密性，即该簇具有较高的置信度。具体地，引入基于置信因子 $\gamma$ 来度量簇间的分离性和紧密性，评判聚类簇的可信程度。

定义特征 $f_i$ 经过聚类后所在簇的簇间分离性 $\alpha(\boldsymbol{f}_i)$ 为:

$$\alpha(\boldsymbol{f}_i)=\frac{\mathcal{Q}(\boldsymbol{f}_i)\bigcap\mathcal{Q}^{+}(\boldsymbol{f}_i)}{\mathcal{Q}(\boldsymbol{f}_i)\bigcup\mathcal{Q}^{+}(\boldsymbol{f}_i)} \tag{1}$$

其中,$\mathcal{Q}(\boldsymbol{f}_i)$表示特征 $\boldsymbol{f}_i$ 经过聚类算法后所在的簇,$\mathcal{Q}^{+}(\boldsymbol{f}_i)$ 表示特征 $\boldsymbol{f}_i$ 在较宽松的设置后聚类所在的簇。更大的 $\alpha$ 值表示 $\mathcal{Q}^{+}$ 和 $\mathcal{Q}$ 的差异更小,体现聚类簇良好的簇间可分离性。

同样地,定义特征 $\boldsymbol{f}_i$ 经过聚类后所在簇的簇内紧密性 $\beta(\boldsymbol{f}_i)$ 为:

$$\beta(\boldsymbol{f}_i)=\frac{\mathcal{Q}(\boldsymbol{f}_i)\bigcap\mathcal{Q}^{-}(\boldsymbol{f}_i)}{\mathcal{Q}(\boldsymbol{f}_i)\bigcup\mathcal{Q}^{-}(\boldsymbol{f}_i)} \tag{2}$$

其中,$\mathcal{Q}^{-}(\boldsymbol{f}_i)$表示特征 $\boldsymbol{f}_i$ 在较严格的设置后聚类所在的簇。$\beta$ 的取值范围同样在 [0,1] 之间,更大的 $\beta$ 值表明即便在聚类要求更严格的情况下,该簇依旧表现出良好的簇内紧密性。对于划为离群值的样本,它的特征对应的 $\beta$ 值为 0。最终,在初始参数设置下聚类得到的簇集合 $C=\{C_1,C_2,\cdots,C_n\}$ 中,分别对每个簇 $C_i$ 中包含的特征 $\boldsymbol{f}^{C_i}=\{\boldsymbol{f}_1^{C_i},\boldsymbol{f}_2^{C_i},\cdots,\boldsymbol{f}_m^{C_i}\}$ 衡量其簇间分离性和簇内紧密性,最后取平均值,即对于簇 $C_i$,其簇间分离性表示为:

$$\boldsymbol{\alpha}^{C_i}=\frac{\alpha(\boldsymbol{f}_1^{C_i}),\alpha(\boldsymbol{f}_2^{C_i}),\cdots,\alpha(\boldsymbol{f}_m^{C_i})}{m} \tag{3}$$

相应地,其簇内紧密性表示为:

$$\boldsymbol{\beta}^{C_i}=\frac{\beta(\boldsymbol{f}_1^{C_i}),\beta(\boldsymbol{f}_2^{C_i}),\cdots,\beta(\boldsymbol{f}_m^{C_i})}{m} \tag{4}$$

定义置信度因子 $\gamma$,根据簇间分离性 $\alpha$ 和簇内紧密性 $\beta$ 衡量一个簇的可信程度,对于簇 $C_i$,其最终的置信度因子 $\gamma^{C_i}$ 表示为:

$$\boldsymbol{\gamma}^{C_i}=\boldsymbol{\alpha}^{C_i}\cdot\boldsymbol{\beta}^{C_i} \tag{5}$$

在得到每个簇的置信因子后,设置置信度阈值 $\rho$ 作为置信程度的下限。当 $\gamma\geqslant\rho$ 时,该簇会作为训练样本由模型训练;反之,该簇的样本将被视为离群值而不对训练产生影响。通过对置信因子的筛选,生成的簇将具有更高的可信度,从而有效提升伪标签的准确率,缓解了聚类算法初期可能会产生大量噪声的问题。随着网络训练的进行,其提取的特征将更具判别力,能帮助后续聚类算法减少离群值,优化簇间分离性和簇内紧密性,从而保证网络朝着正确的梯度方向进行更新。

### 3.2　基于图嵌入的跨模态特征表达框架

基于图的表达方法为数据集中样本关系提供了一个良好的表达框架,图方法将样本视作图的节点,将样本间关系通过边的建立彼此联系起来,并通过对边权值的设置实现对样本间的约束。

图嵌入方法能有效地将经过特征提取器降维后的特征数据在特征空间中建立关系图矩阵。对于样本集合 $\boldsymbol{X}=\{x_1,\cdots,x_N\}\in\mathbb{R}^{D\times N}$,通过特征编码后得到的降维数据表示为 $\boldsymbol{Z}=\{z_1,\cdots,z_N\}\in\mathbb{R}^{1\times N}$。为了建立样本间的正向关系图矩阵,将建立图 $\mathbb{G}=(\boldsymbol{X},\boldsymbol{W})$,其中 $\boldsymbol{X}$ 中的列表示每个样本的特征向量,$\boldsymbol{W}$ 矩阵表示样本之间的对应的关系。对于 $\boldsymbol{W}^{+}$ 矩阵中的一个元素 $\boldsymbol{w}^{(i,j)}$,它对应地表示样本 $x_i$ 和 $x_j$ 之间的边,描述它们在特征空间中的关系。本工作将 $\boldsymbol{W}^{+}$ 定义为样本间的正向关系矩阵,用来鼓励属于同种类别的样本彼此靠近,表示为 $\mathbb{G}^{+}=(\boldsymbol{X},\boldsymbol{W}^{+})$。定义逆向惩罚矩阵 $\boldsymbol{W}^{-}$,用来建立惩罚图 $\mathbb{G}^{-}=(\boldsymbol{X},\boldsymbol{W}^{-})$,一维向量 $\boldsymbol{z}^{*}$ 优化更新方式遵循文献[15]中的规则。

$$\boldsymbol{z}^{*}=\arg\min_{\boldsymbol{z}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{z}=c}\sum_{i\neq j}\parallel z_i-z_j\parallel_2^2\boldsymbol{W}^{(i,j)}=\arg\min_{\boldsymbol{z}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{z}=c}\boldsymbol{z}^{\mathrm{T}}\boldsymbol{L}\boldsymbol{z} \tag{6}$$

其中,$c$ 为常数,$\boldsymbol{L}^{+}$ 和 $\boldsymbol{L}^{-}$ 分别为正向图 $\mathbb{G}^{+}$ 和逆向图 $\mathbb{G}^{-}$ 的拉普拉斯矩阵,表示为:

$$\begin{aligned}\boldsymbol{L}^{+}&=\boldsymbol{D}^{+}-\boldsymbol{W}^{+}\\ \boldsymbol{L}^{-}&=\boldsymbol{D}^{-}+\boldsymbol{W}^{-}\end{aligned} \tag{7}$$

$\boldsymbol{D}$ 矩阵为 $\boldsymbol{W}$ 矩阵每行的和,作为该行元素产生的对角矩阵,表示与样本 $x_i$ 相关的其他样本的数量,用公式表示为 $\boldsymbol{D}=\sum_j\boldsymbol{W}^{(i,j)}$。定义跨模态数据集 $\boldsymbol{X}=[\boldsymbol{X}_r+\boldsymbol{X}_f]$,其中 $\boldsymbol{X}_r\in\mathbb{R}^{\boldsymbol{D}\times\boldsymbol{N}_r}$ 表示可见光图像样本集合,$\boldsymbol{X}_f\in\mathbb{R}^{\boldsymbol{D}\times\boldsymbol{N}_f}$ 表示红外图像样本集合,假定特征编码器 $\varphi(\cdot)$ 将输入样本从原始 D×N 维映射到隐式空间中变为 d×N,建立包含所有映射后的模态特征的矩阵集合为 $\boldsymbol{\Phi}=[\boldsymbol{\Phi}(\boldsymbol{X}_r),\boldsymbol{\Phi}(\boldsymbol{X}_f)]$。将样本间的成对关系编码成图关系矩阵,建立正向关系图矩阵 $\boldsymbol{G}^{+}=(\boldsymbol{X},\boldsymbol{W}^{+})$,其拉普拉斯矩阵的计算[16]可以转换为:

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\parallel\boldsymbol{\Phi}^{(i)}=\boldsymbol{\Phi}^{(j)}\parallel_2^2\boldsymbol{W}^{+}=\mathrm{Tr}(\boldsymbol{\Phi}\boldsymbol{L}^{+}\boldsymbol{\Phi}^{\mathrm{T}}) \tag{8}$$

同样地,对需要惩罚的样本对间建立逆向惩罚图矩阵 $\boldsymbol{G}^{-}=(\boldsymbol{X},\boldsymbol{W}^{-})$,其拉普拉斯矩阵的计算表示为:

$$\sum_{i=1}^{N}\sum_{j=1}^{N}\parallel\boldsymbol{\Phi}^{(i)}-\boldsymbol{\Phi}^{(j)}\parallel_2^2\boldsymbol{W}^{-}=\mathrm{Tr}(\boldsymbol{\Phi}\boldsymbol{L}^{-}\boldsymbol{\Phi}^{\mathrm{T}}) \tag{9}$$

由于图嵌入的目的是对同身份样本进行拉近,对不同身份样本进行疏远,即最小化式(8),最大化式(9),可用迹比表示图嵌入的优化目标。

$$\varphi^{*}=\arg\min_{\varphi}\frac{\mathrm{Tr}(\boldsymbol{\Phi}\boldsymbol{L}^{+}\boldsymbol{\Phi}^{\mathrm{T}})}{\mathrm{Tr}(\boldsymbol{\Phi}\boldsymbol{L}^{-}\boldsymbol{\Phi}^{\mathrm{T}})} \tag{10}$$

进一步地,图嵌入的损失函数可以表示为:

$$\mathcal{L}_g=\frac{\mathrm{Tr}(\boldsymbol{\Phi}\boldsymbol{L}^{+}\boldsymbol{\Phi}^{\mathrm{T}})}{\mathrm{Tr}(\boldsymbol{\Phi}\boldsymbol{L}^{-}\boldsymbol{\Phi}^{\mathrm{T}})} \tag{11}$$

图嵌入的方法通过建立样本间的图边际关系实现对实例特征空间层面上的拉近和疏远,其中拉普拉斯矩阵将样本间的先验知识编码到特定的模式中。值得注意的是,图嵌入结构的核心是对 $\boldsymbol{W}$ 矩阵的设置。下一节将详细讨论 $\boldsymbol{W}$ 矩阵对图嵌入结果的影响。

### 3.3　图嵌入结构的表达方法

#### 3.3.1　线性判别的图嵌入结构

倘若不考虑模态差异,仅针对不同身份的样本建立图嵌入结构,则其可看做被一个线性判别器,即定义正向关系矩阵为:

$$\boldsymbol{W}^{+}=\begin{cases}1, & \text{if } \ell_i=\ell_j\\ 0, & \text{otherwise}\end{cases} \tag{12}$$

其中,$\ell_i$ 和 $\ell_j$ 分别表示样本 $x_i$,$x_j$ 的身份标签,该公式表明正向关系图矩阵仅对相同身份标签建立边际关系。同理逆向惩罚矩阵可定义为:

$$\boldsymbol{W}^{-}=\begin{cases}1, & \text{if } \ell_i\neq\ell_j\\ 0, & \text{otherwise}\end{cases} \tag{13}$$

该公式表明逆向惩罚图矩阵仅对相同模态的不同身份标签建立边际关系。根据 3.2 节介绍的图嵌入优化规则,可利用迹比目标函数表示线性图嵌入结构的损失函数为:

$$\mathcal{L}_a=\frac{\mathrm{Tr}(\boldsymbol{\Phi}\boldsymbol{L}^{+}\boldsymbol{\Phi}^{\mathrm{T}})}{\mathrm{Tr}(\boldsymbol{\Phi}\boldsymbol{L}^{-}\boldsymbol{\Phi}^{\mathrm{T}})} \tag{14}$$

#### 3.3.2　跨模态判别的图嵌入结构

线性判别没有考虑到特征空间中模态差异带来的影响。

在跨模态问题中，特征空间层面模态间距离要显著大于不同身份类别间距离。为了解决这一矛盾，针对跨模态场景建立图嵌入结构，对相同身份样本建立正向关系矩阵，对相同模态的不同身份样本建立逆向惩罚矩阵。对于正向关系图矩阵，定义如下：

$$\boldsymbol{W}^{+}=\begin{cases}1, & \text{if } \ell_i=\ell_j \text{ and } \mathcal{D}_i\neq\mathcal{D}_j\\ 0, & \text{otherwise}\end{cases} \tag{15}$$

其中，$\ell_i$，$\ell_j$ 分别表示样本 $x_i$，$x_j$ 的身份标签，$\mathcal{D}_i$，$\mathcal{D}_j$ 表示样本 $x_i$，$x_j$ 的模态标签。该公式表明正向关系图矩阵仅对不同模态的相同身份标签建立边际关系。同样地，对于逆向惩罚图矩阵，定义如下：

$$\mathcal{L}_b=\frac{\mathrm{Tr}(\boldsymbol{\Phi}\boldsymbol{L}^{+}\boldsymbol{\Phi}^{\mathrm{T}})}{\mathrm{Tr}(\boldsymbol{\Phi}\boldsymbol{L}^{-}\boldsymbol{\Phi}^{\mathrm{T}})} \tag{16}$$

3.3.3 最远最近的图嵌入结构

在特征空间中，对于每一个簇，都有相距其他样本最远的样本，为了明显增加聚类簇的紧密性，可对同身份样本中相距最远的特征向量进行惩罚。同样，可通过疏远相距其他类别最近的样本来保持类别之间的距离。根据这一思想，对于目标样本，惩罚不同模态相同身份的最远距离，鼓励不同模态不同身份的最近距离，损失函数可表示为：

$$\mathcal{L}_c=\sum_{x_j\in\mathcal{D}_v}\max_{\substack{x_i\in\mathcal{D}_I\\ \ell_i=\ell_j}}\{a\mid a\in d_{ij}^2\}-\min_{\substack{x_i\in\mathcal{D}_I\\ \ell_i\neq\ell_j}}\{b\mid b\in d_{ij}^2\} \tag{17}$$

其中，$d_{ij}$ 表示样本 $x_i$ 和 $x_j$ 之间的距离。转换成图嵌入的结构，则设置 $\boldsymbol{W}^{+}$ 和 $\boldsymbol{W}^{+}$ 为：

$$\boldsymbol{W}^{+}=\begin{cases}1, & \text{if } d_{ij}=\max\limits_{x_k\in\mathcal{D}_I}\{a\mid a\in d_{kj}\}\\ & \ell_j=\ell_i=\ell_k \text{ and } \mathcal{D}_i\neq\mathcal{D}_j\\ 0, & \text{otherwise}\end{cases}$$

$$\boldsymbol{W}^{-}=\begin{cases}1, & \text{if } d_{ij}=\min\limits_{x_k\in\mathcal{D}_I}\{b\mid b\in d_{kj}\}\\ & \ell_j\neq\ell_i=\ell_k \text{ and } \mathcal{D}_i\neq\mathcal{D}_j\\ 0, & \text{otherwise}\end{cases} \tag{18}$$

损失函数的形式定义为迹差：

$$\mathcal{L}_c=\mathrm{Tr}(\boldsymbol{\Phi}\boldsymbol{L}^{+}\boldsymbol{\Phi}^{\mathrm{T}}-\lambda\boldsymbol{\Phi}\boldsymbol{L}^{-}\boldsymbol{\Phi}^{\mathrm{T}}) \tag{19}$$

其中，$\lambda$ 为可调节的平衡参数。

3.3.4 线性判别的图嵌入结构

在无监督聚类过程中，聚类簇会产生噪声样本，可能在图矩阵中建立错误的连接。为了缓解这一情况，近邻区域的图嵌入结构为每个目标样本建立其 $k$ 个近邻集合 $\mathcal{N}(i)$。其基本思想是，对每个目标样本，相距越近的样本在本次聚类中越可信，因此仅需考虑该批次中较为可信的样本间距离即可。近邻区域的图嵌入结构可表示为：

$$\mathcal{L}_{\text{neighbour}}=\sum_{\substack{x_i\in\mathcal{D}_v\\ x_j\in\mathcal{N}(i)}}\parallel\varphi_n(\boldsymbol{x}_i)-\varphi_n(\boldsymbol{x}_j)\parallel_2\mathcal{K}_{\mathrm{RBF}}(\boldsymbol{x}_i,\boldsymbol{x}_j) \tag{20}$$

其中，$\mathcal{K}_{\mathrm{RBF}}(\boldsymbol{x},\boldsymbol{x}')=\exp(-\parallel\boldsymbol{x}-\boldsymbol{x}'\parallel_2^2/2\sigma^2)$ 是径向基核函数，可被看作边的权重，它从距离角度衡量了特征向量间相似性，可被看作图嵌入结构中边的权重，表示为图嵌入结构为：

$$\boldsymbol{W}^{+}=\begin{cases}V\dfrac{\mathcal{K}_{\mathrm{RBF}}(\boldsymbol{x},\boldsymbol{x}')}{d_{ij}}, & \text{if } j\in\mathcal{N}(i) \text{ and } \mathcal{D}_i=\mathcal{D}_j=\mathcal{D}_V\\ \dfrac{1}{2}, & \text{if } \mathcal{D}_i\neq\mathcal{D}_j\\ 0, & \text{otherwise}\end{cases} \tag{21}$$

对于逆向惩罚矩阵，设定阈值 $\epsilon$，在 $d_{ij}$ 距离小于 $\epsilon$ 时进行惩罚，表示为：

$$\begin{aligned}\mathcal{L}&=-\sum_{\substack{x_i\in\mathcal{D}_I\\ x_j\in\mathcal{D}_V\\ \ell_i\neq\ell_j\\ d_{ij}<\epsilon}}\frac{1}{2}(d_{ij}-\epsilon)^2\\ &=-\sum_{\substack{x_i\in\mathcal{D}_I\\ x_j\in\mathcal{D}_V\\ \ell_i\neq\ell_j\\ d_{ij}<\epsilon}}d_{ij}^2\frac{1}{2}\left(1+\frac{\epsilon^2}{d_{ij}^2}-\frac{2\epsilon}{d_{ij}}\right)\\ &=\sum_{\substack{x_i\in\mathcal{D}_I\\ x_j\in\mathcal{D}_V}}\parallel\varphi_n(\boldsymbol{x}_i)-\varphi_n(\boldsymbol{x}_j)\parallel_2^2\boldsymbol{W}^{-}\\ &=-\mathrm{Tr}(\boldsymbol{\Phi}\boldsymbol{L}^{-}\boldsymbol{\Phi}^{\mathrm{T}})\end{aligned} \tag{22}$$

表示为图嵌入形式为：

$$\boldsymbol{W}^{-}=\begin{cases}\dfrac{1}{2}+\dfrac{\epsilon^2}{d_{ij}^2}-\dfrac{\epsilon}{d_{ij}}, & \text{if } d_{ij}<\epsilon \text{ and } \ell_i\neq\ell_j \text{ and } \mathcal{D}_i\neq\mathcal{D}_j\\ 0, & \text{otherwise}\end{cases} \tag{23}$$

最终损失函数形式定义为迹差：

$$\mathcal{L}_d=\mathrm{Tr}(\boldsymbol{\Phi}\boldsymbol{L}^{+}\boldsymbol{\Phi}^{\mathrm{T}}-\lambda\boldsymbol{\Phi}\boldsymbol{L}^{-}\boldsymbol{\Phi}^{\mathrm{T}}) \tag{24}$$

后两种方法定义为迹差的原因是，考虑到这两种方法在优化过程中都是针对批次中固定个数的样本约束，实际训练时其值通常不稳定，迹比问题容易造成指数发散，选用迹差更稳妥。

## 4 实验结果与分析

### 4.1 线性判别的图嵌入结构

RegDB 数据集由一个双相机采集系统捕获，即一个可见光相机和一个红外相机，共包含 412 个行人身份，每个身份都有对应的 10 张红外模态图像和 10 张可见光模态图像。该数据集的红外图片和可见光图片拍摄于同一时刻，因此两种模态的数据具有相同的行人形状和姿态信息。

SYSU-MM01 由 6 个相机采集，包含 4 部可见光相机和 2 部红外相机，共包含 491 个行人身份，共计 287 628 张可见光图像和 15 792 张红外图像。SYSU-MM01 数据集由于采集相机数量增多，场景信息更加丰富，并且每张图像行人的姿态都是随机的，因此该数据集在封闭世界设置下与真实场景更相似，同时也带来更多挑战。

本实验采用行人重识别任务中常用的评价标准累计匹配曲线（CMC）[17]、平均精度（mAP）和平均逆负惩罚[18]（mINP）来衡量模型性能。对于 SYSU-MM01 数据集，分别采用全搜索模式和室内搜索模式两种策略，前者考虑所有室内室外图像，后者仅考虑由室内相机采集的图像。对于 RegDB 数据集，412 个行人身份中随机选择 206 个行人身份作为训练集，剩余的 206 个行人身份被选作测试集。由于该数据集的体量较小，可能导致实验效果不稳定，因此采取模型对其随机分割 10 次，将测试结果取平均的策略来获得较稳定的结果。

### 4.2 实现细节

本文提出的模型使用 PyTorch 框架构建，使用 4 块 GTX-2080Ti 显卡进行训练。训练过程持续约 10 h，共迭代 60 次。采用在 ImageNet 上预训练的 ResNet-50 作为网络主干，在全局池化层后连接 1D 批正则化层和 L2 正则化层。该部分构成网络的特征编码模块。使用衰减率为 0.000 5 的 Adam 优化器对模型参数进行优化，设置学习率的初始值为

0.00035,并且每 20 代将衰减 1/10。

开始训练前,SYSU-MM01 和 RegDB 数据集的图片尺寸被裁剪为 224×128,并进行随机数据增强,包括水平翻转、裁剪和随机擦除。后经过 DBSCAN 聚类算法生成伪标签,使用置信因子筛选值得信赖的簇进行训练。每个批次中跨模态图像被一同送入模型,每个批次图像数目为 64,包含 32 张可见光图像和 32 张红外图像。

使用 DBSCAN 聚类、杰卡德距离和 $k$-相互近邻算法对初始无标签数据进行伪标签赋值,$k$ 的初始值设置为 30,DBSCAN 聚类近邻节点之间的最大距离半径设置为 $d=0.6$,每个密集点的最小近邻数量设置为 4。所提出的基于置信因子的聚类算法中宽松和收紧的动态调整范围为 $\Delta d=0.02$,即 $d=0.62$,$d-=0.58$,置信度因子的 $\gamma$ 的阈值 $\rho$ 设为 0.9。

### 4.3　实验结果与分析

#### 4.3.1　实验对比

本小节评估所提方法与近几年各类无监督方法的性能对比,以及所提出的置信因子聚类方法和基于图嵌入的跨模态特征处理方法的有效性。表 1 和表 2 分别列出了在跨模态行人重识别数据集 SYSU-MM01 和 RegDB 上,本模型与近几年数十种模型的多项指标对比。

表 1　本文模型在 RegDB 数据集上与近几年提出的各算法的指标对比

Table 1　Comparison of the indicators of the proposed model and various algorithms proposed in recent years on RegDB dataset

| 方法 | 红外-可见光模式 | | |
|---|---|---|---|
| | $r=1$ | $r=10$ | $mAP$ |
| HOMO[19] | 4.05 | 11.23 | 5.11 |
| CycleGAN[20] | 4.73 | 14.81 | 4.86 |
| SSG[21] | 1.91 | 5.14 | 3.18 |
| ECN[22] | 2.17 | 8.38 | 2.90 |
| H2H[5] | 13.91 | 30.39 | 12.72 |
| 本文模型 | **43.11** | **60.97** | **34.08** |

表 2　本文模型在 SYSU-MM01 数据集上与近几年提出的各算法的指标对比

Table 2　Comparison of the indicators of the proposed model and various algorithms proposed in recent years on SYSU-MM01 dataset

| 方法 | 红外-可见光模式 | | |
|---|---|---|---|
| | $r=1$ | $r=10$ | $mAP$ |
| HOG[23] | 3.59 | 18.39 | 4.91 |
| histLBP[24] | 1.51 | 12.43 | 3.49 |
| GOG[25] | 1.25 | 12.23 | 3.48 |
| HOMO[19] | 10.02 | 35.23 | 11.13 |
| CycleGAN[20] | 8.34 | 31.56 | 10.55 |
| SSG[21] | 2.32 | 17.23 | 5.00 |
| ECN[22] | 8.07 | 32.49 | 12.68 |
| H2H[5] | 25.49 | 63.85 | 25.16 |
| 本文模型 | **27.40** | **72.39** | **28.33** |

在 RegDB 数据集上的实验表明,所提方法在处理跨模态行人重识别任务中表现良好,在红外-可见光查询模式中 rank-1 精度达到 43.11%,mAP 达到 34.08%,证明了网络对于处理跨模态无监督问题有正确的导向。同时该实验证明,与近几年的方法相比,所提模型在各项指标上都大幅领先,其他的方法由于主要关注于单模态无监督问题,因此在面临跨模态数据集的情形时表现较差,这也说明了对无监督跨模态行人重识别问题建立一个专用的网络框架的重要性。

在 SYSU-MM01 数据集上的实验表明,所提方法在应对较为复杂的跨模态识别场景中仍表现不俗。相比目前最先进的方法,所提方法实现了 rank-1 精度 1.91%、mAP 精度 3.17%的提升。其中 H2H[5]方法采用两阶段的模型结构,模型复杂程度上是本文所用模型的两倍,整个处理流程复杂,训练繁琐。而本工作提出的方法是单阶段的,两种模态图像共用相同的网络主干,因此本文方法无论是在模型参数量和计算复杂度上,都具有较大优势。

#### 4.3.2　消融实验

为了更详尽的展示本文方法的有效性,表 3 列出了在 SYSU-MM01 和 RegDB 数据集上使用基于置信因子聚类方法和图嵌入的跨模态特征处理方法带来的提升。

表 3　本文提出的置信因子聚类算法和图嵌入跨模态特征处理方法有效性消融实验

Table 3　Effectiveness ablation experiment of the confidence factor clustering algorithm and graph embedding cross-modal feature processing method proposed in this paper

| 方法 | SYSU-MM01 数据集 | | | RegDB 数据集 | | |
|---|---|---|---|---|---|---|
| | $r=1$ | $r=10$ | $mAP$ | $r=1$ | $r=10$ | $mAP$ |
| B | 24.85 | 69.50 | 26.27 | 35.72 | 53.46 | 29.01 |
| B+D | 26.40 | 71.15 | 27.96 | 37.52 | 55.49 | 30.30 |
| B+G($\mathcal{L}_a$) | 25.12 | 69.63 | 25.43 | 38.79 | 57.15 | 30.79 |
| B+G($\mathcal{L}_b$) | 26.21 | 70.48 | 27.56 | 38.91 | 58.92 | 31.31 |
| B+G($\mathcal{L}_c$) | 26.08 | 70.45 | 27.08 | 39.07 | 59.08 | 32.15 |
| B+G($\mathcal{L}_d$) | 26.17 | 71.08 | 27.71 | 41.11 | 60.10 | 33.09 |
| B+D+G($\mathcal{L}_b$) | 27.40 | 72.39 | 28.33 | 43.11 | 60.97 | 34.08 |

网络主干由 B 表示,基于置信因子的聚类方法由 D 表示,基于图嵌入的跨模态特征处理方法由 G 表示。如表 3 所列,在 SYSU-MM01 数据集上,基于置信因子的聚类算法相比初始模型,rank-1 和 mAP 精度分别提升了 1.55% 和 1.65%;基于图嵌入的跨模态特征处理方法相比初始模型,rank-1 和 mAP 精度分别提升了 1.27% 和 1.29%;当二者共同使用时,模型的性能进一步提升。在 RegDB 数据集上,基于置信因子的聚类方法相比较初始模型,rank-1 和 mAP 分别提升了 1.80% 和 1.29%,基于图嵌入的跨模态特征处理方法相比较初始模型,rank-1 和 mAP 分别提升了 3.07% 和 1.78%。

另外,表 3 还展示了 4 种图嵌入表示方法对模型精度的影响,其中近邻区域的图嵌入方法整体上优于其他方法,但考虑到其较高的训练代价,所以最终的框架中选择效果相当的跨模态判别的图嵌入结构。消融实验验证了本文方法的有效性,基于置信因子的聚类方法提升了对值得信赖的聚类簇的筛选效果,通过迭代训练不断提升伪标签的准确率;基于图嵌入的跨模态特征处理方法通过建立图关系矩阵对样本间实现鼓励和惩罚,有效地引导网络提取模态不变和有判别力的特征。二者共同辅助模型挖掘无监督场景下行人身份的潜在关联信息,完成更深层次的推理任务。

#### 4.3.3　试验参数分析

本文方法中共有 3 个敏感参数,即置信度阈值 $\rho$,DBSCAN 聚类的半径 $d$,簇间分离性和簇内紧密性的动态调整范围 $\Delta d$。图 2 给出了在 SYSU-MM01 数据集上对所提出的置信因子聚类算法核心参数置信度阈值 $\rho$ 的敏感性分析。当 $\rho=0.9$ 时,模型的性能最优。$\rho$ 反应了聚类算法的包容程度,更大的 $\rho$ 值将最值得信赖的聚类簇筛选出来,使得伪标签的

准确度更可靠，但同时训练可用的伪标签数量减少，可能造成在训练初期没有值得信赖的簇的现象，导致训练迭代失败。

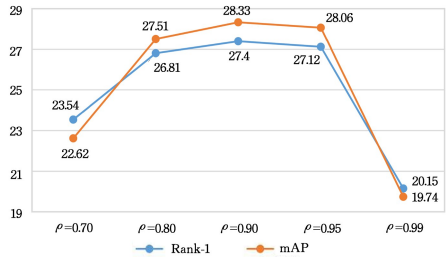


图 2 SYSU-MM01 数据集上置信度阈值 $\rho$ 敏感度分析

Fig. 2 Sensitivity analysis of confidence threshold $\rho$ on SYSU-MM01 dataset

如图 3 所示，DBSCAN 聚类半径 $d$ 决定了在聚类时能成立簇的最小范围，较大的 $d$ 值会将更多的离群值纳入聚类范围或产生多个簇的合并，较小的 $d$ 可能丢失部分难样本，不利于模型对难样本特征的学习，且受数据集在特征空间分布的稀疏程度影响，针对不同任务，$d$ 的最佳取值不同。本文使用的方法当 $d=0.6$ 时，模型性能最佳。

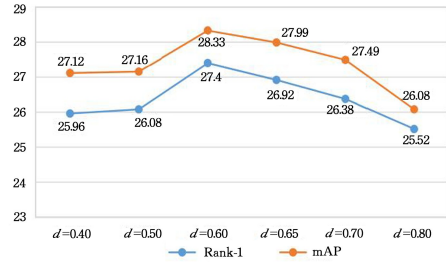


图 3 SYSU-MM01 数据集上 DBSCAN 聚类半径 $d$ 参数敏感度分析

Fig. 3 Sensitivity analysis of DBSCAN clustering radiusparameterd on SYSU-MM01 dataset

如图 4 所示，簇间分离性和簇内紧密性的动态调整范围 $\Delta d$ 代表了放宽和收紧的程度，更大的 $\Delta d$ 对簇的要求更高，理想的聚类簇应当在放宽聚类范围时，仍然没有新的样本添加进来，在收紧聚类范围时，该簇不应丢失任何聚类样本且不应分裂成几个较小的簇。

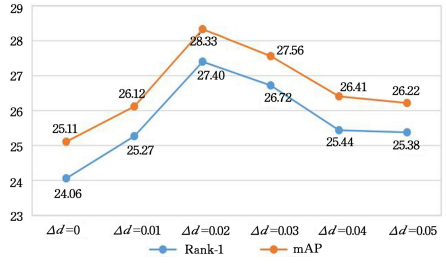


图 4 SYSU-MM01 数据集上簇间分离性和簇内紧密性的动态调整范围 $\Delta d$ 敏感度分析

Fig. 4 Sensitivity analysis of dynamic adjustment range $\Delta d$ of inter-cluster separation and intra-cluster closeness on SYSU-MM01 dataset

图 5 中展示了检索结果的可视化。模型测试以单查询模式进行，选择红外图像作为查询图，可见光图像作为图库。同时，对于每个查询图，图库中最多有 4 个具有相同 ID 标签的图像。查询过程中，网络对于每个查询的图像，都将其与图库中的每个图像进行相似度比较。通过对相似度结果进行排序，模型选取相似度最高的前 20 张图像，判断是否包含正例样本。正例样本被绿色边框包围，而负例样本被红色边框包围。第一行是排序结果，第二行是图库中属于查询图像身份的所有身份图像。

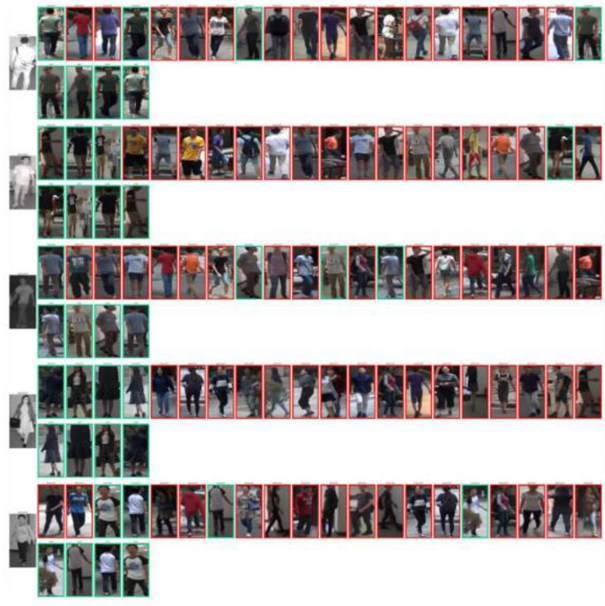

图 5 行人重识别检索结果可视化

Fig. 5 Visualization of pedestrian re-identification search results

**结束语** 本文工作针对无监督跨模态行人重识别任务提出了一个有效的单阶段模型。该模型首先使用聚类算法对无标签数据赋值伪标签，并通过置信因子分析聚类簇的稳定性，筛选更值得信赖的聚类簇，有效提高无监督训练的性能。其次，本工作针对跨模态样本提出了基于图嵌入的处理方法，在特征空间中对向量建立图关系矩阵，并设计了 4 种不同的图嵌入结构表达方法，分别从不同角度对跨模态样本实现约束。最终，通过大量的实验验证了该模型的有效性，并对未来仍需改进的地方进行了展望。总体来讲，当前针对无监督跨模态行人重识别的研究较少，对此本文工作进行大胆的尝试，并取得了有竞争力的结果。

## 参 考 文 献

[1] LIN Y, DONG X, ZHENGL, et al. A bottom-up clustering approach to unsupervised person reidentification[C]// AAAI. 2019:8738-8745.

[2] LIN Y, XIE L, WU Y, et al. Unsupervisedperson re-identificationvia softened similaritylearning[C]// CVPR. 2020: 3390-3399.

[3] ZENG K, NING M, WANG Y, et al. Hierarchical Clustering With Hard-Batch Triplet Loss for Person Re-Identification [C]// CVPR. 2020:13657-13665.

[4] YANG B, YE M, CHEN J, et al. Augmented Dual Contrastive Aggregation Learning for Unsupervised Visible-Infrared Person Re-Identification[C]// Proceedings of the 30th ACM International Conference on Multimedia(MM'22). 2022.

[5] LIANG W, WANG G, LAIJ, et al. Homogeneous-to-Heteroge-

neous: Unsupervised Learning for RGB-Infrared Person Re-Identification[C]// IEEE Transactions on Image Processing. 2021:6392-6407.

[6] HARDOON D R, SZEDMAK S, SHAWE-TAYLOR J. Canonical correlation analysis: An overview with application to learning methods [J]. Neural Computation, 2004, 16(12): 2639-2664.

[7] RASIWASIA N, PEREIRA J C, COVIELLO E, et al. A new approach to cross-modal multimedia retrieval[C]// ACM MM, 2010: 251-260.

[8] ANDREW G, ARORA R, BILMES J, et al. Deep canonical correlation analysis[C]// ICML, 2013: 1247-1255.

[9] FENG F, WANG X, LI R. Cross-modal retrieval with correspondence autoencoder[C]// ACM MM. 2014: 7-16.

[10] ZHAN Y, YU J, YU Z, et al. Comprehensive distance-preserving autoencoders for cross-modal retrieval[C]// ACM MM. 2018: 1137-1145.

[11] HE L, XU X, LU H, et al. Unsupervised cross-modal retrieval through adversarial learning[C]// ICME. 2017: 1153-1158.

[12] CHUNG Y A, WENG W H, TONG S, et al. Unsupervised cross-modal alignment of speech and textembedding spaces [C]// Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS'18). 2018: 7365-7375.

[13] LI C, DENG C, LI N, et al. Self-super visedadversarial hashing networks fo rcross modal retrieval[C]// CVPR. 2018: 4242-4251.

[14] WANG B, YANG Y, XU X, et al. Adversarial cross-modal retrieval[C]// ACM MM, 2017: 154-162.

[15] YAN S, XU D, ZHANG B, et al. Graph embedding and extensions: A general framework for dimensionality reduction [J]. IEEE TPAMI, 2006, 29(1): 40-51.

[16] MORSING L H, SHEIKH-OMARO A, IOSIFIDIS A. Supervised domain adaptation: A graph embedding perspective and a rectified experimental protocol [J]. arXiv: 2004. 11262, 2020.

[17] MOON H, PHILLIPSP J. Computational and performance aspects of PCA-based face-recognition algorithms [J]. Perception, 2001, 30(3): 303-321.

[18] YU S, LI S, CHEN D, et al. COCAS: A Large-Scale Clothes Changing Person Dataset for ReIdentification[C]// CVPR. 2020: 3400-3409.

[19] SONG L, WANG C, ZHANG L, et al. Unsupervised domain adaptive re-identification: Theory and practice [J]. Pattern Recognition, 2020, 102: 107173.

[20] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]// ICCV. 2017: 2223-2232.

[21] FU Y, WEI Y, WANG G, et al. Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification[C]// ICCV. 2019: 6112-6121.

[22] ZHONG Z, ZHENG L, LUO Z, et al. Invariance matters: Exemplar memory for domain adaptive person re-identification[C]// CVPR. 2019: 598-607.

[23] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]// CVPR. 2005: 886-893.

[24] XIONG F, GOU M, CAMPS O, et al. Person re-identification using kernel-based metric learning methods[C]// ECCV. 2014: 1-16.

[25] MATSUKAWA T, OKABE T, SUZUKIE, et al. Hierarchical gaussian descriptor for person re-identification[C]// CVPR. 2016: 1363-1372.

**LOU Ren**, born in 1982, bachelor, senior engineer. His main research interests include transportation Internet of Things and artificial intelligence.

**ZHAO Sanyuan**, born in 1985, Ph.D, associate professor. Her main research interests include computer vision, deep learning and virtual reality.