



# 基于特征细化的多标签学习无监督行人重识别

陈元妹,王凤随,钱亚萍,王路遥

(安徽工程大学,a.电气工程学院;b.高端装备先进感知与智能控制教育部重点实验室,安徽芜湖 241000)

**摘要:**针对无监督行人重识别中行人特征表达不充分以及训练过程产生噪声标签等问题,提出了一种基于特征细化的多标签学习无监督行人重识别方法。首先,为提高网络对关键区域信息的利用能力,设计多尺度通道注意力模块(Multi-scale channel attention module, MCAM),嵌入 ResNet50 网络的不同层来构建特征细化网络,并利用该网络对输入图像通道维度上的关键信息进行强化和关注,以获得更丰富的特征信息;其次,为降低训练过程中产生的噪声标签对网络的负面影响,设计多标签学习模块(Multi-label learning module, MLM),通过该模块进行正标签预测以生成可靠的伪标签;最后,利用多标签分类损失和对比损失进行无监督学习。在数据集 Market-1501 和 DukeMTMC-reID 上进行实验,结果表明该方法在这两个数据集上的平均精度均值分别达到 82.8%和 70.9%,首位命中率分别达到 92.9%和 83.9%。该方法使用注意力机制强化图像的特征信息,并通过正标签预测减少噪声标签,有效提升了无监督行人重识别的准确率,为无监督行人重识别领域提供了更鲁棒的方法。

**关键词:**行人重识别;无监督;特征细化;多尺度通道注意力;多标签学习

**中图分类号:** TP391.4

**文献标志码:** A

**文章编号:** 1673-3851 (2023) 11-0755-09

**引文格式:**陈元妹,王凤随,钱亚萍,等.基于特征细化的多标签学习无监督行人重识别[J].浙江理工大学学报(自然科学),2023,49(6):755-763.

**Reference Format:** CHEN Yuanmei, WANG Fengsui, QIAN Yaping, et al. Multi-label learning unsupervised person re-identification based on feature refinement[J]. Journal of Zhejiang Sci-Tech University, 2023, 49(6): 755-763.

## Multi-label learning unsupervised person re-identification based on feature refinement

CHEN Yuanmei, WANG Fengsui, QIAN Yaping, WANG Luyao

(a. School of Electrical Engineering; b. Key Laboratory of Advanced Perception and Intelligent Control of High-end Equipment, Ministry of Education, Anhui Polytechnic University, Wuhu 241000, China)

**Abstract:** Aiming at the issue of inadequate expression of person features and noise labels generated in the training process in unsupervised person re-identification, we proposed a multi-label learning unsupervised person re-identification method based on feature refinement. Firstly, to improve the network's ability to use key area information, a multi-scale channel attention module (MCAM) was designed. We embedded it into different layers of ResNet50 to construct a feature refinement network. This network was used to strengthen and focus the information on the channel dimension of the input image to obtain richer feature descriptions. Secondly, to reduce the detrimental effects of noise labels produced during network training, we designed a multi-label learning module (MLM). Positive label prediction was performed through this module to generate reliable pseudo-labels. Finally, unsupervised learning was carried out by using multi-label classification loss combined with contrast loss. We conducted experiments

收稿日期: 2023-04-14 网络出版日期: 2023-07-07

基金项目: 安徽省自然科学基金项目(2108085MF197);安徽高校省级自然科学研究重点项目(KJ2019A0162);安徽工程大学国家自然科学基金预研项目(Xjky2022040)

作者简介: 陈元妹(1998—),女,安徽宿州人,硕士研究生,主要从事智能信息处理方面的研究。

通信作者: 王凤随, E-mail: fswang@ahpu.edu.cn

on Market-1501 and DukeMTMC-reID datasets. The results show that the Rank-1 hit rate is 92.9% and 83.9%, while the mean average precision reaches 82.8% and 70.9%, respectively. This method uses the attention mechanism to enhance the feature information of the image and reduces the noise label by positive label prediction. It effectively improves the accuracy of unsupervised person re-identification and provides a more robust method for unsupervised person re-identification fields.

**Key words:** person re-identification; unsupervised; features refinement; multi-scale channel attention; multi-label learning

0 引 言

行人重识别任务旨在不同的摄像机中识别出同一行人,这项任务在预防犯罪和维护公共安全方面具有重要意义<sup>[1-3]</sup>。目前在行人重识别方法中,有监督行人重识别的精准度相对较高,但该方法需要完整的真实标签,成本高昂且标注耗时;无监督行人重识别可以弥补有监督行人重识别需要真实标签的不足,能够从未标记的数据集中学习行人身份的类型差异特征和类内相似特征,降低了标注成本,更适合实际应用。

基于深度学习的行人重识别方法通常利用卷积神经网络(Convolutional neural network, CNN)来提取深层次的特征<sup>[4-5]</sup>。但由于图像模糊、行人姿态、行人遮挡、光线等因素的影响, CNN 无法从行人图像中精确提取关键区域信息,导致行人重识别精度较低。因此,研究基于深度学习的行人重识别方法,首要任务是提升复杂场景下行人图像的特征提取能力,以减少网络训练过程中的噪声标签,提高训练速度和精度。

目前已有大量基于深度学习的行人重识别研究。刘紫燕等<sup>[6]</sup>提出了一种基于注意力机制的行人重识别方法,利用注意力机制缓解了环境因素的影响,有效提取了整个行人的特征。Cho 等<sup>[7]</sup>提出了一种基于部件的伪标签细化(Part-based pseudo label refinement, PPLR)框架,将交叉一致性分数作为特征相似性,并通过部分特征的预测信息来细化全局特征,以此减少全局特征聚类中的噪声标签。Ge 等<sup>[8]</sup>提出了一种自步对比学习(Self-paced contrastive learning, SPCL)框架,结合源域和目标域信息进行联合特征学习;与其他对比学习框架不同的是,SPCL 可以为源域类级、目标域簇级和非簇实例级提供监督信号,并动态更新混合内存,充分挖掘了所有训练数据的信息,提高了目标重识别准确率。为充分获取行人信息, Lin 等<sup>[9]</sup>提出了一种自下而上的聚类方法(Bottom up clustering, BUC),在聚类过程中使用多样性正则项来平衡每个聚类中

的样本数量,实现了多样性和相似性的有效平衡。孙义博等<sup>[10]</sup>提出了一种基于通道注意力机制的行人重识别方法,有效提取了更鲁棒的行人特征,使设计的网络模型达到了更高的识别精度。Yu 等<sup>[11]</sup>提出了一种基于深度聚类的无监督非对称距离度量学习方法,该方法有效降低了由于光线、遮挡等问题引起的图像偏差,从而有利于无监督行人重识别网络挖掘潜在的行人特征信息。Zhang 等<sup>[12]</sup>为充分挖掘行人图像中的判别性信息,并保持跨域标签的一致性,提出了一种基于软多标签和复合注意力模块的无监督行人重识别方法。Xuan 等<sup>[13]</sup>提出了一种相机内和相机间的相似度计算(Intra-inter camera similarity, IICS)方法,相机内的相似度直接利用 CNN 提取每个相机内的特征,相机间的相似度利用每个样本在不同相机上的分类分数计算;这种相似度计算方法有效缓解了相机间的样本分布差异问题,使网络生成更可靠的伪标签。以上研究虽然取得了一定的成果,可以在复杂背景下提取行人的特征信息,但忽略了行人图像中的细微特征。此外,如果无监督行人重识别网络模型没有充分提取样本中的特征信息,易产生噪声伪标签,使得行人重识别准确率降低。

为充分提取行人图像中的关键信息,减少训练过程中产生的噪声标签,本文提出了一种基于特征细化的多标签学习无监督行人重识别方法。首先设计了多尺度通道注意力模块(Multi-scale channel attention module, MCAM),使得行人重识别模型对全局和局部通道上特征进行加权 and 强化;其次,设计了多标签学习模块(Multi-label learning module, MLM),利用 MLM 模块进行正标签预测,以减少噪声标签的影响;最后,利用多损失联合监督网络学习,提高行人重识别模型的训练效率。本文提出的方法强化了行人图像中全局和局部通道中的重要信息,在有效提取行人关键特征的基础上通过 MLM 模块提高了伪标签的质量;无监督行人重识别网络模型在训练过程中无需额外的信息标注,在节约成本的同时提高了无监督行人重识别的准确率。

1 本文方法

1.1 整体网络结构

本文提出的基于特征细化的多标签学习无监督

行人重识别方法,采用 ResNet50 作为基础网络,并在基础网络的 Layer3、Layer4 之间和 Layer4、Layer5 之间添加 MCAM 模块作为主干网络,整体网络结构示意图如图 1 所示。

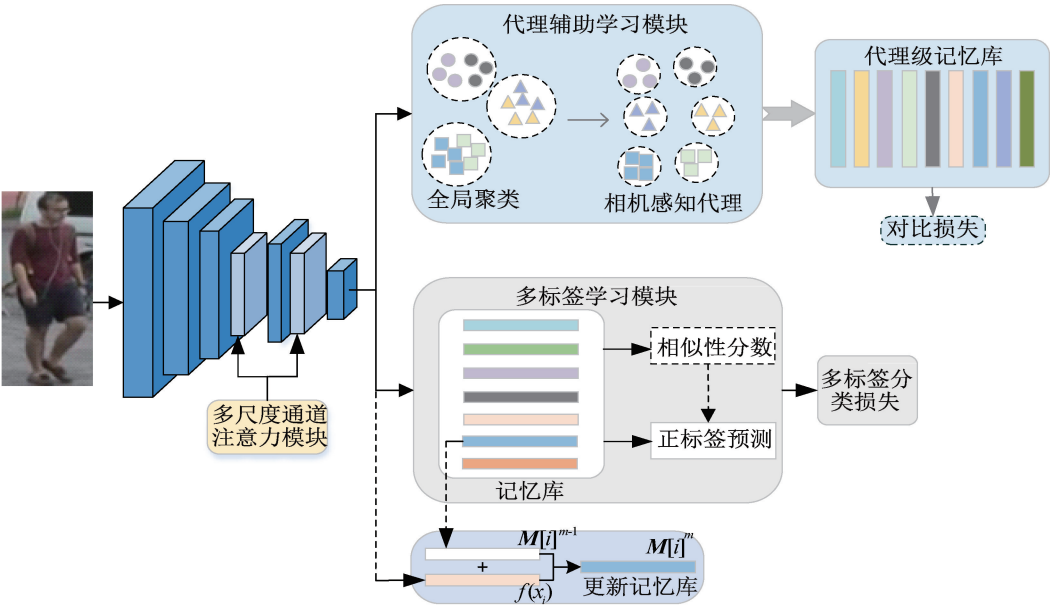


图 1 整体网络结构示意图

整体网络结构主要分为两个模块:代理辅助学习模块和多标签学习模块。

在代理辅助学习模块,首先对网络提取的特征进行全局聚类;其次,将每个聚类作为多个相机感知代理,并生成一组新的伪标签;最后,将代理特征存储代理级记忆库中,并利用代理特征更新代理级记忆库,同时计算对比损失。

在多标签学习模块中,将主干网络提取的 2048 维特征存储在记忆库中,并计算输入特征与其他特征之间的相似性分数。同时根据相似性分数进行正

标签预测,提高伪标签质量。最后计算多标签分类损失,多标签分类损失在迭代训练过程中不断更新网络。

1.2 多尺度通道注意力模块

MCAM 模块的核心思想是通过添加全局平均池化层来实现多个尺度上的通道注意。MCAM 模块由全局和局部通道上下文构成,其中全局通道又分为两个分支。一个分支不改变通道维度,另一个分支先降维后升维,从而学习到不同维度的特征图。多尺度通道注意力模块结构示意图如图 2 所示。

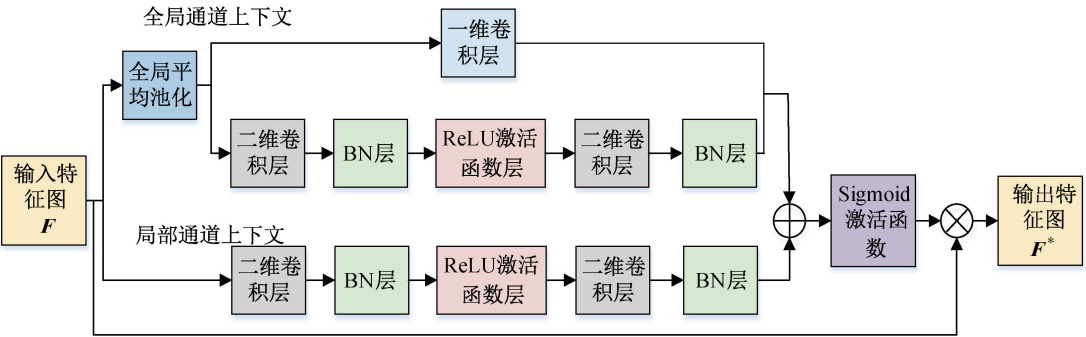


图 2 多尺度通道注意力模块结构示意图

将 MCAM 模块的输入特征图记为  $F \in \mathbf{R}^{H \times W \times C}$ ,输出特征图记为  $F^* \in \mathbf{R}^{H \times W \times C}$ 。

在全局通道上下文结构中,首先,将特征图  $F$  输入全局平均池化层 (Global average pooling, GAP),得到  $1 \times 1 \times C$  的特征描述。其次,将  $1 \times 1 \times$

$C$  的特征信息分别输入两个分支。其中一个分支仅经过一维卷积层,另一个分支依次经过二维卷积层、BN 层、ReLU 激活函数层、二维卷积层、BN 层。最后,输出全局通道上下文信息  $g$ ,  $g$  可用式 (1) 计算:

$$g = V_{\text{Conv1}}(G(\mathbf{F})) + B(V_{\text{Conv22}}(\delta(B(V_{\text{Conv21}}(G(\mathbf{F})))))) \quad (1)$$

其中:  $V_{\text{Conv1}}()$  表示一维卷积层,  $G()$  表示全局平均池化,  $B()$  表示 BN 层,  $V_{\text{Conv21}}()$ 、 $V_{\text{Conv22}}()$  均表示二维卷积层,  $\delta()$  表示 ReLU 激活函数层。

在局部通道上下文结构中, 首先, 将  $H \times W \times C$  的特征图  $\mathbf{F}$  输入  $V_{\text{Conv21}}$ , 由于  $V_{\text{Conv21}}$  的内核大小为  $1 \times 1 \times (C/r) \times C$ , 其输出特征图的大小变为  $H \times W \times C/r$ 。其次, 将  $H \times W \times C/r$  的特征图输入  $V_{\text{Conv22}}$ , 由于  $V_{\text{Conv22}}$  的内核大小为  $1 \times 1 \times C \times (C/r)$ , 其输出的特征图大小变为原来的  $H \times W \times C$ 。可以看出, 局部通道上下文的输出特征图大小与输入特征图大小相同, 它可以保留输入特征图中的细微细节。最后, 输出局部通道上下文信息  $l$ ,  $l$  可用式(2)计算:

$$l = B(V_{\text{Conv22}}(\delta(B(V_{\text{Conv21}}(\mathbf{F})))))) \quad (2)$$

将全局通道上下文信息  $g$  和局部通道上下文信息  $l$  进行叠加, 通过 Sigmoid 激活函数输出注意权重, 然后将注意权重值与输入的特征图  $\mathbf{F}$  做乘法运算, 输出细化特征  $\mathbf{F}^*$ :

$$\mathbf{F}^* = \mathbf{F} \otimes \sigma(g \oplus l) \quad (3)$$

其中:  $\sigma$  表示 Sigmoid 激活函数,  $\otimes$  表示元素级乘法,  $\oplus$  表示元素级加法。

### 1.3 多标签学习模块

给定未标记的行人图像数据集  $D = \{x_i\}_{i=1}^N$ , 其中:  $x_i$  代表第  $i$  个行人图像,  $N$  代表图像数。由于在正标签预测中需要未标记行人图像的特征信息, 本文使用记忆库  $\mathbf{M} \in \mathbf{R}^{n \times d}$  对其进行存储, 其中:  $n$  表示图像数,  $d$  表示特征维度。记忆库  $\mathbf{M}$  由主干网络提取的特征进行初始化, 并在训练期间不断更新, 更新方式为:

$$\mathbf{M}[i]^m = \mu f(x_i) + (1 - \mu) \mathbf{M}[i]^{m-1} \quad (4)$$

其中:  $m$  表示第  $m$  个训练时期,  $\mu$  为更新速率,  $f(x_i)$  表示提取主干网络的特征。

为能够在数据集  $D$  上进行训练, 首先为每个行人图像分配单标签  $y_i$ , 将  $y_i$  作为正标签预测的输入信号, 并输出基于记忆库  $\mathbf{M}$  的多标签预测。在正标签预测过程中, 本文需要计算输入特征与其他特征之间的相似性分数, 根据相似性分数得到排名列表  $R_i$ , 计算公式为:

$$s_{i,j} = \mathbf{M}[i]^T \times \mathbf{M}[j] \quad (5)$$

$$R_i = \text{argsort}(s_{i,j}) \quad (6)$$

其中:  $R_i$  为匹配输入特征的候选标签, 即相似性分数靠前的标签。

为保证标签的预测质量, 减少训练过程中的噪声标签, 受 Xie 等<sup>[14]</sup> 的启发, 本文采用  $k$  倒数最邻

近方法从候选标签中选择正类标签, 并过滤难以正确分类的负类标签。正标签集的计算公式为:

$$P_i^* = P_i[1:t] \quad (7)$$

其中:  $P_i$  是通过给定的相似性阈值筛选出的候选标签集合,  $t$  表示正标签集中的标签数量。

### 1.4 代理辅助学习模块

本文在代理辅助学习模块中, 将每个全局聚类划分为多个相机特定的感知代理, 每个代理都代表来自同一台相机的实例, 并生成一组新的伪标签。用  $D_c = \{(x_i, q_i, z_i, c_i)\}_{i=1}^{N_c}$  表示第  $c$  台相机的新标签数据集, 其中:  $x_i$  代表第  $i$  个行人图像;  $q_i$  表示全局聚类生成的伪标签;  $z_i \in \{1, \dots, Z_c\}$  表示相机感知代理生成的伪标签;  $c_i = c \in \{1, \dots, C\}$  表示相机标签标注,  $C$  表示相机数量;  $N_c$  和  $Z_c$  分别代表相机  $c$  中的图像数量和相机感知代理数量。则整个标记数据集表示为  $D' = \bigcup_{c=1}^C D_c$ 。

此外, 本文在代理辅助学习过程中使用代理级记忆库  $\mathbf{M}' \in \mathbf{R}^{d \times Z}$  存储特征, 其中:  $d$  表示特征维度,  $Z$  表示所有相机中的感知代理总数。代理级记忆库的每个条目存储一个相机感知代理, 并利用代理图像对其进行更新, 更新方式为:

$$\mathbf{M}'[j] = a\mathbf{M}'[j] + (1-a)f(x_i) \quad (8)$$

其中:  $a$  表示动态更新速率。

### 1.5 损失函数

本文基于代理辅助学习方法, 使用基于记忆库的代理特征构造对比损失。相机间的对比损失  $L_{\text{Inter}}$  计算公式为:

$$L_{\text{Inter}} = - \sum_{i=1}^{N'} \frac{1}{|P|} \sum_{p \in P} \log \frac{S(p, x_i)}{\sum_{u \in P} S(u, x_i) + \sum_{q \in Q} S(q, x_i)} \quad (9)$$

其中:  $P$  和  $Q$  分别代表正代理集和难以正确分类的负代理集,  $N'$  为所选聚类中包含的图像数,  $S(p, x_i) = \exp(\mathbf{M}'[p]^T f(x_i)/\tau)$ 。

相机内的对比损失  $L_{\text{Intra}}$  计算公式为:

$$L_{\text{Intra}} = - \sum_{c=1}^C \frac{1}{N_c} \sum_{i \in D_c} \log \frac{\exp(\mathbf{M}'[j]^T f(x_i)/\tau)}{\sum_{k=A+1}^{A+Z_{c_i}} \exp(\mathbf{M}'[k]^T f(x_i)/\tau)} \quad (10)$$

其中:  $A = \sum_{c=1}^{c-1} Z_c$  为从第一台相机累加到  $c_i - 1$  台相机的总代理数量,  $\tau$  为温度因子。

总对比损失  $L_v$  计算公式为:

$$L_v = L_{\text{Inter}} + \lambda L_{\text{Intra}} \quad (11)$$

其中:  $\lambda$  是两项损失的平衡因子。

在 MLM 模块中,考虑到存在样本不平衡的问题,本文采用多标签分类损失深度挖掘难以正确分类的负类样本,以此提高行人重识别模型在 MLM 模块中的训练效率。多标签分类损失  $L_{\text{MCL}}$  计算公式为:

$$L_{\text{MCL}} = \sum_{i=1}^n \frac{\omega}{|P_i^*|} \sum_{p \in P_i^*} \mathbb{1}(p | x_i) + \frac{1}{|N_i|} \sum_{s \in N_i} \mathbb{1}(s | x_i) \quad (12)$$

其中:  $N_i$  是难以正确分类的负类标签集,  $|N_i| = (n - |P_i^*|) \times 1\%$ ,  $n$  为可靠的候选标签数,并选择前 1% 的类作为难以正确分类的负类标签;  $\omega$  为衡量多标签分类损失的参数。

最终,总损失计算公式为:

$$L_{\text{Total}} = L_v + \beta L_{\text{MCL}} \quad (13)$$

其中:参数  $\beta$  是两种损失的平衡因子。

## 2 实验与分析

### 2.1 实验数据与评价指标

本文在 Market-1501<sup>[15]</sup> 和 DukeMTMC-reID<sup>[16]</sup> 两个数据集上进行实验。Market-1501 数据集由 6 个不同的摄像机中捕获的图像组成,共有 1501 个行人身份,其中:训练集行人数量为 751 人,包含 12936 张行人图像;测试集行人数量为 750 人,包含 19732 张行人图像。DukeMTMC-reID 数据集由 8 个不同的摄像机中捕获的图像组成,共有 1404 个行人身份,其中:训练集行人数量为 702 人,包含 16522 张行人图像;测试集行人数量为 702 人,包含 17661 张行人图像。

本文采用累积匹配特性(Cumulative matching characteristics, CMC)曲线和平均精度均值(Mean average precision, mAP)评估模型性能。CMC 表示识别准确率最高的前  $n$  张图像的准确率,记为 Rank- $n$ 。本文使用 Rank-1、Rank-5、Rank-10 作为主要评价指标,其中 Rank-1 也称为首位命中率;用  $R_1$ 、 $R_5$ 、 $R_{10}$  分别表示 Rank-1、Rank-5、Rank-10 值。mAP 是所有类别平均识别精度的均值,反映数据库中所有行人重识别正确的图像排在识别准确率顺序列表前面的程度。本文 Rank-1 和 mAP 均使用百分制表示。mAP 的计算公式为:

$$P_{\text{mAP}}/\% = \frac{\sum_{i=1}^n P_{A,i}}{C} \times 100 \quad (14)$$

$$P_{A,i} = \sum_{k=1}^n P(k) \Delta r(k) \quad (15)$$

其中:  $P_{\text{mAP}}$  表示 mAP 值;  $C$  表示类别数;  $P_{A,i}$  表示第  $i$  个类别的平均精度;  $P(k)$  代表 top- $k$  检索结果的精准率;  $\Delta r(k) = R(k) - R(k-1)$ ;  $R(k)$  为 top- $k$  结果的召回率,  $R(0) = 0$ 。每次实验采用随机测试集重复 10 次,计算平均性能。

### 2.2 实验设置

本文实验选用的 GPU 为 NVIDIA GeForce RTX 2080Ti(11 GiB),操作系统为 Ubuntu 16.04,处理器为英特尔 Core i9-10900@3.7 GHz,深度学习框架为 Fytorch 1.2.0。本文采用 ImageNet 预训练的 ResNet50 作基础网络。在每个 epoch 的开始,本文计算  $k$  倒数最邻近的 Jaccard 距离,并使用 Density-based spatial clustering of applications with noise (DBSCAN) 进行全局聚类。模型的训练批次由从 8 个代理中随机抽取的 32 张图像组成,每个代理有 4 张图像。采用随机翻转、裁剪和擦除等数据增强方法。epoch 大小设置为 50,在前 10 个 epoch 采用预热方案,初始学习率为 0.00035,每 20 个 epoch 后除以 10。正标签预测中相似性阈值设置为 0.6,超参数  $\mu = \alpha = 0.5$ ,  $\tau = 0.07$ ,  $\lambda = 0.5$ ,  $\omega = 5$ 。总损失函数采用对比损失和多标签分类损失,前 5 个 epoch 仅使用相机内对比损失和多标签分类损失,在剩余的 epoch 中,相机间对比损失与其共同作用,训练过程中总损失曲线如图 3 所示。

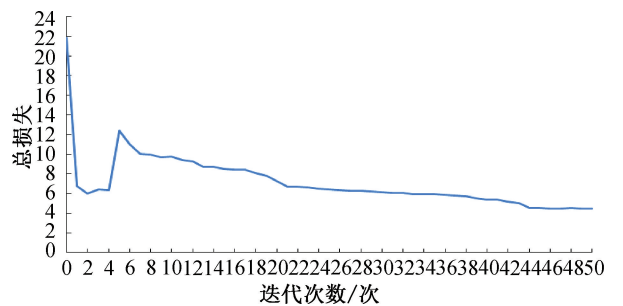


图 3 训练过程中总损失曲线

### 2.3 消融实验

#### 2.3.1 MCAM 模块嵌入层实验

为分析 MCAM 模块嵌入不同层的实验效果,本文在 Market-1501 数据集进行实验,实验结果见表 1。其中 Layer2、Layer3、Layer4、Layer5 分别表示在 ResNet50 的第 2 层、3 层、4 层、5 层之后嵌入 MCAM 模块。Layer3+Layer4 表示在第 3 层和第 4 层之后同时嵌入 MCAM 模块。

表 1 MCAM 模块嵌入层实验评价指标

模块嵌入层	$P_{mA}/\%$	$R_1/\%$
Layer2	79.0	91.0
Layer3	79.8	91.3
Layer4	79.8	91.1
Layer5	79.2	90.9
Layer3+Layer4	80.0	91.9

由表 1 可以看出,与分别在 Layer2 和 Layer5 之后嵌入 MCAM 模块相比,在 Layer3 和 Layer4 之后分别嵌入 MCAM 模块的实验结果较好,并且在 Layer3 和 Layer4 之后分别嵌入 MCAM 模块的实验结果基本相同。因此本文在 Layer3 和 Layer4 之后同时嵌入 MCAM 模块进行一次实验,结果发现,该方式比分别在 Layer3 和 Layer4 之后嵌入 MCAM 模块的效果更佳。由该实验结果可知,将 MCAM 模块同时嵌入基础网络的 Layer3 和 Layer4 之后,无监督行人重识别网络挖掘行人关键特征的能力更优。因此本文在第 3 层和第 4 层之后都嵌入 MCAM 模块,将此网络作为本文的主干特征提取网络。

2.3.2 MCAM 模块和 MLM 模块有效性验证实验

为进一步验证 MCAM 模块和 MLM 模块的有效性,在 Market-1501 数据集进行消融实验。实验结果见表 2,其中:Baseline+MCAM 表示在基线网络基础上仅添加 MCAM 模块,Baseline+MLM 表示仅添加 MLM 模块,Baseline+MCAM+MLM 为本文最终的模型。

表 2 MCAM 模块和 MLM 模块有效性验证实验评价指标

加入模块	$P_{mA}/\%$	$R_1/\%$
Baseline	78.0	90.5
Baseline+MCAM	80.0	91.9
Baseline+MLM	81.0	91.8
Baseline+MCAM+MLM	82.8	92.9

实验过程如下:首先,对基线网络进行测试,基线网络的  $P_{mA}$  为 78.0%, $R_1$  为 90.5%;其次,在基线网络 ResNet50 中嵌入 MCAM 模块,由表 2 可以看出, $P_{mA}$  和  $R_1$  分别提高了 2.0%和 1.4%;再次,在基线网络中添加 MLM 模块,其  $P_{mA}$  和  $R_1$  分别为 81.0%和 91.8%,相比基线网络分别提高了 3.0%和 1.3%;最后,将 MCAM 模块和 MLM 模块同时添加至基础网络结构中,其  $P_{mA}$  达到了 82.8%,与原先网络相比大幅提升了 4.8%, $R_1$  达到了 92.9%,在原先网络基础上提升了 2.4%。

通过实验结果可以看出,同时添加 MCAM 模块和 MLM 模块的评价指标最好,表明联合使用

MCAM 模块和 MLM 能够有效提升无监督行人重识别的精确度,本文所提出的改进的方法有效。

2.3.3 参数  $\beta$  取值实验

为探究本文总损失函数中参数  $\beta$  的取值对模型识别精度的影响,本文选取了  $\beta=0.4, 0.5, \cdots, 0.8$  等 5 个不同的数值,在 Market-1501 和 DuleMTMC-reID 数据集上进行实验,实验结果见表 3。

表 3 参数  $\beta$  取值实验评价指标

$\beta$	Market-1501		DukeMTMC-reID	
	$P_{mA}/\%$	$R_1/\%$	$P_{mA}/\%$	$R_1/\%$
0.4	82.1	92.5	70.1	83.2
0.5	82.4	92.4	70.6	83.3
0.6	82.8	92.9	70.9	83.9
0.7	82.5	92.3	70.8	83.5
0.8	82.4	92.0	70.1	83.1

表 3 的实验结果显示,当选取不同  $\beta$  值时,Market-1501 和 DuleMTMC-reID 数据集的  $P_{mA}$  和  $R_1$  均呈先升后降的趋势。当  $\beta$  取 0.6 时,两个数据集均呈现最佳效果,因此本文中  $\beta$  取 0.6。

2.4 可视化结果分析

为比较本文方法与基线方法的行人重识别效果,本文在 Market-1501 数据集上进行实验,其中行人 a、行人 b、行人 c 的识别结果如图 4 所示。图 4 中第 1 列为待识别行人;后 10 列为行人重识别准确率最高的前 10 张行人图像,即  $R_1$  至  $R_{10}$  对应的行人重识别结果。每张识别图像上方均标有数字,其中:未标方框的图像表示行人识别正确;标有方框的图像表示误识别为其他行人,即行人识别错误。

由图 4 可知,对于行人 a,应用基线方法的行人重识别结果中有 5 张图像识别错误,而应用本文方法的行人重识别结果仅出现 1 张图像识别错误;对于行人 b 和 c,应用基线方法分别出现 2 张和 3 张图像识别错误,而本文方法未出现识别错误。上述实验结果表明,本文方法可以充分提取行人图像的细节特征,减少噪声标签,有效提升了无监督行人重识别的准确率;同时该结果表明,多尺度通道注意力模块和多标签学习模块联合使用的行人重识别方法具有较强的鲁棒性。

为验证 MCAM 模块的有效性,本文利用梯度加权类激活热力图技术(Gradient-weighted class activation mapping, Grad-CAM)<sup>[17]</sup>对 MCAM 模块输出的区域注意力特征在 Market-1501 数据集上进行可视化分析,热力图如图 5 所示。

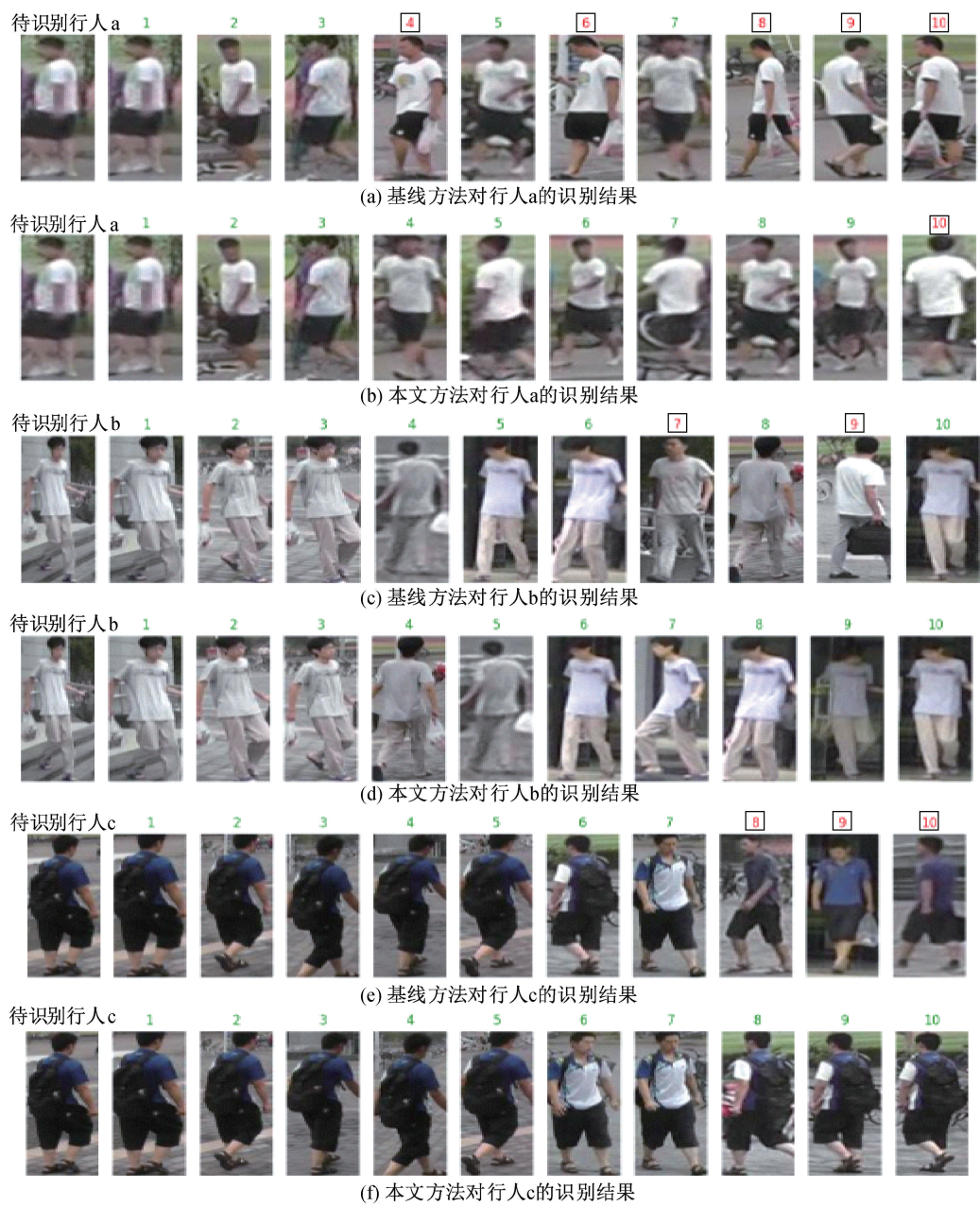


图 4 基线方法与本文方法的行人重识别结果



图 5 MCAM 模块输出区域注意力特征的可视化热力图

图 5 结果表明,MCAM 模块强化了行人图像中多样化的细节特征,如背包、手持物体、衣服图案等,提高了无监督行人重识别网络的特征提取能力。

2.5 与其他无监督行人重识别方法对比实验

本文方法与其他典型无监督行人重识别方法在

Market-1501 和 DukeMTMC-reID 两个大数据集上进行了对比,对比方法包括 ACAN-GRL<sup>[18]</sup>、CVSE<sup>[19]</sup>、MSC-GDC<sup>[20]</sup>、CAP<sup>[21]</sup>、CAPL<sup>[22]</sup>等,对比结果见表 4,表身中的空白表示文献中没有报告相关数据。

表 4 与其他无监督行人重识别方法对比试验评价指标

方法	Market-1501				DukeMTMC-reID			
	$P_{mA}/\%$	$R_1/\%$	$R_5/\%$	$R_{10}/\%$	$P_{mA}/\%$	$R_1/\%$	$R_5/\%$	$R_{10}/\%$
Attention+RAAdam	46.7	71.6	86.1		52.3	69.4	81.5	
ACAN-GRL	50.6	73.3	87.6	91.8	46.6	65.1	80.6	85.1
CVSE	63.2	84.1	92.8	95.0	56.1	75.3	82.9	85.4
IICS	72.9	89.5	95.2	97.0	64.4	80.0	89.0	91.6
SpCL	73.1	88.1	95.1	97.0				
MSC-GDC	76.1	90.1	95.7	97.5	66.4	80.1	89.9	92.3
CAP	79.2	91.4	96.3	97.7	67.3	81.1	89.3	91.8
CAPL	80.4	92.8	97.3		67.1	81.7	90.3	
本文方法	82.8	92.9	97.0	97.9	70.9	83.9	91.7	93.8

由表 4 可以看出,在同一评估指标下,本文方法相较于其他方法取得了更好的识别结果。将本文方法与 Attention+RAAdam 方法相比,在 Market-1501 和 DukeMTMC-reID 数据集下, $P_{mA}$  与  $R_1$  均大幅提升。Attention+RAAdam 方法与本文方法均使用注意力机制来辅助网络关注行人图像中的关键信息。Attention+RAAdam 方法仅对全局信息进行关注,而本文的 MCAM 模块对全局和局部通道上的关键信息同时施加关注,获得了更加丰富的行人特征信息。本文方法与 IICS 方法相比, $P_{mA}$  分别提高了 9.9% 和 6.5%, $R_1$  分别增加了 3.4% 和 3.9%。与 CAP 方法相比, $P_{mA}$  均提高了 3.6%, $R_1$  分别增加了 1.5% 和 2.8%。CAP 方法与本文方法均采用相机感知代理学习策略,不同的是,本文将相机感知代理学习与多标签学习模块置于同一架构进行联合学习,通过进行正标签预测减少噪声标签,提高伪标签质量。CVSE、MSC-GDC 方法在两个数据集上的识别准确率都远低于本文方法,进一步验证了本文所提方法的有效性和先进性。

3 结 语

本文提出了一种基于特征细化的多标签学习无监督行人重识别方法。在特征提取过程中使用多尺度通道注意力模块融合了全局和局部通道信息,更加关注行人重识别网络所需要的信息;在多标签学习中通过标签预测生成了可靠伪标签,有效提高了行人重识别的准确率。实验结果显示,Market-1501 和 DukeMTMC-reID 数据集的平均精度均值分别

达到了 82.8% 和 70.9%,首位命中率分别达到 92.9% 和 83.9%,表明本文方法能够在充分提取行人特征的同时减少噪声标签,提升了无监督行人重识别模型的性能。

为进一步提升行人重识别网络提取特征的能力,后续研究可考虑同时使用空间注意力和通道注意力,使行人重识别网络模型分别从位置和通道上关注行人图像中的关键信息,更加准确地定位到目标区域,从而获得更丰富的特征信息。

参考文献:

[1] 黄新宇,许娇龙,郭纲,等. 基于增强聚合通道特征的实时行人重识别[J]. 激光与光电子学进展, 2017, 54 (9): 119-127.

[2] 罗浩,姜伟,范星,等. 基于深度学习的行人重识别研究进展[J]. 自动化学报, 2019, 45(11): 2032-2049.

[3] Wang Y Y, Li X A, Jiang M X, et al. Cross-view pedestrian clustering via graph convolution network for unsupervised person re-identification [J]. Journal of Intelligent & Fuzzy Systems, 2020, 39(3): 4453-4462.

[4] Kim G, Shu D W, Kwon J. Robust person re-identification via graph convolution networks [J]. Multimedia Tools and Applications, 2021, 80 (19): 29129-29138.

[5] 潘海鹏,郝慧,苏雯. 基于注意力机制与多尺度特征融合的人脸表情识别[J]. 浙江理工大学学报(自然科学版), 2022, 47(3): 382-388.

[6] 刘紫燕,万培佩. 基于注意力机制的行人重识别特征提取方法[J]. 计算机应用, 2020, 40(3): 672-676.

- [7] Cho Y, Kim W J, Hong S, et al. Part-based pseudo label refinement for unsupervised person re-identification [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA. IEEE, 2022; 7308-7318.
- [8] Ge Y X, Zhu F, Chen D P, et al. Self-paced contrastive learning with hybrid memory for domain adaptive object re-ID [C] // Proceedings of the 34th International Conference on Neural Information Processing Systems. Vancouver, BC, Canada. New York; ACM, 2020; 11309-11321.
- [9] Lin Y T, Dong X Y, Zheng L A, et al. A bottom-up clustering approach to unsupervised person re-identification[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2019, 33(1): 8738-8745.
- [10] 孙义博, 张文靖, 王蓉, 等. 基于通道注意力机制的行人重识别方法[J]. 北京航空航天大学学报, 2022, 48(5): 881-889.
- [11] Yu H X, Wu A C, Zheng W S. Unsupervised person re-identification by deep asymmetric metric embedding [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(4): 956-973.
- [12] Zhang B H, Zhu S Y, Zhou Y F, et al. A novel unsupervised person re-identification algorithm based on soft multi-label and compound attention model[J]. Multimedia Tools and Applications, 2022, 81(17): 24081-24098.
- [13] Xuan S Y, Zhang S L. Intra-inter camera similarity for unsupervised person re-identification[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, TN, USA. IEEE, 2021; 11921-11930.
- [14] Xie K, Wu Y, Xiao J, et al. Unsupervised person re-identification via  $K$ -reciprocal encoding and style transfer[J]. International Journal of Machine Learning and Cybernetics, 2021, 12(10): 2899-2916.
- [15] Zheng L, Shen L Y, Tian L, et al. Scalable person re-identification: a benchmark [C] // 2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile. IEEE, 2016; 1116-1124.
- [16] Zheng Z D, Zheng L, Yang Y. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro [C] // 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy. IEEE, 2017; 3774-3782.
- [17] Selvaraju R R, Cogswell M, Das A, et al. Grad-cam: Visual explanations from deep networks via gradient-based localization [C] // 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy. IEEE, 2017; 618-626.
- [18] Qi L, Wang L, Huo J, et al. Adversarial camera alignment network for unsupervised cross-camera person re-identification [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(5): 2921-2936.
- [19] Zhou S R, Wang Y, Zhang F, et al. Cross-view similarity exploration for unsupervised cross-domain person re-identification [J]. Neural Computing and Applications, 2021, 33(9): 4001-4011.
- [20] Pang Z Q, Guo J F, Ma Z Q, et al. Median stable clustering and global distance classification for cross-domain person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2022, 32(5): 3164-3177.
- [21] Wang M L, Lai B S, Huang J Q, et al. Camera-aware proxies for unsupervised person re-identification[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2021, 35(4): 2764-2772.
- [22] Liu Y X, Ge H W, Sun L, et al. Camera-aware progressive learning for unsupervised person re-identification[J]. Neural Computing and Applications, 2023, 35(15): 11359-11371.

(责任编辑:康 锋)