Opto-Electronic Engineering
光 电 工 程

Article

2020 年，第 47 卷，第 12 期

DOI: 10.12086/oee.2020.190636

软多标签和深度特征融合的无监督行人重识别

张宝华[1,3*]，朱思雨[1]，吕晓琪[2,3]，谷　宇[1,3]，王月明[1,3]，
刘　新[1,3]，任　彦[1]，李建军[1,3]，张　明[1,3]

[1]内蒙古科技大学信息工程学院，内蒙古自治区 包头 014010；
[2]内蒙古工业大学信息工程学院，内蒙古自治区 呼和浩特 010051；
[3]内蒙古自治区模式识别与智能图像处理重点实验室，内蒙古自治区 包头 014010

摘要：跨摄像头场景中依赖面向标签映射关系的学习以提高识别精度，有监督行人重识别模型虽然识别精度较好，但存在可扩展问题，诸如算法识别精度严重依赖有效的监督信息，算法实时性差等；针对上述问题，提出一种基于软多标签的无监督行人重识别算法。为了提高标签匹配精度，首先利用软多标签逼近真实标签，通过计算参考数据集和参考代理在软多标签函数中的损失函数，预训练参考数据集，并构建预训练与训练结果的映射模型。再通过生成数据和真实数据分布的最小距离的期望即简化的 2-Wasserstein 距离计算相机视图中软多标签均值和标准差得到损失函数，解决跨视域标签一致性问题。为了提高软多标签对未标记目标数据集的有效性，计算联合嵌入损失，挖掘不同类别间的相似对，纠正跨域分布错位。针对残差网络训练时长和无监督学习精度低的问题，通过结合压缩激励网络(SENet)和多层级深度特征融合改进残差网络的结构，提高训练速度和精度。实验结果表明，该方法在标准数据集下的首位命中率和平均精度均值优于先进相关算法。

关键词：残差网络；行人重识别；软多标签；无监督；深度特征

中图分类号：TP391.4　　**文献标志码：**A

引用格式：张宝华，朱思雨，吕晓琪，等. 软多标签和深度特征融合的无监督行人重识别[J]. 光电工程，2020，**47**(12): 190636

Soft multilabel learning and deep feature fusion for unsupervised person re-identification

Zhang Baohua[1,3*], Zhu Siyu[1], Lv Xiaoqi[2,3], Gu Yu[1,3], Wang Yueming[1,3],
Liu Xin[1,3], Ren Yan[1], Li Jianjun[1,3], Zhang Ming[1,3]

[1]School of Information Engineering, Inner Mongolia University of Science and Technology, Baotou, Inner Mongolia 014010, China;
[2]School of Information Engineering, Mongolia Industrial University, Huhehaote, Inner Mongolia 010051, China;
[3]Inner Mongolia Key Laboratory of Pattern Recognition and Intelligent Image Processing, Baotou, Inner Mongolia 014010, China

收稿日期：2019-10-24；　收到修改稿日期：2020-03-02

基金项目：国家自然科学基金资助项目(61962046，61663036，61841204)；内蒙古杰青培育项目(2018JQ02)；内蒙古草原英才，内蒙古青年科技创新人才项目(第一层次)；内蒙古自治区自然科学基金资助项目(2015MS0604，2018MS06018)；内蒙古自治区高等学校科学技术研究项目资助(NJZY145)

作者简介：张宝华(1981-)，男，博士，教授，硕士生导师，主要从事数字图像处理及应用、目标识别与跟踪的研究。
E-mail：zbh_wj2004@imust.edu.cn

Abstract: In cross-camera scenarios, it relies on the learning of label mapping relationships to improve recognition accuracy. The supervised person re-identification model has better recognition accuracy, but there are scalability problems. For example, the accuracy of algorithm identification relies heavily on effective supervised information. When adding a small amount of data in the classification process, all data needs to be reprocessed, resulting in poor real-time performance. Aiming at the above problems, an unsupervised person re-identification algorithm based on soft label is proposed. In order to improve the accuracy of label matching, first, learn soft multilabel to make it close to the real label, and obtain the reference agent by calculating the loss function of the reference data set to achieve the purpose of pre-training the reference data set. Then, calculate the expected value of the minimum distance between the generated data and the real data distribution (using the simplified 2-Wasserstein distance), calculate the mean and standard deviation vector of the soft multilabel in the camera view, and the resulting loss function can solve cross-view domain label consistency issues. In order to improve the validity of the soft tag on the unmarked target data set, the joint embedding loss is calculated, the similar pairs between different categories are mined, and the cross-domain distribution misalignment is corrected. In view of the problem that the residual network training duration and the unsupervised learning accuracy are low, the structure of the residual network is improved by combining the SENet and fusing multi-level depth feature to improve the training speed and accuracy. The experimental results show that the rank-1 and mAP are better than advanced correlation algorithms.

Keywords: resnet; person re-identification; soft multilabel; unsupervised; depth feature

Citation: Zhang B H, Zhu S Y, Lv X Q, *et al.* Soft multilabel learning and deep feature fusion for unsupervised person re-identification[J]. *Opto-Electronic Engineering*, 2020, **47**(12): 190636

1 引 言

行人重识别(Person re-identification, ReID)主要用于跟踪跨摄像头场景中所拍摄的无重叠区域内的行人，即在摄像头所拍摄的图像中检取感兴趣的行人，然后在跨摄像头场景中检索与感兴趣行人图像相似的目标[1]。利用该技术去查找行人数据库中的嫌疑人图像，可以节省大量的时间和人力[2]。在智能安防、刑侦工作、搜寻走失人员以及图像检索等方面有良好的应用前景。应用场景包括我国的“天网行动”和“天眼工程”等。

行人重识别方法可以分为有监督学习和无监督学习两种。有监督学习中的跨视域变化、不同行人间的高度相似等问题，会降低标注精度，导致相关方法可拓展性较差[3]。无监督学习可以解决有监督模型的可扩展性问题，但目前无监督学习的识别精度低且跨摄像头图像没有映射标签，从而使得无监督行人重识别受到限制[4]。

在无监督行人重识别方面，典型方法有基于伪标签的方法和域自适应方法。在伪标签方法中 Yu 等人[5]提出了无监督非对称度量学习，旨在基于交叉视图行人图像的非对称聚类来学习非对称度量，通过模型学习找到一个共享空间，可以减小特定视图的偏差，从而实现更好的匹配性能。Fan 等人[6]提出渐进式无监督学习方法，将预训练的深度表示转移到无标记的数据集以实现更好的识别精度。但基于伪标签学习的模型通过直接比较视觉特征(例如通过 K 均值聚类)来分配伪标签，而没有关注到潜在的判别性信息。在域自适应的方法中，Wang 等人[7]提出了可转移联合属性-身份深度学习(transferable joint attribute-identity deep learning)，用于将现有数据集的标记信息转移到新的未标记目标域，无需在目标域中进行任何监督学习。Wei 等人[7]提出了一种针对行人重识别的生成对抗网络(person transfer GAN)，实现不同行人重识别数据集的行人图片迁移，在保证行人本体前景不变的情况下，将背景转换成期望的数据集风格，还提出一个大型的行人重识别数据集 MSMT17，这个数据集包括多个时间段多个场景，包括室内和室外场景，是一个应用指向明确的数据集。Deng 等人[9]提出了一种风格迁移学习的框架以及一种生成对抗网络(similarity preserving GAN)，以无监督学习的方式将有标记图像从源域迁移到目标域，然后通过有监督学习的方法训练迁移图像的行人重识别模型。Zhong 等人[10]提出了异构同质学习(heterogeneous homogenous learning)的方法，将目标域和源域混合训练，提高目标测试集上的行人重识别模型的泛化能力。但基于无监督自适应的方法仅专注于从源域转移或适应判别性信息，而忽略了未标记目标域中的判别性标签信息的挖掘，甚至在适应之后，

源域中的判别性信息在目标域中的有效性也较低。因而本文提出了一种基于软多标签和深度特征融合的无监督行人重识别方法。

该方法主要创新点包括：

1) 提出软多标签解决无监督行人重识别目标数据集无标签问题；

2) 使用与软多标签对应的损失函数，提高软多标签准确度，解决跨摄像头标签一致以及纠正跨域分布错位问题；

3) 改进残差网络，将多层级深度特征进行融合，通过特征互补解决无监督行人重识别的识别效果较差的问题；

4) 结合压缩激励网络提高训练速度，解决深度学习行人重识别训练速度较慢的问题。

2 基于改进残差网络的软多标签无监督行人重识别算法

2.1 软多标签学习

软多标签学习是通过比较目标人和参考人，给无标签目标数据集加上软多标签。并且引入参考代理的概念，用参考代理代表每个参考人。为了提高软多标签准确度，参考 Yu[11]提出的损失函数(如图 1)，学习判别性的深度特征嵌入，即利用软多标签判别视觉上相似的目标对，减小不同视图间软多标签的差异，解决跨视域标签一致性以及纠正跨域分布错位等问题。

2.1.1 软多标签函数

通过将深度特征 $f(x)$ 与参考代理 $\{a_i\}_{i=1}^{N_p}$ 进行比较，得到如下软多标签函数 y：

$$y^{(k)}=l(f(x),\{a_i\}_{i=1}^{N_p})^{(k)}=\frac{\exp(a_k^{\mathrm{T}}f(x))}{\sum_i\exp(a_i^{\mathrm{T}}f(x))}\quad,\tag{1}$$

式中：$X=\{x_i\}_{i=1}^{N_u}$ 为无标签的目标数据集；$l(\cdot)$ 为软多标签要学习的映射函数；$f(\cdot)$ 为要学习的有区分力深度特征嵌入；$\{a_i\}_{i=1}^{N_p}$ 为引入的参考代理，其中每个代理代表一个共享联合特征嵌入的参考人。

在行人重识别中，两张图像是同一人为正对，不是同一人为负对，视觉相似度高但不是同一人为硬负对。为了拉近正对距离，推远负对距离，设计的损失函数为

$$L_{\mathrm{MDL}}=-\log\frac{\bar{P}}{\bar{P}+\bar{N}}\quad,\tag{2}$$

其中：

$$\bar{P}=\frac{1}{|p|}\sum\nolimits_{(i,j)\in p}\exp(-\|f(x_i)-f(x_j)\|_2^2)\quad,\tag{3}$$

$$\bar{N}=\frac{1}{|n|}\sum\nolimits_{(k,l)\in n}\exp(-\|f(x_k)-f(x_l)\|_2^2)\quad。\tag{4}$$

p 为所有正对的集合；n 为所有硬负对的集合。

$$p=\{(i,j)\,|\,f(x_i)^{\mathrm{T}}f(x_j)\geq S,A(y_i,y_j)\geq T\}\quad,\tag{5}$$

$$n=\{(k,l)\,|\,f(x_k)^{\mathrm{T}}f(x_l)\geq S,A(y_k,y_l)<T\}\quad。\tag{6}$$

S 为特征相似度阈值(用内积简化)；T 为软多标签相似度阈值；$A(\cdot,\cdot)$ 为基于 L_1 距离的软多标签一致性：

$$A(y_i,y_j)=y_i\wedge y_j=\sum\nolimits_k\min(y_i^{(k)},y_j^{(k)})$$
$$=1-\frac{\|y_i-y_j\|_1}{2}\quad。\tag{7}$$

通过最小化 L_{MDL} 来学习有区分力的嵌入，更精确判别目标对。

2.1.2 跨视图标签一致

为了让不同摄像头中相同的行人图像的软多标签

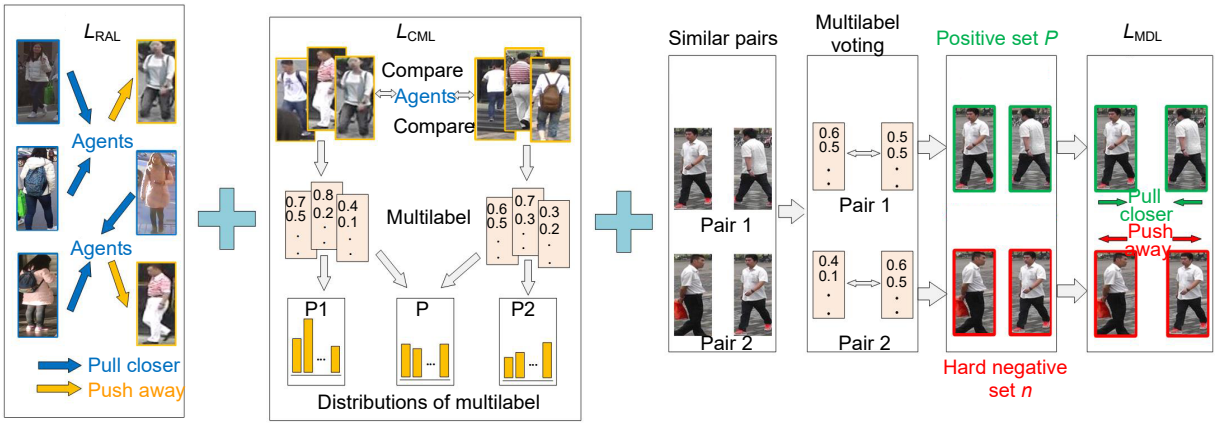


图 1 软多标签学习损失函数

Fig. 1 Soft multilabel learning loss function illustrate

是一致的，损失函数为

$$L_{\mathrm{CML}}=\sum_v d(P_v(y),P(y))^2,$$

和：

$$L_{\mathrm{CML}}=\sum_v\left\|\mu_v-\mu\right\|_2^2+\left\|\sigma_v-\sigma\right\|_2^2 \quad 。 \tag{8}$$

软多标签近似遵循对数正态分布，式中 $P(y)$ 是数据集 X 的软多标签分布，$P_v(y)$ 是数据集 X 中第 v 个摄像头视图中的软多标签分布，$d(\cdot,\cdot)$ 是两个分布之间的距离，采用简化的 2-Wasserstein 距离[12]。μ 和 σ 分别代表软多标签的均值和方差，μ_v 和 σ_v 是第 v 摄像头视图中软多标签的均值和方差。损失函数 L_{CML} 目的是减小同一人在不同摄像头下软多标签之间的差异。

2.1.3 参考代理学习

为了减小参考代理和参考人之间的差距，设计参考代理损失函数如下：

$$\begin{aligned}L_{\mathrm{AL}}&=\sum_k-\log l(f(z_k),\{a_i\})^{(w_k)}\\&=\sum_k-\log\frac{\exp(a_{w_k}^{\mathrm{T}}f(z_k))}{\sum_j\exp(a_j^{\mathrm{T}}f(z_k))},\end{aligned} \tag{9}$$

其中：有标签的辅助参考数据集为 $Z=\{z_i,w_i\}_{i=1}^{N_a}$，$w_i=1,...,N_p$ 是每个参考人 z_i 对应的标签；z_k 是辅助数据集中带有标签 w_k 的第 k 个人，w_{ak} 是标签为 w_k 的人的特征表达。$\{a_i\}$为引入的参考代理。通过最小化 L_{AL} 可以学习参考代理，也增强了软多标签函数 $l(\cdot)$的有效性。

为了进一步提高参考代理的有效性和解决跨域分布错位问题，针对每个代理 a_i 找出与之接近的无标签人 $f(x)$，损失函数为

$$L_{\mathrm{RJ}}=\sum_i\sum_{j\in M_i}\sum_{k:w_k}i\left[m-\left\|a_i-f(x_j)\right\|_2^2\right]_+\left\|a_i-f(z_k)\right\|_2^2, \tag{10}$$

其中：$M_i=\{j\left|\left\|a_i-f(x_j)\right\|_2^2<m\right.\}$ 表示与第 i 个代理 a_i 相关联的挖掘数据，m=1[13]，$[\cdot]_+$ 是铰链函数。

参考代理学习为 $L_{\mathrm{RAL}}=L_{\mathrm{AL}}+\beta L_{\mathrm{RJ}}$，其中 β 为平衡损失大小的参数。

整体损失函数为 $L_{\mathrm{MAR}}=L_{\mathrm{MDL}}+\lambda_1L_{\mathrm{CML}}+\lambda_2L_{\mathrm{RAL}}$，其中 λ_1 和 λ_2 分别是跨视图标签一致和参考代理学习的超参数。

2.2 改进的残差网络

2.2.1 SENet

SENet(Squeeze-and-excitation networks，压缩激励网络)模型[14]从特征通道间的关系层面提升网络性能。通过模型学习来得到每个特征通道的重要程度，再依据重要程度提升有用特征。

在 SENet 中，首先是压缩操作，将每个二维的特征通道变成一个实数。其次进行激励操作，通过类似于循环神经网络中门的机制为每个特征通道生成权重。最后是重标定权重(reweight)操作，将经过激励操作输出的权重看做是经过特征选择后的每个特征通道的重要程度，再针对每个通道对之前的特征进行加权计算[15]。

2.2.2 改进的残差网络

在图像处理领域，深度学习使用卷积神经网络的堆叠，能够提取图像更深层次的特征[16]。但是随着网络深度增加，会出现退化问题，也就是当网络变深时，训练准确率趋于平缓，但训练误差变大，为了解决这种退化现象，ResNet[17]被提出。

残差网络中的残差块在网络中使用了跳跃连接，不会引进额外的参数以及提高计算复杂度。同时上个残差块的信息直接流入下个残差块，提高了信息流通，也避免了由于网络过深引起的梯度消失和退化问题，使得网络在不断加深的同时也能良好地训练。

在残差网络中，有两个基本的残差块[18]，分别是卷积残差块和恒等残差块。ResNet50 残差网络的结构(如图 2)首先都通过一个 7×7 的卷积层，接着是一个 3×3 的最大池化，之后就是堆叠残差块。堆叠的残差块为四层，每层分别有 3、4、6、3 个残差块。每层只有一个卷积残差块，其余均为恒等残差块。在网络的最后连接全局平均池化、归一化层和全连接层。

在残差网络中，不同层对应不同语义层次的特征，将不同层特征进行融合，实现不同层特征之间的信息互补，达到提高特征鲁棒性的目的。本文中的深度特

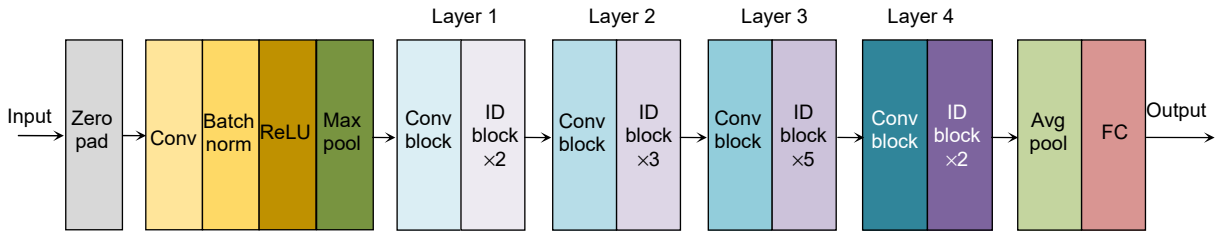


图 2 ResNet-50 结构图

Fig. 2 ResNet-50 illustrate

征融合，是在每层后加全局平均池化之后对其融合。

在残差网络基础上融合 SENet，是在每一个残差块中融合 SE 模块，在 SE 模块中先进行平均池化，再经过两个全连接将特征维度降低和提高，最后进行归一化并将归一化后的权重逐通道加权到特征上。融合后的 SE_ResNet 网络[19]结构如图 3 所示。

2.3 算法步骤

本文提出一个基于软多标签无监督和深度特征融合的行人重识别，其结构如图 4 所示，算法步骤如下：

1) 在整个网络中，首先由一个在 ImageNet 上预训练的 ResNet50 网络训练参考数据集，在预训练中只保留参考数据集相关部分，仅使用 L_{AL} 损失函数训练参考代理。

2) 再把参考图像和目标图像共同输入网络模型中进行训练，batchsize 设置为 368，一半是随机的无标签图像 X，另一半是随机的参考样本 Z。选择 ResNet50 作为基础网络，去掉基础网络原有的最后一层，连接一个 2048 维的全连接层，在每个残差块中加入 SE 模块，提升有用特征抑制其他特征。

3) 将每层后加一层平均池化，再进行深度特征融合并赋值给 feature_map，由于融合后尺寸会发生变化，故需将最后一层维度改为融合后的维度。再将 feature_map 展平成一维再规则化得到 feature，然后将权重取出进行规则化再相乘，得到相似度 sim，返回值有三个分别是 feature_map、feature、sim。

4) 最后通过计算损失函数，挖掘负样本信息，以及实现跨域标签一致，最后通过设置阈值，对高于阈值的正样本进行排序。

3 实验结果与分析

3.1 数据集与评价指标

实验采用目前行人重识别领域普遍采用的

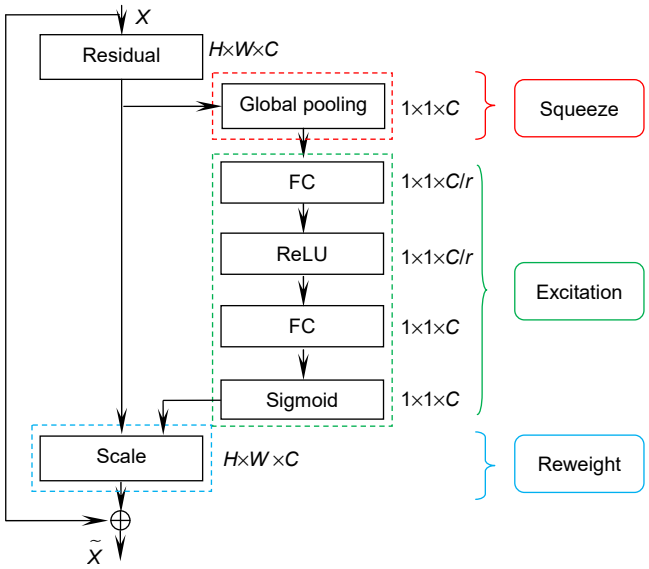


图 3 SE_ResNet 网络结构图

Fig. 3 SE_ResNet illustrate

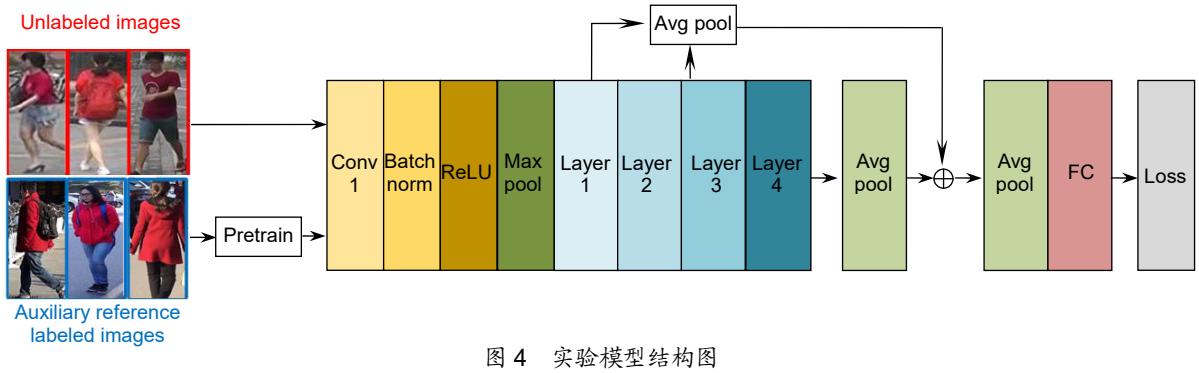


图 4 实验模型结构图

Fig. 4 Experimental model illustrate

Market-1501[20]与 DukeMTMC-reID[21]数据集作为目标数据集，MSMT17[8]作为参考数据集，Market-1501、DukeMTMC-reID 和 MSMT17 数据集涵盖的现实采集情况最为广泛，行人图像多，接近实际情况。其中 MSMT17 的训练图像有 32621 张，Market-1501 和 DukeMTMC-reID 的训练图像分别为 12936 和 16522 张。在实验中需要参考数据集大约是训练人数的两倍，性能趋于稳定，本文选用 Market-1501 与 DukeMTMC-reID 数据集作为目标数据集，MSMT17 作为参考数据集。为了得到更准确的数据，实验中参考数据集与目标数据集完全不重叠，训练集的行人身份和测试集的行人身份也是没有重复的。本文使用 Rank1、Rank5、Rank10 和平均精度均值(mAP)来评估性能。

3.2 实验配置

实验基于 PyTorch1.0 框架，使用 Tesla V100GPU 的 NVIDIA DGX Station 服务器的 Linux 系统，用四块 GPU 进行训练和测试，预训练和训练时间均约为 8 h。使用改进的残差网络作为骨干网络来进行实验，使用 ImageNet 作为实验中 pre-train 的预训练来训练参考数据集。

3.3 实验结果

3.3.1 改进残差实验

首先实验验证预训练参考数据集对训练精度有很大的影响，在 ResNet50 的基础网络上对预训练进行消融实验。使用不同的网络进行预训练和训练，实验结果如表 1 所示。

本实验中预训练中参数设置训练次数 epoch 为 60、一次训练的样本数为 256、学习率为 0.001、权重衰减率为 0.025，预训练结果最优；在训练中参数设置训练次数 epoch 为 20、一次训练的样本数为 368、学习率为 0.0002、权重衰减率为 0.025，训练结果最优。

由实验结果可知，在相同网络下，没有预先训练参考数据集的相比有预先训练的最终训练精度低 17%，有预先训练参考数据集，但没有使用 imagenet 预训练的相比使用了的，精度低 6%。

改变训练的骨干网络，选择 SE_ResNet50 进行调参，不同的 Epoch、学习率(learning rate)以及权重衰减(weight decay)实验所得结果如图 6。

实验结果表明，SE_ResNet50 的精度比 ResNet50

图 5 数据集图片。

(a) Market‐1501 数据集行人图片；(b) DukeMTMC‐reID 数据集行人图片；(c) 参考数据集 MSMT17 行人图片

Fig. 5 Dataset picture. (a) Person pictures in the Market-1501 dataset; (b) Person pictures in the DukeMTMC-reID dataset; (c) Person pictures in the reference dataset MSMT17

表 1 预训练对训练结果的影响

Table 1 The effect of pre-trained on training results

Train	Pre-train	Market			
		R1	R5	R10	mAP
ResNet50	ResNet50(loss_total=0.622)	66.627	81.977	86.609	39.361
ResNet50	ResNet50, imageNet=None	61.401	77.316	82.245	33.930
ResNet50	None	42.548	60.778	69.032	22.139
SE_ResNet50	SE_ResNet50	51.989	69.893	77.078	28.330
SE_ResNet50	ResNet50(loss_total=0.622)	64.371	81.621	86.461	38.236
SE_ResNet50	None	35.362	51.425	58.462	17.403

的精度低一个百分点，但是训练速度有明显提高，在Epoch 均为 20 次的情况下，SE_ResNet50 的训练时间比 ResNet50 低一个小时左右。其主要原因是 SENet 对通道层面的特征进行重标定，提升有用特征，抑制其他特征。但在软多标签中，抑制特征过程存在误差，导致损失精度，但加入 SE 模块后提高了学习效率，使训练时间减少。在预训练过程中加入 SE 模块会影响参考代理的学习，因为预训练是训练参考数据集为参考代理的过程，训练过程中需要学习尽量多的特征。

3.3.2 深度特征融合

根据上一小节实验结果可知，在网络中加入 SE 模块，提高训练速度的同时损失部分特征，本节深度特征融合实验的目的是融合更多特征，减小信息传递过程的损失，因此本实验中只对 ResNet50 进行深度特征融合，实现特征信息互补，以提高识别精度。

实验结果表明(表 2)，将残差网络中第三层和第四层的深度特征进行融合，融合后具有更好的特征表达，提高了识别精度，与直接输出第四层深度特征相比识别精度提高约 1%。在第三、四层深度特征融合基础上再融入第一层的浅层特征，达到深度特征与浅层特征互补的目的，最终识别精度共提高 2%左右。

3.3.3 消融实验

在基本的 ResNet50 网络中，按去掉损失函数 L_{CML}、去掉损失函数 L_{CAL} 和二者均去掉 3 种方案进行实验，结果如表 3 所示。

3.3.4 实验分析

将实验结果与最先进的无监督行人重识别模型进行比较，其中包括基于伪标签学习的：用于无监督人员重新识别的跨视图非对称度量学习[5]、无监督行人重识别：聚类和微调[6]；基于无监督域自适应的：基于可转移的联合属性-身份深度学习的无监督人的重新识别[7]、基于人迁移生成对抗网络缩小域差距的行人重识别[8]、基于自相似性和域相似性的图像-图像域自适应的行人重识别[9]和异质同构概括行人的检索模型[10]；以及基于多任务中级特征对齐网络的无监督的跨数据集行人重识别[22]，深度非对称度量嵌入的无监督行人重识别[23]，适应和重新识别网络：用于行人重识别的无监督深度迁移学习[24]，基于无监督行人重识别的自底向上聚类方法[25]和基于软多标签学习的无监督行人重识别[11]。结果见表 4。

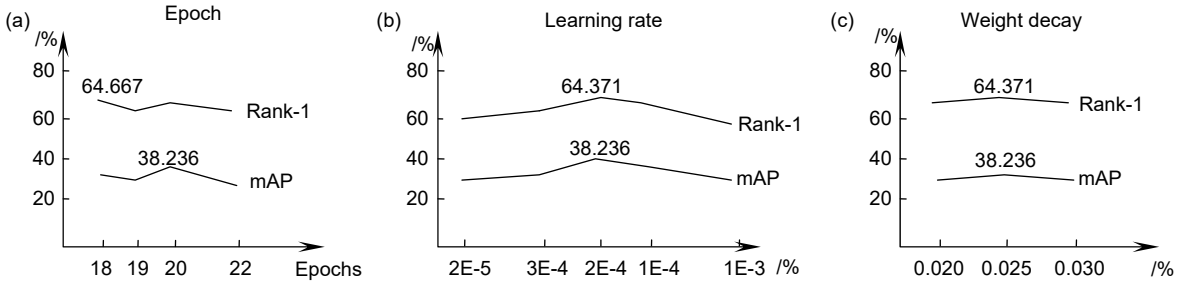


图 6 对 SE_ResNet-50 进行调参实验结果。(a) 调纪元数 epoch 参数实验结果曲线图；(b) 调学习率 learning rate 实验结果曲线图；(c) 调权重衰减 weight decay 实验结果曲线图

Fig. 6 Results of adjusting hyperparameters for SE_ResNet-50. (a) Adjusting the experimental results of the epoch parameters; (b) Adjusting the learning rate experimental results; (c) Adjusting the weight decay experimental results

表 2 特征融合实验结果

Table 2 Feature fusion experimental results

Methods	Market			
	R1	R5	R10	mAP
ResNet50+layer1+layer3	52.049	69.567	76.395	28.279
ResNet50+layer1+layer4	66.330	81.473	86.520	39.581
ResNet50+layer2+layer3	51.306	69.240	76.306	28.113
ResNet50+layer2+layer4	62.500	77.316	83.254	35.884
ResNet50+layer3+layer4	67.102	81.977	86.876	40.036
ResNet50+layer1+layer3+layer4	68.973	82.601	86.995	41.188

表 3 消融实验

Table 3 Ablation study

Methods	Market-1501				DukeMTMC-reID			
	R1	R5	R10	mAP	R1	R5	R10	mAP
w/o L_{CML}	60.0	75.9	81.9	34.6	63.2	77.2	82.5	44.9
w/o L_{RAL}	59.2	76.4	82.3	30.8	57.9	72.6	77.8	37.1
w/o L_{CML} & L_{RAL}	53.9	71.5	77.7	28.2	60.1	73.0	78.4	40.4
ResNet50+L_{MAR}	66.627	81.977	86.609	39.361	67.1	79.8	84.2	48.0

表 4 与相关方法无监督行人重识别精度对比

Table 4 Comparison of unsupervised person recognition accuracy with related methods

Methods	Reference	Market			Duke		
		R1	R5	mAP	R1	R5	mAP
CAMEL[5]	ICCV'17	54.5	73.1	26.3	40.3	57.6	19.8
PUL[6]	ToMM'18	45.5	60.7	20.5	30.0	43.4	16.4
TJ-AIDL[7]	CVPR'18	58.2	74.8	26.5	44.3	59.6	23.0
PTGAN[8]	CVPR'18	38.6	57.3	15.7	27.4	43.6	13.5
SPGAN[9]	CVPR'18	51.5	70.1	27.1	41.1	56.6	22.3
HHL[10]	ECCV'18	62.2	78.8	31.4	46.9	61.0	27.2
MMFA[22]	BMVC'18	45.3	-	24.7	56.7	-	27.4
DECAMEL[23]	SCI'18	60.24	-	32.44	-	-	-
ARN[24]	CVPRW'19	70.3	80.4	39.4	60.2	73.9	33.4
BUC[25]	AAAI'19	66.2	79.6	38.3	47.4	62.6	27.5
MAR[11]	CVPR'19	67.7	81.9	40.0	67.1	79.8	48.0
The proposed method	This work	68.97	82.6	41.2	68.6	80.6	50.1

从对比结果可以看出，本文方法明显优于相关无监督行人重识别方法。与基于伪标签学习的无监督行人重识别模型相比，本文的软多标签参考学习可以利用辅助参考信息挖掘潜在的区分性信息，当直接比较一对视觉相似的人的视觉特征时，这些信息很难被检测到；与基于无监督域自适应的行人重识别模型相比，本文的模型在未标记的目标数据中挖掘区分性信息，这对行人重识别任务具有直接的有效性；与最先进的无监督学习方法相比，精度提高 1%到 23%；与基于多软标签的方法相比，精度提高 1%到 2%。实验证明该方法是有效的。

4 结 论

本文提出了基于软多标签无监督和深度特征融合的行人重识别模型，相比于其他的无监督行人重识别方法，本文的无监督使用了软多标签，并结合压缩激励网络以及深度特征融合，在两个行人重识别普遍采用的数据集上都取得了较好的效果。从研究结果来看，软多标签的设计在无监督行人重识别中具有很大的作用，其精度远远超过其他无监督行人重识别方法，改进残差结构能够在软多标签无监督基础上提升训练速度，取得较好的结果。但精度低于有监督学习的方法，因为图像无标签需要计算机自学习，后续可以继续提高软多标签的准确率，使无监督同有监督一样识别。加入 SE 模块是对通道层面特征进行重排序，说明通道层面特征对行人重识别影响不明显，后续可以研究其他的注意力机制方法对实验结果的影响。

参考文献

[1] Xiong F, Xiao Y, Cao Z G, *et al*. Good practices on building effective CNN baseline model for person re-identification[J]. *Proceedings of SPIE*, 2019, **11069**: 110690I.

[2] Wang S Q, Xu X, Liu L, *et al*. Multi-level feature fusion model-based real-time person re-identification for forensics[J].

Journal of Real-Time Image Processing, 2020, **17**(1): 73–81.

[3] Bak S, Carr P, Lalonde J F. Domain adaptation through synthesis for unsupervised person re-identification[J]. *ECCV*, 2018: 189–205.

[4] Ye M, Li J W, Ma A J, *et al*. Dynamic graph co-matching for unsupervised video-based person re-identification[J]. *IEEE Transactions on Image Processing*, 2019, **28**(6): 2976–2990.

[5] Yu H X, Wu A C, Zheng W S. Cross-view asymmetric metric learning for unsupervised person re-identification[C]// *Proceedings of 2017 IEEE International Conference on Computer Vision*, Venice, 2017: 994–1002.

[6] Fan H H, Zheng L, Yan C G, *et al*. Unsupervised person re-identification: clustering and fine-tuning[J]. *ACM Transactions on Multimedia Computing*, *Communications*, *and Applications*, 2018, **14**(4): 83.

[7] Wang J Y, Zhu X T, Gong S G, *et al*. Transferable joint attribute-identity deep learning for unsupervised person re-identification[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 2018: 2275–2284.

[8] Wei L G, Zhang S l, Gao W, *et al*. Person transfer GAN to bridge domain gap for person re-identification[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 2018: 79–88.

[9] Deng W J, Zheng L, Ye Q X, *et al*. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 2018: 994–1003.

[10] Zhong Z, Zheng L, Li S Z, *et al*. Generalizing a person retrieval model hetero-and homogeneously[C]//*Proceedings of the European Conference on Computer Vision*, Glasgow, 2018: 172–188.

[11] Yu H X, Zheng W S, Wu A C, *et al*. Unsupervised person re-identification by soft multilabel learning[C]//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Long Beach, CA, 2019: 2148–2157.

[12] He R, Wu X, Sun Z N, *et al*. Wasserstein CNN: learning invariant features for NIR-VIS face recognition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, **41**(7): 1761–1773.

[13] Wang F, Xiang X, Cheng J, *et al*. NormFace: L_2 hypersphere embedding for face verification[C]//*Proceedings of the 25th ACM International Conference on Multimedia*, California, Mountain View, 2017: 1041–1049.

[14] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 2018: 7132–7141.

[15] Wang C, Zhang Q, Huang C, *et al*. Mancs: a multi-task attentional network with curriculum sampling for person re-identification[C]//*Proceedings of the 15th European Conference on Computer Vision*, Munich, 2018: 365–381.

[16] Fan H, Zheng L, Yan C, *et al*. Unsupervised Person Re-identification by Deep Learning Tracklet Association[J]. *Acm Transactions on Multimedia Computing Communications & Applications*, 2018, **14**(4): 1–18.

[17] He K M, Zhang X Y, Ren S Q, *et al*. Deep residual learning for image recognition[C]//*Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 2016: 770–778.

[18] Wang Y, Wang L Q, You Y R, *et al*. Resource aware person re-identification across multiple resolutions[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 2018: 8042–8051.

[19] Hu Y, Wen G H, Luo M N, *et al*. Competitive inner-imaging squeeze and excitation for residual network[Z]. arXiv: 1807.08920[cs: CV], 2018.

[20] Zheng L, Shen L Y, Tian L, *et al*. Scalable person re-identification: a benchmark[C]//*Proceedings of 2015 IEEE International Conference on Computer Vision*, Santiago, 2015: 1116–1124.

[21] Zheng Z D, Zheng L, Yang Y. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro[C]//*Proceedings of 2017 IEEE International Conference on Computer Vision*, Venice, 2017: 3754–3762.

[22] Lin S, Li H L, Li C T, *et al*. Multi-task mid-level feature alignment network for unsupervised cross-dataset person re-identification[Z]. arXiv: 1807.01440[cs: CV], 2018.

[23] 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*, 2018, **42**(4): 956–973.

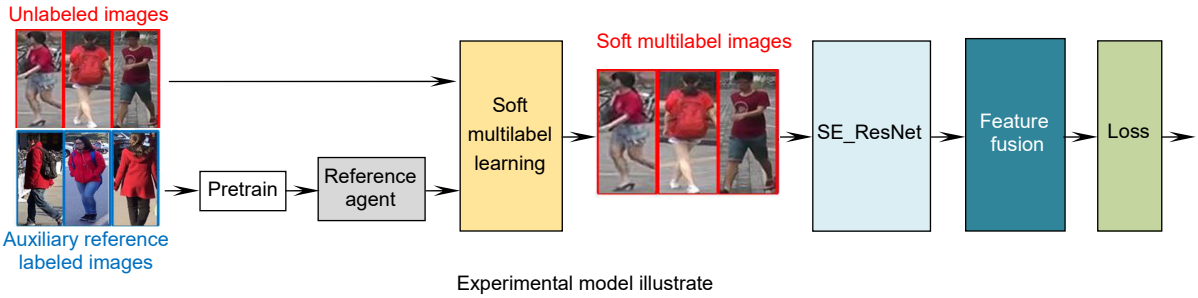
[24] Li Y J, Yang F E, Liu Y C, *et al*. Adaptation and re-identification network: an unsupervised deep transfer learning approach to person re-identification[C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Salt Lake City, 2018: 172–178.

[25] Lin Y T, Dong X Y, Zheng L, *et al*. A bottom-up clustering approach to unsupervised person re-identification[C]// *Proceedings of the 33rd AAAI Conference on Artificial Intelligence*, 2019: 8738–8745.

Soft multilabel learning and deep feature fusion for unsupervised person re-identification

Zhang Baohua[1,3*], Zhu Siyu[1], Lv Xiaoqi[2,3], Gu Yu[1,3], Wang Yueming[1,3],
Liu Xin[1,3], Ren Yan[1], Li Jianjun[1,3], Zhang Ming[1,3]

[1]School of Information Engineering, Inner Mongolia University of Science and Technology, Baotou, Inner Mongolia 014010, China;
[2]School of Information Engineering, Mongolia Industrial University, Huhehaote, Inner Mongolia 010051, China;
[3]Inner Mongolia Key Laboratory of Pattern Recognition and Intelligent Image Processing, Baotou, Inner Mongolia 014010, China



Experimental model illustrate

Overview: People re-identification is mainly used to retrieve pedestrians of interest in the images taken by the camera, and then retrieve targets similar to the people's image. This technology can save a lot of time and manpower in finding the images of the suspect in the pedestrian database, and has good application prospects in intelligent security, criminal investigation, and image retrieval. The supervised person re-identification model has better recognition accuracy, but there are scalability problems. For example, the accuracy of algorithm identification relies heavily on effective supervised information. When adding a small amount of data in the classification process, all data needs to be reprocessed, resulting in poor real-time performance. Aiming at the above problems, an unsupervised person re-identification algorithm based on soft multilabel is proposed. By learning the feature of the target, and then comparing it with the labeled reference datasets, each unlabeled target gets a soft multilabel. In this learning process, in order to obtain more accurate soft multilabel, we introduce the concept of reference agents and in order to reduce the difference between reference agents and labeled reference datasets, we pre-trained the reference datasets. Using a reference agent instead of a labeled reference dataset to compare with an unlabeled target. We also use three loss functions, which are used to mine hard negative pair information, make the cross-camera labels of the same target consistent, and correct cross-domain distribution misalignment. In these three loss functions, the purpose of mining hard negative pair information is to determine negative pairs more accurately and push the distance of negative pairs farther away; The cross-camera label consistency is to reduce the gap between multilabel for the same target under different camera distributions. Using the simplified 2-Wasserstein distance, the mean and standard deviation vectors of soft multilabel in different camera views are calculated; In order to further improve the effectiveness of the reference agent and solve the problem of cross-domain distribution misalignment, for each reference agent, find unlabeled people close to it and design a loss function. In the process of feature extraction, we use multi-level deep feature fusion to complement deep features with shallow features to achieve the purpose of improving feature robustness and thereby improving the recognition accuracy. We also tried to integrate squeeze-and-excitation networks (SENet) into the residual network to achieve a function similar to the attention mechanism to improve the learning speed. Experimental results show that rank-1 and mAP in this paper are superior to advanced correlation algorithms.

Citation: Zhang B H, Zhu S Y, Lv X Q, *et al*. Soft multilabel learning and deep feature fusion for unsupervised person re-identification[J]. *Opto-Electronic Engineering*, 2020, **47**(12): 190636

Supported by National Natural Science Foundation of China (61962046, 61663036, 61841204), Inner Mongolia Jieqing Cultivation Project (2018JQ02), Inner Mongolia Grassland Talents, Inner Mongolia Youth Science and Technology Innovation Talent Project (Level 1), Inner Mongolia Autonomous Region Natural Science Fund (2015MS0604, 2018MS06018), Inner Mongolia Autonomous Region Higher Education Science Funded by the Technical Research Project (NJZY145)

* E-mail: zbh_wj2004@imust.edu.cn