文章编号：2095-2163(2021)08-0146-05 中图分类号：TP391.4 文献标志码：A

无监督行人重识别的判别性特征研究

唐佳敏，韩 华，黄 丽，王春媛
（上海工程技术大学 电子电气工程学院，上海 201620）

摘 要：目前，在计算机视觉方面，大多的监督学习方法用于解决其重要分支；行人重识别问题已经取得了不错的成果，但是此类方法需要对训练数据进行手工标注，特别是对于大容量的数据集，手工标注的成本很高，而且完全满足成对标记的数据难以获得，所以无监督学习成为必选项。此外，全局特征注重行人特征空间整体性的判别性，而局部特征有助于凸显不同部位特征的判别性。所以，基于全局与局部特征的无监督学习框架，使用全局损失函数与局部相斥损失函数共同进行判别性特征学习，并联合优化 ResNet-50 卷积神经网络（CNN）和各个样本之间的关系，最终实现行人重识别。大量实验数据验证了提出的方法在解决行人重识别任务时具有优越性。

关键词：计算机视觉；行人重识别；无监督学习；判别性特征学习

Discriminative feature learning for unsupervised person re-identification

TANG Jiamin, HAN Hua, HUANG Li, WANG Chunyuan
(School of Electronic and Electrical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China)

【Abstract】 At present, most supervised learning methods in computer vision are used to address its important branches: Good results have been achieved in pedestrian recognition, but this kind of method requires manual labeling of training data, especially for large capacity data sets, manual labeling cost is very high, and it is difficult to obtain data that fully meet paired labeling, so unsupervised learning has become a mandatory option. In addition, global features pay attention to the discriminant of the spatial integrity of pedestrian features, while local features help to highlight the discriminant of features in different parts. Therefore, based on the unsupervised learning framework of global and local features, the global loss function and local repulsive loss function are used to jointly learn discriminant features, and the relationship between resnet-50 convolutional neural network (CNN) and each sample is jointly optimized to realize the pedestrian recognition. A large number of experimental data verify the advantages of the proposed method in solving the pedestrian recognition task.

【Key words】 computer vision; person re-identification; unsupervised learning; discriminative feature learning

0 引 言

近年来对“安全防范”与“治安管理”的重视，视频监控因为其实时性与精准性，需求性逐渐增强。但在传统的视频监控中，只有简单的记录、存储、回放等功能，无法起到有效的安全防范和治安管理的作用，且海量的视频数据依靠人工检索非常耗时耗力，还不能保证准确性，由此智能视频应运而生，且发展迅猛，其中的行人重识别问题也发展成为热点话题。行人重识别（Person Re-ID）研究主要从行人跨摄像头跟踪问题开始，是用来判断目标行人在无重叠视域中被拍摄到的图像是否属于同一身份的目标行人。行人重识别研究可以广泛用于智能视频监控，安全防御等领域。由于行人外观易受衣物、姿态和摄像头视角变化以及光照角度、事物遮挡、环境等各种复杂因素的影响，使得行人重识别研究面临了很多挑战与困难。近年来，行人重识别技术引起了各界的广泛关注，提出了很多优秀的技术研究方法[1]。

1 行人重识别

行人重识别问题目前已成为计算机视觉等研究领域的热点，其主要功能就是在不同摄像头下找到目标行人的身份关联信息，以便能准确的识别目标行人，如图 1 所示。

行人重识别早期并没有获得过多关注，只是作为跨摄像机目标跟踪的一个分支。2005 年，行人重识别（Person Re-identification）一词在研究跨摄像机

作者简介：唐佳敏（1995-），女，硕士研究生，主要研究方向：目标识别与跟踪、行人重识别、深度学习和图像处理；韩 华（1983-），女，博士，副教授，主要研究方向：目标识别与跟踪、行人重识别、模式识别等；黄 丽（1993-），女，博士，讲师，主要研究方向：多机器人系统、生物智能算法、动态协同等；王春媛（1983-），女，博士，讲师，主要研究方向：多源信息协同处理、模式识别、机器学习等。
通讯作者：韩 华 Email:2070967@mail.dhu.edu.cn

收稿日期：2021-01-17

目标跟踪问题中第一次被提出；在 2006 年，Gheissari 等人在国际顶级会议上首次将 Person Re-Identification 这一术语提出，将行人重识别当作一个独立的研究方向来开展[2]；特别是在 2007 年，D. Gary 等人公开发布了第一个关于行人重识别的数据集：VIPeR。这一数据集的发布使得越来越多的国内外学者对此感兴趣，纷纷投入研究，使之成为计算机视觉领域的热点研究问题。

图 1　无重叠区域监控网络中的行人重识别

Fig. 1　Person re-identification in non-overlapping area monitoring network

行人重识别的研究，从国内外研究的发展历史来看主要有两大阶段：基于人工设计特征的行人重识别方法和基于深度学习的行人重识别方法。基于人工设计特征的行人重识别方法主要由两部分组成：特征提取和相似性度量。特征提取主要提取鲁棒性强且具有很强区分判别性的特征表示向量；相似性度量主要对目标行人间的特征向量间的相似度进行比对。基于深度学习的行人重识别方法则是将这两部分整合为一个整体，辅以损失函数约束。

行人重识别研究出现了很多优秀的有监督学习算法，虽然有监督学习的发展已经取得了很好的结果，但是其获得标签信息的工作量和难度都很大；而无监督学习由于不需要给数据打标签，通过发现一些潜在的结构来训练数据，可以节省很多人力物力资源，因而受到越来越多的关注。本文的研究也是基于无监督学习的，以提取联合判别性特征为目标。

2　方法

在本项工作中，使用 ResNet-50 作为卷积网络的骨干网络，研究了基于深度学习系统的无监督行人重识别，提出了一种联合判别性特征的无监督框架，如图 2 所示。对于行人图片，使用基于补丁的判别特征学习损失，将类似补丁块的特征拉到一起，并推出不相似的补丁块，来指导未标记数据集学习具有判别性的局部补丁特征。从全局方面，提出使用相斥损失的聚类策略来对样本进行判别性的全局特征学习。

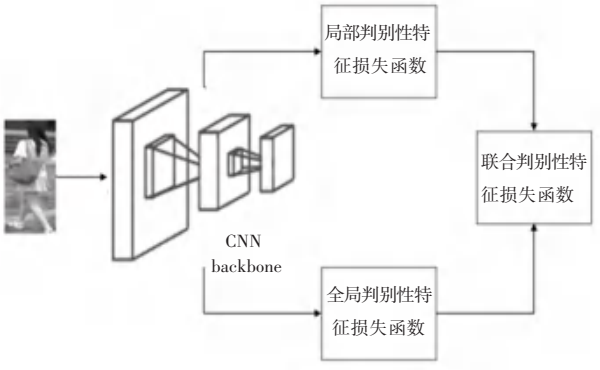

图 2　本方法框架图

Fig. 2　Framework diagram of this method

2.1　局部判别性特征损失函数

局部特征学习旨在指导补丁网络在未标记的数据集上学习判别性补丁特征。从相对较小尺寸的特征图中提取补丁，而不是从图像中采样，这样可以有效地减少特征计算中的计算量和 CNN 网络的复杂度[3]。为此，本文引入了一个空间变换网络来形成补丁网络，可以实现自动地从特征图中提取补丁的功能[4]。补丁网络为每个图像特征映射，生成 M 个补丁块，并且同一图像的这些不同补丁块位于不同的空间区域，这些不同的区域可能包含不同的身体部位，具有不同的语义信息，所以使用不同的 CNN 分支对同一图像的这些不同的补丁进行编码，并对不同的分支独立地进行判别性特征学习，如图 3 所示。

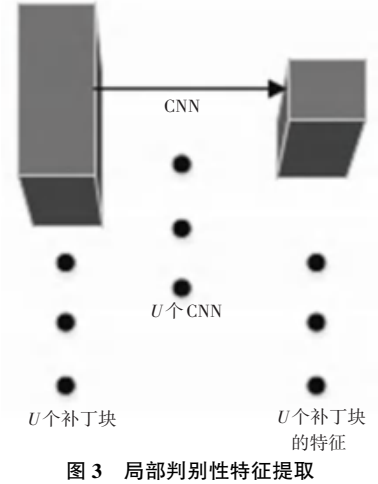

图 3　局部判别性特征提取

Fig. 3　Local discriminative feature extraction

在一般的特征学习中，总是希望同一类的特征在特征空间中更接近，同时远离其它类，这样学习到的特征更具有判别性，所以这里的补丁网络特征学习是将特征空间中相似的补丁块拉近，同时将不相

似的补丁块推远。

本文在本项无监督框架中使用一种基于补丁块的判别特征损失函数,将相似的特征拉到一起,并推出不相似补丁块,来学习未标记数据集中的补丁特征,公式(1)如下:

$$L_s^u = -\log \frac{\sum_{w_j^u \in k_i^u} e^{-\frac{s}{2}\| x_i^u - w_j^u \|_2^2}}{\sum_{j=1, j\neq i}^{N} e^{-\frac{s}{2}\| x_i^u - w_j^u \|_2^2}}. \quad (1)$$

其中, $W^u = \{W_j^u\}_{j=1}^N$ 是用来存储补丁块特征的存储体; N 是训练图像的数量; x_i^u 表示一个体量中的第 i 个图像的第 u 个补丁的特征; k_i^u 是 x_i^u 的 k 个最近补丁的合集,是通过对每个 x_i^u 计算 W^u 的成对距离得到的; s 是缩放参数。

因为将相似的人的图像特征直接拉近,也许会把具有不同身份的,但视觉上相似的人的图像特征拉近,这是由于忽略人的身份信息,导致的识别率的降低。所以通过将人的图像划分为部分,可以让同一图像的不同补丁块包含该人的不同信息,从而挖掘出埋藏其中的潜在信息。

2.2 全局判别性特征损失函数

全局特征学习旨在通过聚类策略,利用特征的相似性,将具有相同身份的图片结合在一起,以此生成聚类,然后利用卷积模型进行最大化聚类中心差异性的操作进行数据的更新。

已知无监督数据集里的每张图片都没有身份标注,因此在开始的时候会将每张图片分配为各自的聚类中心,即 $\{\hat{y}_i = i \mid 1 \leqslant i \leqslant N\}$ ($\hat{y}_i$ 是 x_i 的聚类数量的动态索引)。这种方式能让网络学习识别每一个聚类的训练样本,而不是每一个人,并且可以将每个训练样本间的多样性达到最大化。随着数据参数更新,将类似的行人图片并到同一个身份的聚类中,来表明行人图片身份的同一性。

令一张图片 x 属于第 c 个聚类中心的概率如式(2)所示:

$$p(c \mid x_i, V) = \frac{\exp(V_c^T v_i / \tau)}{\sum_{j=1}^{C}, \exp(V_j^T v_i / \tau)}. \quad (2)$$

其中, C 是当前状态下聚类的数目,在开始状态时 $C = N$,也就是聚类的数目等于图片的数目。随着相似的图像逐渐合并,聚类 C 的数量也逐渐减少; $v = \frac{\varphi(\theta; x_i)}{\| \varphi(\theta; x_i) \|}$ 指代的是数据 x_i 特征空间中的 l_2 范数,即 $\| v_i \| = 1$; $V \in R^{C \times n_\varphi}$ 是一个查询列表,其中存放着每一个聚类的特征; V_j 表示 V 的第 j 列特征; τ 是一个标量参数,引入的目的是为了便于对概率的取值区间有一个控制因素。在后续的实验中,将 τ 设置为 0.1。

在之前的操作中,通过 $V^T \cdot v_i$ 来计算数据 x_i 和其它数据间的余弦相似度,而现在通过 $V_{\hat{y}_i} \leftarrow 1/2(V_{\hat{y}_i} + v_i)$ 来计算表 V 的第 $\hat{y}_i$ 列数据,将原来聚类的特征与新的数据特征求和并求平均值;利用公式(3)的损失函数优化算法的卷积模型,将其作为相斥损失函数,可以让不同身份图片间的差异性扩大。

$$L_r = -\log(p(c \mid x, V)). \quad (3)$$

通过最小化公式(3)的损失函数,可以计算每个图像特征 v_i 与每一个聚类中心特征 $V_{j \neq \hat{y}_i}$ 之间的余弦距离,并将其最大化。还可以计算每个图像特征 v_i 与相对应的聚类中心特征 $V_{j=\hat{y}_i}$ 之间的余弦距离,并将其最小化,这样就可以利用多样性来推远不相似的图片。在优化的步骤中, V_j 列举了第 j 个聚类中心中所包括的全部图片的特征,将其作为该聚类的"中心点"。在模型训练的每一个阶段,对聚类中心的计算操作的时间复杂度非常高,所以可以通过查询表格 V 的方法来节省很多冗余的计算过程,这样带来的好处是在每次训练阶段不需要从所有训练数据中反复地进行提取特征的步骤。

2.3 联合损失函数

基于以上的无标签数据集框架下的局部判别性损失函数和全局判别性损失函数,最终每张图像形成的总的损失函数可以表示为式(4):

$$L = \lambda \frac{1}{U} \sum_{u=1}^{U} L_s^u + L_r. \quad (4)$$

其中, U 表示一张图片的补丁块的个数, λ 是一个控制权重的参数。

3 实验结果与分析

3.1 实验数据集

本次实验的数据集描述见表 1,实验在 Market-1501 数据集和 DukeMTMC-reID 数据集上操作研究。Market1501 数据集包含共 32 668 张行人图片,由分布的 6 个摄像头捕捉的 1 501 个不同行人身份,将总共的 32 668 张图片分为训练集和测试集两部分,其中训练集上有 12 936 张行人图片,测试集上有 19 732 张行人图片。DukeMTMC-reID 数据集共计 36 411 张图像。由 8 个摄像头捕捉 1 404 个行人身份,同样分为训练集的 16 522 张图像和测试集的 17 661 张图像。

表 1　数据集描述

Tab. 1　Description of the datasets

Dataset	Cams	Identities	Images	Train Images	Test images
Market-1501	6	1 501	32 668	12 936	19 732
DukeMTMC-reID	8	1 404	36 411	16 522	19 889

3.2　评测标准

本次实验中，使用两个性能指标来评判此研究方法：

(1) 累积匹配特征(CMC)曲线；

(2) 平均精度均值(*mAP*)。每个被查询图像的平均精度(*AP*)由图像的召回曲线确定，并通过计算查询图像的平均精度的平均值获得平均精度均值(*mAP*)。在积累匹配特性曲线(CMC)中选取 *Rank* - 1，*Rank* - 5 和 *Rank* - 10 的得分来反映检索的精度。

3.3　实验结果

将本算法性能与目前较先进的方法进行了比较，在 Market-1501 数据集上得到的积累匹配特性曲线(CMC)如图 4 所示，在 DukeMTMC-reID 数据集上得到的积累匹配特性曲线(CMC)如图 5 所示。同时，将本文方法与目前较先进方法的 *mAP* 值比较，见表 2，在 Market-1501 数据集上达到了 36.02，和已有的好方法相比提高 8.62 个百分点；在 DukeMTMC-reID 数据集上达到了 40.64，与已有的好方法比提高 15.94 个百分点。在 CMC 曲线中选取了具有代表性的 *Rank* - 1，*Rank* - 5 和 *Rank* - 10 的得分来进行比较，见表3，表4。从表3可以看出，本文的算法在 Market - 1501 数据集上的 *rank* - 1 最终结果达到了59.35，相较于已有的好方法提高了 2.65 个百分点；从表 4 我们可以看出，本文的算法在 DukeMTMC-reID 数据集上的 *Rank* - 1 最终结果达到了 55.75，比已有的好方法提高了 10.45 个百分点。因此，可以看出本文方法可以很好地解决行人重识别的问题，并且由于从局部和全局两个分支全面地解决此问题，使得本文方法具有一定的先进性。

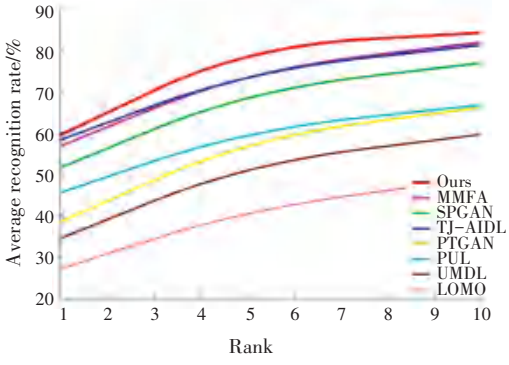

图 4　各算法在 Market-1501 数据集上的积累匹配特性曲线(CMC)

Fig. 4　Cumulative matching characteristic curve (CMC) of each algorithm on the Market-1501 dataset

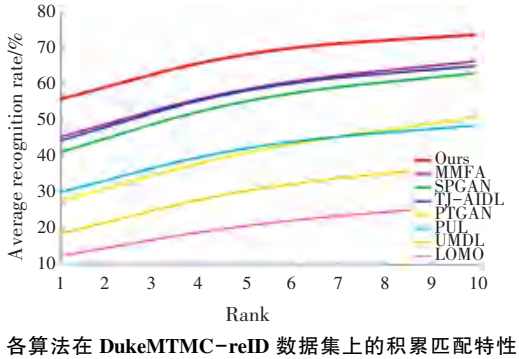

图 5　各算法在 DukeMTMC-reID 数据集上的积累匹配特性曲线(CMC)

Fig. 5　Cumulative matching characteristic curve (CMC) of each algorithm on the DukeMTMC-reID dataset

表 2　各算法在 Market-1501 和 DukeMTMC-reID 数据集上的平均精度均值(mAP)

Tab. 2　The average precision (mAP) of each algorithm on the Market-1501 and DukeMTMC-reID datasets

经典先进算法	LOMO	UMDL	PUL	TJ-AIDL	SPGAN	MMFA	Ours
在 Market-1501 数据集的 *mAP* 值	8.0	12.4	20.5	26.5	22.8	27.4	36.02
在 DukeMTMC-reID 数据集的 *mAP* 值	4.8	7.3	16.4	23.0	22.3	24.7	40.64

表 3　在 Market-1501 数据集的结果

Tab. 3　Results of the Market-1501 dataset

Methods	*Rank* - 1	*Rank* - 5	*Rank* - 10
LOMO	27.2	41.6	49.1
UMDL	34.5	52.6	59.6
PUL	45.5	60.7	66.7
PTGAN	38.4	-	66.1
TJ-AIDL	58.2	74.8	81.1
SPGAN	51.5	70.1	76.8
MMFA	56.7	75.0	81.8
Ours	**59.35**	**80.85**	**84.06**

表 4　在 DukeMTMC-reID 数据集的结果

Tab. 4　Results of the DukeMTMC-reID dataset

Methods	*Rank* - 1	*Rank* - 5	*Rank* - 10
LOMO	12.3	21.3	26.6
UMDL	18.5	31.4	37.6
PUL	30.0	43.4	48.5
PTGAN	27.4	-	50.7
TJ-AIDL	44.3	59.6	65.0
SPGAN	41.1	56.6	63.0
MMFA	45.3	59.8	66.3
Ours	**55.75**	**69.62**	**73.66**

3.4 联合损失函数中权重 λ 的分析

本文还在 Market-1501 和 DukeMTMC-reID 这两个大型数据集上对总损失中参数 λ 的影响进行了实验分析,选取 *Rank* - 1 和 *mAP* 作为评测指标,实验结果如图 6,图 7 所示。可以发现,λ 的区间在[0,1]之间,*Rank* - 1 的结果首先随着 λ 的值呈现平稳上升的趋势,当 λ = 0.7 时,到达最高点之后下降。*mAP* 的结果虽然有所曲折,但也是呈现上升趋势,并且当 λ = 0.7 时取得最好的结果,随之下降。即 λ 值设置为 0.7 可以取得比较好的结果。由于学习到了更有判别力的联合判别性特征,因此将全局损失和局部损失组合起来可以得到更好的结果。

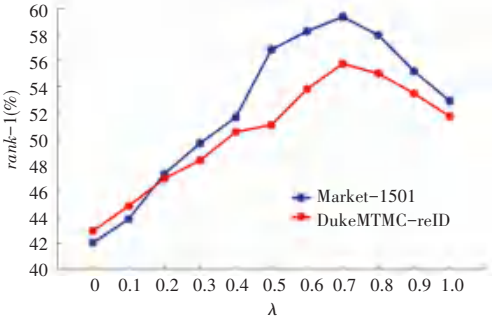

图 6 λ 值对 *rank*-1 的影响

Fig. 6 The effect of λ on *rank*-1

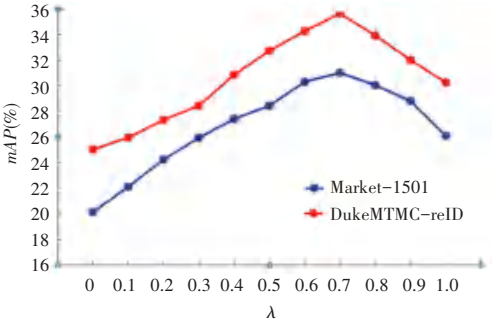

图 7 λ 值对 *mAP* 的影响

Fig. 7 The effect of λ on *mAP*

4 结束语

行人重识别任务随着计算机视觉和模式识别领域的快速发展而发展,成为该研究方向中的一个重要分支。作为智能视频监控方向上的研究支撑,对于实现跨摄像机研究中的目标跟踪和行为分析等一系列面向智能视频监控的应用难题起到非常大的推进作用。本文基于全局与局部特征的无监督学习框架,提出了一种联合判别性特征学习方法来解决重识别任务,并实验验证了方法中每一部分的有效性,证明了所提出的方法对于解决行人重识别任务具有显著的效果。

参考文献

[1] GONG S, CRISTANI M, YAN S, et al. Person Re-Identification [J]. Advances in Computer Vision & Pattern Recognition, 2014, 42(7): 301-313.

[2] GHEISSARI N, SEBASTIAN T B, HARTLEY R. Person Re-Identification Using Spatiotemporal Appearance[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2006: 1528-1535.

[3] SUN Y F, LIANG Z, YANG Y, et al. Beyond part models: Person retrieval with refined part pooling [C]// European Conference on Computer Vision. 2018: 480-496.

[4] JADERBERG M, SIMONYAN K, ZISSERMAN A. Spatial transformer networks [J]. Advances in neural information processing systems. 2015, 28: 2017-2025.

(上接第 145 页)

化人口预测提供了切实可行的方法。

参考文献

[1] WU Y, HUANG T. Vision-based gesture recognition: A review. International Gesture Workshop, 1999, 1739(1):103-115.

[2] 余乐安. 基于最小二乘近似支持向量回归模型的电子商务信用风险预警[J]. 系统工程理论与实践,2012,32(3):508-514.

[3] HOLLAND J H. Adaptation in natural and artificial systems[M]. Ann Arbor,MI:University of Michigan Press,1975:228-234.

[4] GOLDBERG D E. Genetic algorithms in search, optimization and machine learning[M]. New Jersey: Addison-Wesley, 1989: 31-35.

[5] DEJONG T M, DASILVA D, VOS J, et al. Using functional-structural plant models to study, understand and integrate plant development and ecophysiology[J]. Annals of Botany, 2011, 108(6):987-989.