

基于难样本混淆增强特征鲁棒性的行人重识别

郝玲, 段继忠, 庞健

(昆明理工大学信息工程与自动化学院, 昆明 650504)

摘要: 随着深度学习的兴起, 行人重识别逐渐成为计算机领域的热门话题。它通过给定的查询行人图像进行跨摄像机检索, 找出与查询身份相匹配的行人。然而, 由于受到不同视角下的背景、光照等因素影响, 采集到的行人图像中存在大量的难样本, 利用这些难样本训练得到的模型识别性能低下, 缺乏鲁棒性。因此, 为了提高模型对难样本的鉴别能力, 设计了一种新颖的通过混淆因子合成具有难样本信息图像的方法。对于每批输入图片, 通过相似性度量寻找每张图像对应的难样本, 结合混淆因子合成具有难样本信息的新图像再以有监督的方式促使模型挖掘难样本信息, 从而提高模型鲁棒性。大量对比实验表明, 所提方法在主流数据集上达到了较高的识别率, 消融实验证明了所提方法的有效性。

关键词: 深度学习; 行人重识别; 混淆因子; 相似性度量; 难样本

中图分类号: TP391 **文献标志码:** A

Person Re-identification Based on Hard Negative Sample Confusion to Enhance Robustness of Features

HAO Ling, DUAN Jizhong, PANG Jian

(School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650504, China)

Abstract: With the rise of deep learning, person re-identification has gradually become a hot topic in the computer vision field. It performs cross-camera retrieval through a given query image, and finds the images that match the query identity. However, due to the factors such as background and illumination under different cameras, there are a large number of hard negative samples in the collected pedestrian datasets, and the performance of the model trained using these samples is bad and lacks robustness. Therefore, in order to improve the ability of the model to discriminate such negative samples, a novel method of synthesizing images with hard negative samples information through confusion factors is designed. For each input batch images, the similarity measurement is used to find the hard negative sample corresponding to each image, the new generated images with the clues of negative samples are synthesized through the confusion factor, and the model is prompted to mine the negative samples information in a supervised manner thus improving the model robustness. A large number of comparative experiments show that the proposed method achieves high performance on the mainstream datasets. The ablation study proves the effectiveness of the proposed method.

Key words: deep learning; person re-identification; confusion factor; similarity measurement; hard negative sample

引言

随着深度学习的出现,行人重识别技术得到了广泛的发展^[1],它是利用计算机视觉的方法判断图像或视频中是否存在特定行人的技术,由此也常被认为是图像检索的子问题。具体来说是将同一个人不同环境、不同摄像头下所捕捉到的图像关联起来,从而实现跨域跨设备的检索。行人重识别技术^[2-3]的出现弥补了目前固定摄像头的视觉局限性、人工搜索时间长等问题,由此在刑事侦查、智能安防、无人超市等领域得到了广泛的应用。

目前,行人重识别过程主要分为两步:(1)行人图片的有效特征提取;(2)对有效特征进行距离度量。然而由于数据集中大量难样本的干扰(图1),使得模型不能有效地优化类内、类间距离。三元组损失函数^[4]是一种被广泛用于难样本信息挖掘的方法,即选择3张图片作为输入,分别命名为固定样本 a ,难正样本 p 和难负样本 n ,它在特征层面减小类内距离的同时增大类间距离,然而该方法只关注如何改进网络提取出来的特征,忽略了数据集中存在的大量难样本图像中的信息。



图1 锚点图像集合及难负样本集合

Fig.1 Positive sample sets and negative sample sets

随着深度学习的不断发展,2014年Goodfellow等^[5]提出了一种新的深度学习模型即生成式对抗网络(Generative adversarial networks, GAN),2018年Wei等^[6]首次将GAN应用到Re-ID中,将利用GAN生成的不同相机风格的行人图片以有监督的方式重新训练模型。随后提出的CycleGAN^[7-8]使用两个生成器和两个判别器来进行不同域间行人图像的风格转换,SPGAN^[9]则是在CycleGAN的基础上加入前景约束,在尽可能多地保留行人外观的前提下改变图片的背景风格。然而上述基于GAN的方法都需要建立十分复杂的网络,而且训练过程不稳定存在模型崩塌的风险。这类方法生成不同风格的行人图像仅仅是一种对数据的增强,提升了样本多样性,没有赋予模型挖掘难样本信息的能力。

针对上述方法中存在的问题,本文提出一种难样本混淆增强特征鲁棒性的行人重识别方法。主要工作如下:(1)使用基于ResNet-50的主体框架,充分利用难样本在图像层面的信息,赋予网络提取更具有鉴别性的特征的能力;(2)将距离度量的结果作为相似性找到与原图最为相似的难样本,在图像层面进行难样本混淆,增加了训练样本中信息的多样性,提高了模型泛化能力;(3)不需要额外的生成器和鉴别器,在端到端的模型中深度挖掘难样本信息,增强模型鲁棒性的同时让网络本身的结构相对简化。在不同数据集上的实验结果表明了本文方法的优越性,能够有效提升行人重识别性能。

1 基于深度学习的行人重识别

1.1 标签平滑交叉熵函数

在基于深度学习的行人重识别中,一般使用交叉熵损失函数^[9]对图像的特征进行分类,在距离度量中通常采用三元组损失函数^[4]来进行距离约束。定义数据集中的图像集合为 $X = \{x_i\}_{i=1}^{N_{\text{img}}}$, N_{img} 表示数据集中的图像总数, $M = \{m_i\}_{m=1}^N$ 表示图像标签, N 为数据集中的行人身份总数。首先利用交叉熵损失函数优化模型,让网络对不同行人具有初始的鉴别能力,如式(1)所示。

$$L_{\text{id}} = \frac{1}{B \times N} \sum_{b=1}^B \sum_{i=1}^N -q_i \ln(p_i(x_b)) \quad (1)$$

式中: B 为每批训练图像数量, $p_i(x_b)$ 为将输入图像 x_b 预测为第 i 个行人的概率, q_i 为指示函数,用来判断网络的预测结果是否正确。

$$q_i = \begin{cases} 0 & m_b \neq i \\ 1 & m_b = i \end{cases} \quad (2)$$

式中 m_b 表示与输入图像 x_b 对应的真实行人身份标签。由于常用的交叉熵损失函数过度地依赖训练集中正确的行人标签,这使得网络容易出现过拟合的现象。通常解决上述问题,标签平滑(Label smoothing, LS)^[10]是最常用的方法,因此对 q_i 进行改写,具体过程为

$$q_i = \begin{cases} \frac{\epsilon}{N} & m_b \neq i \\ 1 - \frac{N-1}{N} \epsilon & m_b = i \end{cases} \quad (3)$$

式中: ϵ 为平滑超参数,在训练时作为容错因子引入到交叉熵损失函数中,实验中设置 $\epsilon=0.1$ 。由于训练集和测试集中的行人是互不重合的,标签平滑可以有效防止训练过程中出现的过拟合,进一步提升模型泛化能力。

1.2 三元组函数

常用的另一种优化特征的损失函数为三元组损失函数^[4]。三元组由3张图片构成,分别命名为:锚点 a ,难正样本 p ,难负样本 n 。其中难正样本表示一批图像中与锚点图像身份一致,但与锚点具有较低相似度的行人图像,难负样本表示与锚点图像身份不一致,但与锚点具有较高相似度的行人图像。三元组损失函数表达为

$$L_t = (d_{a,p} - d_{a,n} + \delta)_+ \quad (4)$$

式中 δ 为训练过程中设置的阈值参数。式(4)表示括号内的数值与0相比较,选择大的数值作为损失函数的结果。本文选用常用的欧式距离作为距离度量方式,后续实验中会分析不同度量方式产生的结果。三元组损失的学习目标是增大锚点与难负样本(类间)的距离,减小锚点与难正样本(类内)的距离,并采用欧氏距离来表示样本之间的距离。在一个欧氏距离空间 R^d 中,度量学习的目的是得到一个能够将 x 映射到 R^d 中的函数 $f(x)$ 以实现三元组学习的目标,具体表示为

$$d_{a,p} = \|f(x_i^a) - f(x_i^p)\|_2^2 \quad (5)$$

$$d_{a,n} = \|f(x_i^a) - f(x_i^n)\|_2^2 \quad (6)$$

式中: x_i^a 为选定的一个样本作为锚点。 x_i^p 和 x_i^n 分别为通过欧式距离选出来的难正样本和难负样本。

2 基于难样本混淆增强特征鲁棒性的行人重识别

2.1 难样本的挑选和混淆

图2为本文的网络结构图,本实验选择ResNet-50^[11]作为编码器进行特征提取,在特征层面利用欧氏距离找出每个输入图像对应的难样本,通过混淆因子对输入图像和与之对应的难样本图像进行样本混淆,利用新生成的混淆样本重新优化网络,使网络进一步挖掘难样本信息从而提取更加具有鉴别性的特征。

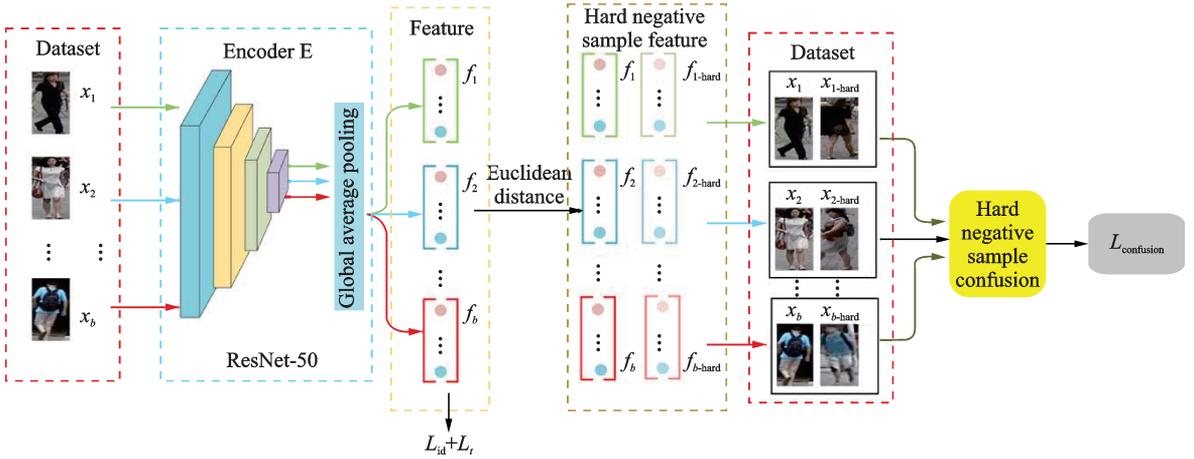


图2 网络结构示意图

Fig.2 Schematic diagram of network structure

所提方法的一个关键点在于寻找图像的难样本,数据集中的难样本很多,要挑选与到目标图像最相似的难样本是本节主要目标。给定一个锚点样本 x_i^a , a 表示该样本的身份,将其通过网络得到的特征 $f(x_i^a)$ 与训练批次中其他身份的图像特征逐一进行距离度量,找到与 x_i^a 最为相似的难样本。具体过程如下

$$k = \arg \min_j \|f(x_i^a) - f(x_j^l)\|_2^2 \quad l \neq a, j = 1, 2, 3, \dots, B \quad (7)$$

式中: B 为每个训练批次的图像数量 (Batch size), 定义 $x_{i_k}^l$ 为第 i 个样本 x_i^a 对应的难样本, 其身份为 l 。通过式 (7) 可以找到每张图像对应的难样本图像, 接下来进行难样本混淆。难样本混淆的目的是使锚点样本 x_i^a 带有其难样本图像 $x_{i_k}^l$ 的部分信息, 从而使训练得到的网络更加具有鉴别性不易被难样本干扰。每批训练图像和其对应的难样本都需进行如下难样本混淆过程

$$x_{i\text{-new}} = x_i^a \times \lambda + x_{i_k}^l \times (1 - \lambda) \quad i = 1, 2, \dots, B \quad (8)$$

式中: $x_{i\text{-new}}$ 为生成的混淆样本, λ 为混淆因子。不同的 λ 会影响生成的混淆样本 $x_{i\text{-new}}$ 中难样本信息含量, 后续实验中会展示不同 λ 的生成图像可视化结果, 并且讨论 λ 的变化对性能的影响。对于混淆样本 $x_{i\text{-new}}$ 对应的标签, 应与锚点图像 x_i^a 的标签保持一致, 同样为 m_i 。因为所提出的难样本混淆方法目的是让当前锚点图像携带有部分难样本图像信息, 从而生成一个新的混淆图像, 通过这类混淆图像提升模型鉴别性, 使网络在识别行人时不容易被难样本信息干扰。

2.2 鲁棒性特征的提取

通过 2.1 节的训练过程找到难样本图像, 并进行难样本混淆生成一批含有难样本信息的新图像, 将

生成的含有难样本信息的新图像重新利用 1.1 节介绍的平滑交叉熵函数以有监督的方式训练网络,以提高模型的鲁棒性,公式如下

$$L_{\text{confusion}} = \frac{1}{B \times N} \sum_{b=1}^B \sum_{i=1}^N -q_i \ln(p_i(x_{b\text{-new}})) \quad (9)$$

目前大部分行人重识别研究都在特征层面挖掘难样本信息^[12],很少有根据难样本的图像信息降低其对模型性能的干扰,而且所提方法不需要额外的生成和鉴别网络,相较于基于 GAN^[6]的方法训练速度更快且更稳定。最后综合上述本文的损失函数表示为

$$L_{\text{total}} = L_{\text{id}} + L_t + \beta \times L_{\text{confusion}} \quad (10)$$

式中超参数 β 用来衡量所提方法对整体性能的影响,后续实验会进一步讨论 β 的变化对整体模型的影响。

3 实验结果与分析

3.1 数据集及评估指标

为了验证难样本混淆算法的有效性,本文在两个大型主流数据集 DukeMTMC-ReID 和 Market-1501 以及两个小数据集 GRID^[13]和 PRID2011^[14]上进行了测试,采样图像如图 3 所示。4 个数据集都由训练集和测试集两部分组成,见表 1。并且这两部分的行人身份互不重合,查询数据集(Query)由测试集中抽取的部分样本组成。Market-1501^[15]数据集于 2015 年在清华大学采集得到,数据集中的图像来自于 5 个高分辨率摄像头和 1 个低分辨率摄像头,共计 1 501 个行人,32 668 张图片。DukeMTMC-ReID^[16]在杜克大学内采集得到,图像来自 8 个不同的摄像头,共计 1 404 个行人和 34 183 张图片。PRID2011 数据集提供了 2 个不同静止监控摄像头下拍摄到的行人图像,也是一个规模不大的数据集。这个数据集中行人图像的检测框都是人工手动提取,并且两个摄像头拍摄到的相同身份行人有 200 个,这些行人随机分为两组,分别作为训练集和查询集。GRID 数据集采集于 6 个不重叠相机,共包含 250 对行人图像。其中该数据集还有 775 张剩余图像,这类图像均为干扰样本,没有对应的身份信息,在训练过程中不予使用。

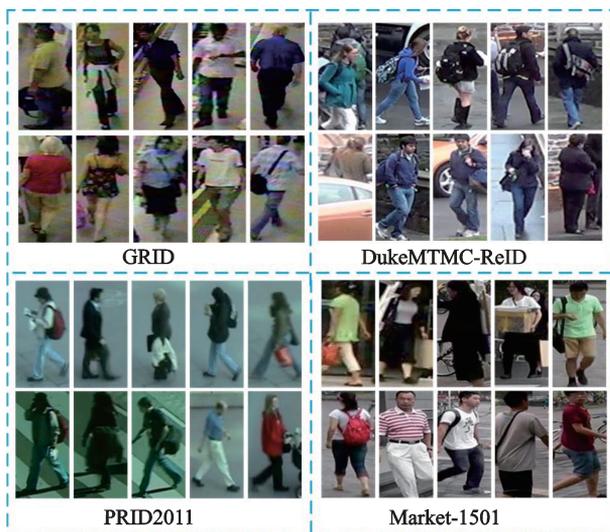


图 3 4 个不同数据集下的行人样本

Fig.3 Samples from four different datasets

表 1 数据集介绍

Table 1 Introduction to the datasets

Data set	Date	Number of samples	Number of ID
Market-1501	Test	19 732	750
	Train	12 936	751
	Query	3 368	750
DukeMTMC-ReID	Test	17 661	702
	Train	16 522	702
	Query	2 228	702
PRID2011	Test	649	649
	Train	200	100
	Query	100	100
GRID	Test	900	126
	Train	250	125
	Query	125	125

实验中,本文采用累计匹配曲线(Cumulative match characteristic, CMC)中的Rank-1、Rank-5、Rank-10以及平均精度均值(Mean average precision, mAP)^[15]进行性能评估,对于所有数据集,使用单个查询图像(Query)在样本库(Gallery)中进行检索,没有使用重排序(Re-ranking)或随机擦除(Random erasing)等数据增强的方法。

3.2 实验环境和参数设置

本文采用在Image-Net预训练过的ResNet-50^[11]作为主干网络进行特征提取,最后额外增加了一个全局平均池化层。本文实验所采用的Tricks参考文献[17],对学习率进行了Warm up,初始学习率为0.000 35,前10代线性增加学习率至0.003 5,在40代和70代分别降低至当前值的1/10。将主干网络的最后一个空间下采样的步长(Last stride)改为2。在最后一层的分类器前面加入Batch normalization(BN)层,消除特征分布的偏差。实验中采用Pytorch深度学习平台,CPU为I7-8700,内存为32 GB,显卡为11 GB的GTX1080Ti。所有训练图像剪裁为256×128作为网络的输入,训练批次的大小设为64,其中为了满足三元组的要求,每个身份采样4张图像。使用Adam^[18]对网络参数进行优化,其中权重衰减因子设为0.000 5,初始学习率为0.000 35,在40代和70代学习率下降为当前值的1/10,总共训练120代。交叉熵损失函数中的 ϵ 设为0.1,三元组损失函数中的 δ 设为0.3,混淆因子 λ 设为0.98。

3.3 与其他方法比较

为了验证本文提出的算法性能,首先在大数据集Market-1501和DukeMTMC-ReID上与当前性能较为优异的方法进行了比较。

首先在Market-1501数据集上与3类方法进行了比较:(1)基于深度学习和属性的方法。PAN^[19]、APR^[19]、TFusion-sup^[20]、FPO^[21]、DAI^[22]、LWL^[23];(2)基于注意力机制和特征融合的方法。HACNN^[24]、MCFE^[25];(3)基于三元组改进的方法。TriNet^[12]、IDTL^[26]。表2中,本文提出的算法在Market-1501数据集上Rank-1和mAP分别达到了94.0%和82.5%,性能上超过了所有对比的方法,相较于性能较高的挖掘难样本信息的方法IDTL,所提方法在Rank-1和mAP上分别高于其2.6%和6.1%。对于传统的优化特征类内类间距离的三元组方法TriNet,本文方法在性能上大幅领先,说明了所提算法的优越性。

同样地,将本文提出的算法在DukeMTMC-ReID数据集上与目前较为主流的上述3类算法进行对比,额外加入行人掩膜引导的方法Human parsing^[27]进行比较,该方法在图像层面降低了背景对行人的干扰。性能对比如表3所示,相较于目前性能较高的行人掩膜引导方法Human parsing,本文算法在Rank-1和mAP上分别提升了1.1%和0.5%。在和难样本信息挖掘方法IDTL的比较中,本文方法在Rank-1和mAP上优势明显。尽管两项指标提升的效果没有在Market-1501数据集上那么优越,但是也足以说明模型的鲁棒性。

接下来,在小数据集PRID和GRID上与其他方法进行比较。小数据与大数据集Market-1501和DukeMTMC-ReID有显著差别,主要体现在行人图像数量较少,一般方法都是用Rank-1、Rank-5、Rank-10和Rank-20进行性能对比,而不选用mAP,因为有的数据集中同一个人可能只有一张图像,因此无法计算mAP。在PRID上(表4),所对比的方法有RPLM^[28]、KCCA^[29]、MFA^[30]、kLFDA^[30]、NFST^[31]、aMTL-LORAE^[32]。与基于度量学习最好的方法RPLM相比,所提方法在Rank-1、Rank-5、Rank-10、Rank-20上分别高出26.6%、36.8%、35.9%和31.5%。和基于核的鉴别性空间学习方法中最优异的kLFDA相比,所提方法的Rank-1达到了41.6%,而kLFDA仅为22.4%。与属性信息辅助的方法aMTL-LORAE对比,所提方法在Rank-1上提高了23.6%。和其他方法对比,所提方法均体现了大幅领先。

在 GRID 上的实验结果如表 5 所示,所对比的方法包括 GOG+XQDA^[33]、MHF^[34]、LOMO+DMLV^[35]、MKFSL^[36]、CRAFT-MFA^[37]、DMVFL^[38]、CSPL^[39]、SRR^[40]和 DSRPDL^[41]。在 Rank-1、Rank-5、Rank-10、Rank-20 的对比中,所提方法比手工提取特征的方法 GOG+XQDA 分别高出 11.86%、9.1%、7.12% 和 7.84%,证明了基于深度学习的方法在行人重识别数据集 GRID 上的性能远远超出手工提取特征的方法。和当前最好的利用字典学习进行特征学习的方法 DSRPDL 相比较,所提方法在 Rank-1 等 4 个评价指标上都大幅领先。综上证明了所提方法在小数据集上同样具有很强的竞争力。

表 2 Market-1501 结果对比

Table 2 Results comparison of Market-1501 %

方法	Rank-1	Rank-5	Rank-10	mAP
TriNet	84.9			69.1
TFusion-sup	73.1	86.4	90.50	
PAN	82.8			63.4
APR	84.3	93.2	95.20	64.7
LWL	86.2			79.8
HA-CNN	91.2			75.7
MCFE	91.4			75.7
IDTL	91.4	96.6	97.70	76.4
DAI	92.0	96.6	97.80	77.2
FPO	93.4		98.45	82.2
本文方法	94.0	97.9	98.80	82.5

表 3 DukeMTMC-ReID 结果对比

Table 3 Comparison results of DukeMTMC-ReID %

方法	Rank-1	Rank-5	Rank-10	mAP
APR	70.7			51.9
PAN	71.6			51.5
LWL	78.6			61.5
HA-CNN	80.5			63.8
IDTL	80.8	89.5	92.0	63.1
DAI	81.3	90.0	92.3	66.0
MCFE	82.1			60.6
Human parsing	84.4	91.9	93.7	71.0
本文方法	85.5	92.7	94.7	71.5

表 4 PRID 结果对比

Table 4 Comparison results of PRID %

方法	Rank-1	Rank-5	Rank-10	Rank-20
RPLM	15.0	32.0	42.0	54.0
KCCA	14.3	37.4	47.6	62.5
MFA	22.3	45.6	57.2	68.2
kLFDA	22.4	46.5	58.1	68.6
NFST	29.8	52.9	66.0	76.5
aMTL-LORAE	18.0	37.4	50.1	66.6
本文方法	41.6	68.8	77.9	85.5

表 5 GRID 结果对比

Table 5 Comparison results of GRID %

方法	Rank-1	Rank-5	Rank-10	Rank-20
GOG+XQDA	24.70	47.0	58.40	69.0
MHF	19.84	40.48	56.88	62.32
LOMO+DMLV	19.80	35.60	46.50	58.10
MKFSL	26.40	47.20	55.80	67.80
CRAFT-MFA	22.40	49.90	61.80	71.70
DMVFL	26.70	49.20	58.60	68.90
CSPL	25.84	46.48	58.40	70.16
SRR	26.56	46.32	56.16	66.80
DSRPDL	31.52	48.95	60.24	72.72
本文方法	36.56	56.10	65.52	76.84

3.4 消融实验

通过图 2 可知,本文提出的难样本混淆算法主要包含 3 项损失函数:使网络提取行人鉴别性特征的身份损失(L_{id}),特征层面约束类内和类间距离进一步增强行人身份信息的三元组损失(L_t)以及难样本混淆后新数据的交叉熵损失($L_{confusion}$)。接下来进行消融实验证明本文所提方法的有效性。实验结果如表 6 所示,可以看出仅使用身份约束 L_{id} 时 Rank-1 和 mAP 分别达到了 93.2% 和 82.0%,加入难样本混淆后 $L_{id} + L_{confusion}$ 性能进一步获得提升。结合所有损失函数($L_{total} = L_{id} + L_t + L_{confusion}$)后性能达到了

表6 消融实验
Table 6 Ablation experiment

Method	Market-1501		DukeMTMC-ReID	
	Rank-1	mAP	Rank-1	mAP
L_{id}	93.2	82.0	84.7	71.1
$L_{id} + L_t$	93.7	82.2	85.1	71.4
$L_{id} + L_{confusion}$	93.5	82.2	85.1	71.3
$L_t + L_{confusion}$	93.0	81.8	84.2	70.9
$L_{total} = L_{id} + L_t + L_{confusion}$	94.0	82.5	85.5	71.5

最优,说明所提方法在加入三元组后能够进一步提高特征的鉴别性,与在特征层面进行约束的三元组函数具有互补性。

3.5 超参数分析

本节内容是对式(10)中的超参数 β 进行性能分析,该超参数用来权衡所提难样本混淆方法对整体性能的影响。在两个较大数据集 Market-1501 和 DukeMTMC-ReID 上进行参数分析,以 Rank-1 和 mAP 进行说明。参数分析结果如图 4 所示,从 Market-1501 数据集可以看出,当 β 取值为 1 时 Rank-1 和 mAP 达到了最优值,当 β 大于 1 时性能显著下降,因此适用于 Market-1501 数据集的 β 最优值为 1。当测试数据集为 DukeMTMC-ReID, β 取值为 0.8 时取得最优性能,同样地; β 越大性能下降越明显,说明所提方法权重过大时可能会对模型造成干扰。

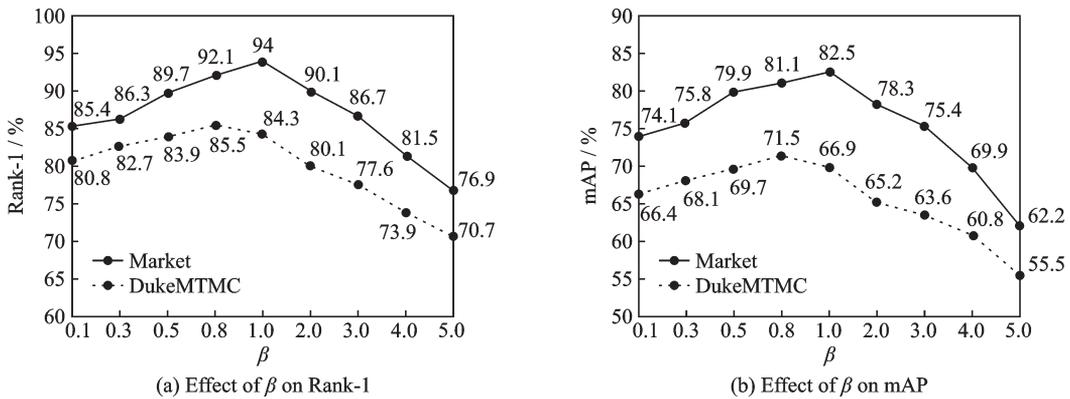


图4 β 对 Rank-1 和 mAP 的影响

Fig.4 Effect of β on Rank-1 and mAP

3.6 混淆因子分析

在本节中,讨论式(8)中控制难样本混淆程度的混淆因子 λ 的作用以及不同 λ 下的可视化图像。 λ 参数的取值范围是 $[0, 1]$,该值越小,生成的混淆图像越趋近于难样本,反之越像原图像。图 5 中,列出了 Market-1501 中两组样本的难样本混淆图像,其中 λ 取值分别为 0.1、0.3、0.5、0.7 和 0.98。从生成的混淆样本可以看出, λ 取值越小时,其难样本在图像中所占的比重越高,例如第一行中 $\lambda=0.1$ 时生成的样本和难样本几乎相同,而随着 λ 的增大,当 $\lambda=0.98$ 时,生成的样本趋近于正样本同时带有部分难样本的信息。可视化进一步说明了随着 λ 的改变,其生成的混淆图像携带有不同含量的难样本信息,接下来讨论不同 λ 对模型性能的影响。

为了分析不同 λ 对模型性能的影响,在 Market-1501 数据集上进行实验,并且以 Rank-1 和 mAP 进行

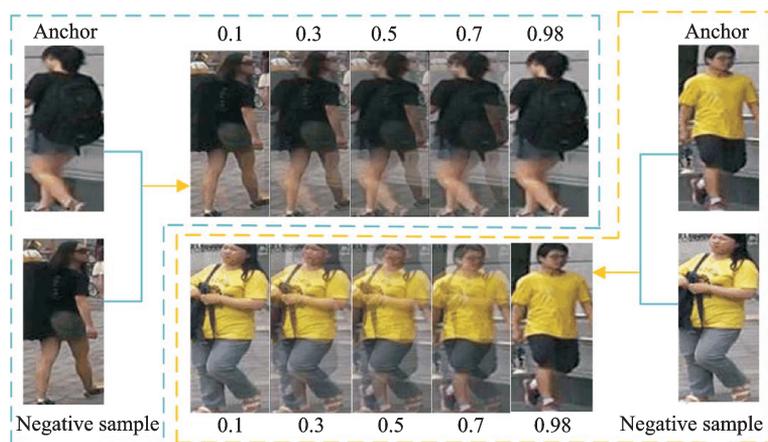
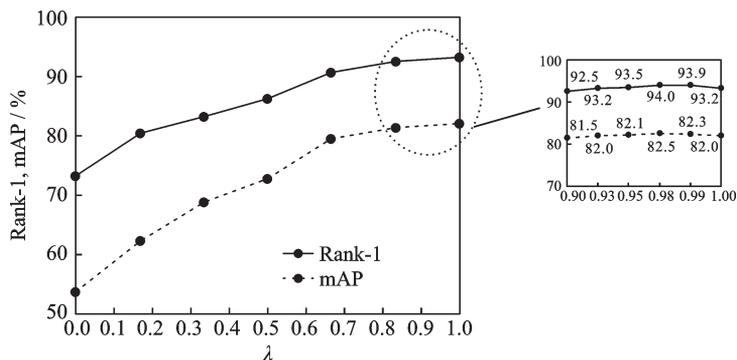


图5 难样本混淆可视化

Fig.5 Visualization of negative sample confusion

说明。实验结果如图6所示,可以看出 λ 越小对模型性能影响越大,因为其生成的混淆样本携带大量难样本信息,而交叉熵损失函数会将该样本误分类为原图像而造成模型性能下降。随着 λ 的增大,混淆样本中正样本所占有的信息越大,其对性能提高也越明显。当 λ 处于(0.9, 1)时,模型的性能逐渐在高点趋于稳定,且在取值为0.98时达到了最优。对于 λ 的分析说明了当生成的混淆样本中带有少量难样本信息时确实能够提升模型的鲁棒性。

图6 λ 对Rank-1和mAP的影响Fig.6 Effect of λ on Rank-1 and mAP

3.7 距离度量分析

本节对难样本选取过程中采用的度量方式进行分析。实验中,拟采用3种不同的度量方式选取训练批次图像中的难样本进行混淆。实验结果如表7所示,在实验过程中,除了度量方式不同,其余参数均保持一致。由表7结果可以看出,曼哈顿距离的效果最差,原因是该距离度量方式衡量的是特征间距离绝对值之和,导致求和后的值相对较大,降低了样本间的区分度。反观欧式距

表7 不同度量方式的结果

Table 7 Results with different measures %

度量方式	Market-1501		DukeMTMC-ReID	
	Rank-1	mAP	Rank-1	mAP
曼哈顿距离	92.0	80.4	84.1	70.9
余弦距离	93.2	81.8	84.9	71.1
欧式距离	94.0	82.5	85.5	71.5

离的结果,其度量方式是特征差的平方和,使特征差异在数值上的体现较为明显,更易挑选出真实的难样本,而余弦距离更多的是反映两个特征间的角度差异,对特征整体分布的差异表现不明显,虽然相较于曼哈顿距离的结果有所提高,但仍逊于欧式距离的结果。

4 结束语

本文针对不同视角下的背景、光照等因素造成数据集中存在大量难样本的问题,提出一种难样本混淆增强特征鲁棒性的行人重识别方法,通过相似性度量寻找每张图像对应的难样本,利用混淆因子合成具有难样本信息的新图像再以有监督的方式促使模型挖掘难样本信息,从而提高模型鲁棒性。在多个数据集上的实验结果表明,本文算法在性能方面优于目前难样本信息挖掘的方法和一些主流的深度学习方法。消融实验进一步证明了所提方法的有效性,通过后续的参数分析选择了适用于不同数据集的最优参数,最后由难样本可视化阐明混淆过程以及混淆因子的作用。未来会研究如何在图像风格更加相似的图像中更加快速有效地找到难样本,进一步提高行人重识别的精度。

参考文献:

- [1] 曹亮,王洪元,戴臣超,等.基于多样性约束和离散度分层聚类的无监督视频行人重识别[J].南京航空航天大学学报,2020,52(5):752-759.
CAO Liang, WANG Hongyuan, DAI Chenchao, et al. Unsupervised video-based person re-identification based on diversity constraint and dispersion hierarchical clustering[J]. Journal of Nanjing University of Aeronautics & Astronautics, 2020, 52(5): 752-759.
- [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.
LUO Hao, JIANG Wei, FAN Xing, et al. A survey on deep learning based person re-identification[J]. Acta Automatica Sinica, 2019, 45(11): 2032-2049.
- [3] 张文文,王洪元,万建武,等.基于稀疏学习的行人重识别算法[J].数据采集与处理,2018,33(5):91-100.
ZHANG Wengwen, WAN Hongyuan, WANG Jianwu, et al. A sparsity-learning-based person re-identification algorithm[J]. Journal of Data Acquisition and Processing, 2018, 33(5):91-100.
- [4] LIU Hao, FENG Jiashi, QI Meibin, et al. End-to-end comparative attention networks for person re-identification[J]. IEEE Transactions on Image Processing, 2017, 26(7):3492-3506.
- [5] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets[C]//Proceedings of the 2014 Advances in Neural Information Processing Systems(NIPS). Montréal, Canada: MIT Press, 2014: 2672-2680.
- [6] WEI Longhui, ZHANG Shiliang, TIAN Qi, 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. [S.l.]: IEEE, 2018: 79-88.
- [7] HUANG Yan, XU Jinsong, WU Qiang, et al. Multi-pseudo regularized label for generated data in person re-identification[J]. IEEE Transactions on Image Processing, 2019, 28(3):1391-1403.
- [8] CHEN Xi, DUAN Yan, HOUTHOOFT R, et al. Info GAN: Interpretable representation learning by information maximizing generative adversarial nets[J]. Advances in Neural Information Processing Systems, 2016, 29: 2172-2180.
- [9] DENG Weijian, ZHENG Liang, KANG Guoliang, et al. Image-image domain adaptation with preserved self-similarity and domain-dissimilarity for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2018: 994-1003.
- [10] SZEGEDY C, VANHOUCHE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA: IEEE, 2016: 2818-2826.
- [11] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition [C]//Proceedings of International Conference on Computer Vision and Pattern Recognition. [S.l.]: [s.n.], 2016:770-778.
- [12] HERMANS A, BEYER L, LEIBE B. In defense of the triplet loss for person re-identification[J/OL]. (2017-05-17) [2020-04-20]. <https://arxiv.org/abs/1703.07737>.
- [13] LOY C C, XIANG T, GONG S. Time-delayed correlation analysis for multi-camera activity understanding[J]. International

- Journal of Computer Vision, 2010, 90(1): 106-129.
- [14] HIRZER M, BELEZNAI C, ROTH P M, et al. Person re-identification by descriptive and discriminative classification[C]// Proceedings of Scandinavian Conference on Image Analysis. [S.l.]: Springer, 2011: 91-102.
- [15] ZHENG Liang, SHEN Liyue, TIAN Lu, et al. Scalable person re-identification: A benchmark[C]//Proceedings of the 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE, 2015: 1116-1124.
- [16] RISTANI E, SOLERA F, ZOU R, et al. Performance measures and a data set for multi-target, multicamera tracking[C]// Proceedings of the European Conference on Computer Vision. Cham: Springer, 2016: 17-35.
- [17] LUO H, GU Y, LIAO X, et al. Bag of tricks and a strong baseline for deep person re-identification[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. [S.l.]: IEEE, 2019.
- [18] KINGMA D P, BA J. Adam: A method for stochastic optimization[J] The International Conference for Learning Representations (ICLR). arXiv preprint arXiv:1412.6980, 2014.
- [19] LIN Yutian, ZHENG Liang, ZHENG Zhedong, et al. Improving person re-identification by attribute and identity learning[J]. Pattern Recognition, 2019, 95:151-161.
- [20] LV Jianming, CHEN Weihang, LI Qing, et al. Unsupervised cross-dataset person re-identification by transfer learning of spatial-temporal patterns[C]//Proceedings of Computer Vision and Pattern Recognition. [S.l.]: [s.n.], 2018:7948-7956.
- [21] TANG Y, YANG X, WANG N, et al. Person re-identification with feature pyramid optimization and gradual background suppression[J]. Neural Networks, 2020, 124: 223-232.
- [22] SI T, ZHANG Z, LIU S. Discrimination-aware integration for person re-identification in camera networks[J]. IEEE Access, 2019, 7: 33107-33114.
- [23] CHEN Y, DUFFNER S, STOIAN A, et al. List-wise learning-to-rank with convolutional neural networks for person re-identification[J]. Machine Vision and Applications, 2021, 32(2): 1-14.
- [24] LI Wei, ZHU Xiadian, GONG Shaogang. Harmonious attention network for person re-identification[C]//Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR). New York: IEEE, 2018:2285-2294.
- [25] 徐龙壮, 彭力. 基于多尺度卷积特征融合的行人重识别[J]. 激光与光电子学进展, 2019, 56(14):221-227.
XU Longzhuang, PENG Li. Person reidentification based on multiscale convolutional feature fusion[J]. Laser & Optoelectronics Progress, 2019, 56(14) :221-227.
- [26] YUAN Ye , CHEN Wuyang , YANG Yang, et al. In defense of the triplet loss again: Learning robust person re-identification with fast approximated triplet loss and label distillation[C]//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). [S.l.]: IEEE, 2020:354-355.
- [27] KALAYEH M M, BASARAN E , GOKMEN M. Human semantic parsing for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2018: 1062-1071.
- [28] HIRZER M, ROTH P M, KÖSTINGER M, et al. Relaxed pairwise learned metric for person re-identification[C]// Proceedings of European Conference on Computer Vision. Berlin, Heidelberg: Springer, 2012: 780-793.
- [29] LISANTI G, MASI I, DEL BIMBO A. Matching people across camera views using kernel canonical correlation analysis[C]// Proceedings of the International Conference on Distributed Smart Cameras. [S.l.]: [s.n.], 2014: 1-6.
- [30] XIONG F, GOU M, CAMPS O, et al. Person re-identification using kernel-based metric learning methods[C]//Proceedings of European Conference on Computer Vision. Cham: Springer, 2014: 1-16.
- [31] ZHANG L, XIANG T, GONG S. Learning a discriminative null space for person re-identification[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2016: 1239-1248.
- [32] SU C, YANG F, ZHANG S, et al. Multi-task learning with low rank attribute embedding for multi-camera person re-identification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(5): 1167-1181.
- [33] MATSUKAWA T, OKABE T, SUZUKI E, et al. Hierarchical gaussian descriptor for person re-identification[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. [S.l.]: IEEE, 2016: 1363-1372.
- [34] AN L, CHEN X, YANG S, et al. Person re-identification by multi-hypergraph fusion[J]. IEEE Transactions on Neural Networks and Learning Systems, 2016, 28(11): 2763-2774.
- [35] SUN C, WANG D, LU H. Person re-identification via distance metric learning with latent variables[J]. IEEE Transactions on Image Processing, 2016, 26(1): 23-34.

- [36] YANG X, WANG M, HONG R, et al. Enhancing person re-identification in a self-trained subspace[J]. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2017, 13(3): 1-23.
- [37] CHEN Y C, ZHU X, ZHENG W S, et al. Person re-identification by camera correlation aware feature augmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(2): 392-408.
- [38] TAO D, GUO Y, YU B, et al. Deep multi-view feature learning for person re-identification[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017, 28(10): 2657-2666.
- [39] DAI J, ZHANG Y, LU H, et al. Cross-view semantic projection learning for person re-identification[J]. Pattern Recognition, 2018, 75: 63-76.
- [40] LEI J, NIU L, FU H, et al. Person re-identification by semantic region representation and topology constraint[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 29(8): 2453-2466.
- [41] LI H, XU J, ZHU J, et al. Top distance regularized projection and dictionary learning for person re-identification[J]. Information Sciences, 2019, 502: 472-491.

作者简介:



郝玲(1996-),女,硕士研究生,研究方向:数字图像处理和模式识别,E-mail:812592939@qq.com。



段继忠(1984-),通信作者,男,博士,副教授,研究方向:图像处理、深度学习和基于GPU的并行计算等,E-mail:duanjz@kust.edu.cn。



庞健(1996-),男,硕士研究生,研究方向:计算机视觉和模式识别等。

(编辑:夏道家)