

文章编号: 2095-2163(2021)04-0065-07

中图分类号: TP391.1

文献标志码: A

融合损失优化的行人重识别方法

张仕远, 丁学明

(上海理工大学 光电信息与计算机工程学院, 上海 200093)

摘要: 行人重识别的目的是通过深度学习检索出数据集中存在特定身份的行人, 在行人重识别算法中最理想的优化目标是最小化类内距离, 最大化类间距离。本文提出一种对损失函数进行优化改进的行人重识别方法。通过在分类损失上添加了额外的监督信号, 使其配合联合训练。在保持类间判别力的基础上, 使训练出的特征有较强的内聚性, 约束了类内紧凑性, 有效提升特征的辨识度。同时在交叉熵损失上增加类内与类间相似度的加权因子, 使学习速率适应具体的优化状态。从而使其优化方式更加灵活, 收敛状态更加明确来提高特征鉴别力。通过在行人重识别的主流数据集 Market-1501 与 DukeMTMC-reID 上的实验结果表明本文方法在性能上取得有效的提升, mAP 值分别达到了 90.5% 和 81.6%, Rank-1 值分别达到了 95.7% 和 91.6%。

关键词: 行人重识别; 深度学习; 度量学习; 残差网络; 损失函数

Person Re-identification based on loss optimization

ZHANG Shiyuan, DING Xueming

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

[Abstract] The purpose of Person Re-identification is to retrieve the pedestrians with specific identities in the data sets through deep learning. In the Person Re-identification algorithm, the ideal optimization goal is to minimize the distance between classes and maximize the distance between classes. This paper presents a pedestrian recognition method which optimizes and improves the loss function. By adding additional monitoring signals to the classification of losses, it can be combined with joint training. On the basis of maintaining the discriminability between the classes, the trained features have strong cohesion, which restricts the compactness within the classes and effectively improves the identification degree of the features. At the same time, the weight factor of similarity between and within classes was added to the cross entropy loss, so that the learning rate could adapt to the specific optimal state. In this way, the optimization mode is more flexible and the convergence state is more explicit to improve the feature discrimination. The experimental results on Market-1501 and DukeMTMC-reID, the mainstream data sets of Person Re-identification, show that the performance of the proposed method is effectively improved. The mAP value reaches 90.5% and 81.6%, respectively, and the Rank-1 value reaches 95.7% and 91.6%, respectively.

[Key words] Person re-identification; Deep learning; Metric learning; Residual network; Loss function

0 引言

近年来计算机视觉领域中的行人重识别(Person Re-identification)的关注度逐渐升高。其作为图像检索问题的分支, 目的是通过深度学习, 检索出数据集中存在特定身份的行人^[1]。随着人工智能科技的飞速发展, 该技术在社区安防、刑事追踪、智能机器人等领域都得到初步的应用。但由于图像中行人的动作和光照强度都会存在的差异, 且不同的图像之间存在遮挡物、分辨率低等问题, 会造成的类内与类间的差异变化, 故行人重识别到目前仍然

存在一些具有挑战性的难点。

随着深度学习的飞速发展以及高性能的计算硬件的不断升级, 深度学习已成为行人重识别任务的主流方法。其通过组合不同网络结构, 形成多层卷积神经网络的非线性深度学习模型; 对数据集中训练集进行训练; 再通过计算欧式距离等方法来比较样本间的相似度; 进而获得较高的性能指标。当前, 基于深度卷积神经网络的行人重识别, 普遍使用的骨干网络有 ResNet^[2]、VGGNet^[3]等。因其在识别任务中能自动提取具有较强鲁棒性行人图像的特征, 从而完成端到端的学习。随着行人重识别的关

基金项目: 国家自然科学基金(61673277)。

作者简介: 张仕远(1996-), 男, 硕士研究生, 主要研究方向: 深度学习和行人重识别; 丁学明(1971-), 男, 博士, 副教授, 主要研究方向: 智能控制、系统辨识以及嵌入式系统。

通讯作者: 丁学明 Email: xuemingding@usst.edu.cn

收稿日期: 2020-12-17

注度逐年递增,基于深度卷积神经网络的许多具有优异性能的行人重识别算法被不断提出。如:文献[4]中提出的三元组损失函数,其优化目标是使与目标样本行人属于不同类行人之间的距离,要比属于同类行人之间的距离大,从而增强模型的辨别能力。此外,还有四元组损失^[5]、柱状图损失^[6]和群组相似性学习^[7]等其它类型的度量学习损失方法。文献[8]中提出了批归一化机制,加快了深度网络的训练,并让网络能承受增加的训练率,同时减少了训练步骤。

综上所述,基于深度学习的方法虽然在一定程度上提升网络性能,但大部分都是对网络最终输出的特征进行预测处理,而没有使类内距离与类间距离达到理想效果。针对这个问题,本文提出一种对损失函数进行优化改进的行人重识别方法。通过在原始交叉熵损失上增加相似度的权重因子,使学习速率适应具体的优化状态。从而使其优化方式更加灵活,收敛状

态更加明确,有助于使特征更具鉴别力。

为了达到缩小同类样本之间距离,本文添加了额外的监督机制,使辅助的损失函数配合 softmax 联合训练。在保持不同类间判别力的基础上,让训练出特征的内聚性增强,约束了类内紧凑性,有效提升特征的辨识度,使模型泛化性和辨别能力有效提升。本文采用 ResNet-50 作骨干网络,结合改进后的 softmax 与三元组损失联合训练,再经过改进后的交叉熵损失进行训练,并结合一些训练技巧对网络加以改进,提升了网络的性能。通过基于 Market-1501 与 DukeMTMC-reID 数据集上大量的实验,结果表明本文方法提升了行人重识别准确率。

1 融合损失优化的行人重识别网络

本文所提融合损失优化方法网络的整体框架如图1所示。网络结构主要由预处理模块、主干网络模块、聚合模块和损失优化模块组成。

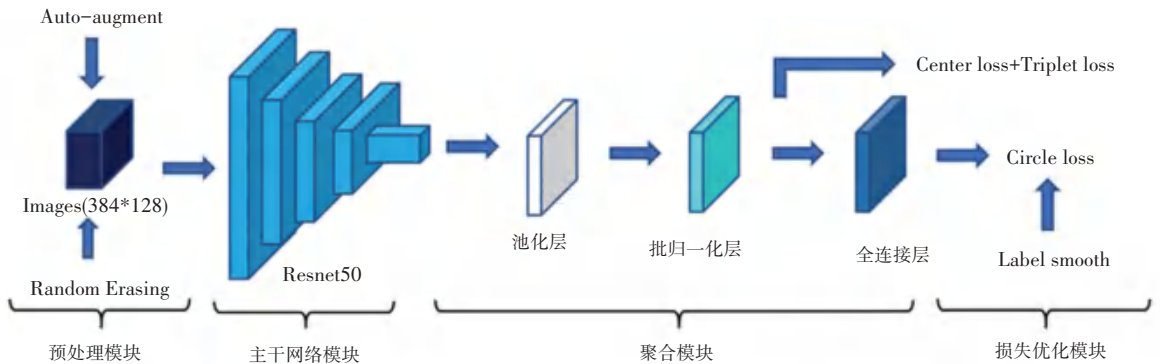


图1 损失优化模型的网络框架

Fig. 1 Network framework of loss optimization model

1.1 预处理模块

首先对数据集通过 Resize 将图像调整为固定大小。使输入行人图像的尺寸大小统一调整 256×128 像素,方便图像批量处理。本文设置不同身份行人的数量 $P = 16$, 同一身份行人的图片数 $K = 4$ 来训练样本,故 batch-size 大小为 64。对参与训练的图片使用随机擦除(Random Erasing, RE)^[9]。该方法是在输入图像中,将一个随机区域的像素更改为随机值的增强方法,从而避免模型出现过拟合的现象,使鲁棒性增强。图1中自动增强(Auto-augment)^[10]是一种提高图像分类准确性的算法,其运用自动搜索算法来寻找最佳的图像处理策略。如:将图像平移、旋转和剪切等操作,从而提高特征表示的鲁棒性。

1.2 主干网络

本文基于 ResNet-50 为骨干网络来提取行人图片的特征,其具有引入残差块和网络层数较深的特

点。其中 ResNet50 可分为六部分,Conv1 为卷积层模块,由一个 7×7 的卷积核提取特征,其步长设置为 2,故图像的长宽均减小为原图的一半。Conv2~Conv5 模块均采用重复的残差块提取特征,每个残差模块包含多层卷积层、下采样层和激活函数。在下采样层中,对特征图进行下采样的步长为 2,使得长宽再度缩减,进一步降低图像的分辨率。为获取更多的特征信息,本文将 ResNet50 结尾空间的下采样操作步长调整为 1,以增大网络输出特征图的尺寸,依次处理虽然会增加一些计算量,但却能在不增加额外训练参数的同时,对特征图分辨率带来较大的提升。

1.3 聚合模块

在骨干网络阶段完成后,对输出特征图进行池化操作。采用全局平均池化(Avg pooling),将骨干网生成的特征图聚合成一个全局特征,得到的特征

向量为 2048 维。再将其经过一个批归一化层 (Batch Normalization, BN), 用于在网络训练过程中, 获得更加平滑的优化空间, 并对学习率等超参具有更强的鲁棒性, 同时避免激活函数靠近其饱和区。其后经全连接层 (Fully Connected, FC), 最终输出全局特征向量。

1.4 损失优化模块

在行人重识别任务中, 使用类标签进行学习时, 普遍采用结合 softmax 与交叉熵损失函数的方法, 来学习同类别之间的信息。从而优化样本和权重向量之间的相似度, 提高正确预测的准确度。softmax 是行人重识别中常用的分类损失函数, 其公式为:

$$L_S = - \sum_{i=1}^m \log \frac{e^{W_j^T x_i + b_{y_i}}}{\sum_{j=1}^n e^{W_j^T x_i + b_j}} \quad (1)$$

其中: x_i 代表 d 维空间中第 i 个深度特征; y 表示其类别; W_j 为全连接层的参数矩阵 W 中的第 j 列; n 为类别的数量; b 为偏置; mini-batch 的大小为 m 。

在 softmax 的监督学习中, 得到的特征仅限于区别出各类别的差异, 无法量化的控制各类别间与同类别内的距离, 且同类样本间特征差异无法约束。本文提出采样中心损失函数^[11]作为辅助损失函数, 用以监督 softmax 在训练时忽略的同类标签样本间的距离, 使训练出的特征具有向一个特征中心聚合的特性。中心损失函数公式为:

$$L_C = \frac{1}{2} \sum_{j=1}^m \|x_i - c_{y_j}\|_2^2 \quad (2)$$

在 mini-batch 中, 第 j 张图片的标签表示为 y_j , c_{y_j} 为深度特征的第 y_j 个类中心。其计算的是各样本的特征与该类别的中心特征的欧式距离之和, 并使其最小化, 即最小化类内距离, 其大小体现的是类内的变化。中心损失原理是根据对某个类别特征中心的学习来约束不同样本特征与类别中心特征的偏移, 形成同一种类样本围绕在样本中心的状态, 则可以增加类内紧凑性, 即使存在某些类内变化大的样本, 也不会影响识别结果的鲁棒性。在原有的 softmax 上增加辅助函数的改进, 更加有利于网络的训练。总体公式为:

$$L = L_S + \lambda L_C = - \sum_{i=1}^m \log \frac{e^{W_j^T x_i + b_{y_i}}}{\sum_{j=1}^n e^{W_j^T x_i + b_j}} + \frac{\lambda}{2} \sum_{i=1}^m \|x_i - c_{y_i}\|_2^2 \quad (3)$$

在结合使用时, softmax 负责区别类间距离, 中心损失约束类内的距离, 使学习到的特征判别度更

高; 再与三元组损失联合训练, 使类间距离远大于类内距离, 即目标样本正距离尽可能小于目标样本负距离。

交叉熵损失 (Cross-Entropy Loss) 是行人重识别中常用的损失函数, 其目标是增加类内相似度 s_p 和缩减类间相似度 s_n , 从而优化样本和权重向量之间的相似度。具体表示为先将 s_n 和 s_p 组合成相似度对, 为使 $(s_n - s_p)$ 的值减小, 则要减小 s_n 或增大 s_p 。然而, 由于二者系数均为 1, 故其优化的梯度幅度相同, 对类内和类间的相似度惩罚程度也是一致的, 不能以不同的速率学习优化。另外这种方式使用的决策边界为 $s_n - s_p = m$, 该边界在 (s_n, s_p) 空间中与 $s_n = s_p$ 平行, 导致通向边界上任意点对应的难度相同, 出现方向不唯一的收敛状态, 导致收敛状态不明确。公式表示为:

$$L = \log \left[1 + \sum_{i=1}^K \sum_{j=1}^L \exp(\gamma(s_n^i - s_p^i + m)) \right] \quad (4)$$

其中, K 为与某个样本比较, 所有类内相似度分数的数量; L 则表示与某个样本比较, 所有类间相似度分数的数量; m 为余量; γ 为松弛变量。本文提出的采用圆损失函数^[12]的方法进行训练, 引入加权因子 α_n 和 α_p , 从而允许 s_n 和 s_p 的优化不完全同步, 其权重的大小与离最优点的距离大小成正相关。例如, 当相似性分数与优化目标偏离越远, 匹配的加权因子则越大, 从而更新梯度越大, 使学习速率适应具体的优化状态。公式中表示为把 $(s_n - s_p)$ 泛化为 $(\alpha_n s_n - \alpha_p s_p)$, 使 α_n 和 α_p 与 s_n 和 s_p 分别为线性关系, 该方式所得决策边界为 $\alpha_n s_n - \alpha_p s_p = m$, 且该分界面呈现的是圆弧形。在该决策边界上损失函数有明确的收敛状态, 使特征更具鉴别力。

$$L_{circle} = \log \left[1 + \sum_{i=1}^K \sum_{j=1}^L \exp(\gamma(\alpha_n^i s_n^i - \alpha_p^i s_p^i + m)) \right] \quad (5)$$

综上, 损失函数部分先由中心损失与三元组损失联合训练, 经一个全连接层降维后, 计算圆损失, 以得到最终损失。其中, 中心损失使网络学习出来一个类中心, 优化类内距离。三元组损失负责使类间距离大于类内距离。圆损失使优化方式更加灵活, 优化目的更明确。三种损失联合对网络进行更好的优化, 保证模型学习到具有辨别性的特征。总体损失函数为:

$$L = L_{Triplet} + L_{circle} + \beta L_{Center} \quad (6)$$

其中, β 是中心损失的权重系数, 本文实验默认为 0.005。由于在分类任务中, 训练时常有过拟合的

风险,因此本文对身份标签采取了标签平滑(Label smoothing, LS)^[13]处理,用来更好的提升泛化能力。

2 实验结果与分析

2.1 实验环境与参数设置

本文方法在基于 Pytorch 深度学习框架工具包上进行网络搭建,使用 4 个 NVIDIA GeForce RTX 2080 GPU 进行数据并行加速。实验平台基于 64 位的 Ubuntu16.04 操作系统;基于 python3.6.4 的深度学习框架 Pytorch1.1.0 完成程序编程;骨干网络选用 ResNet-50;输入行人图像的像素均为 256×128 ,并采用标签平滑训练模型,平滑因子 $\varepsilon = 0.1$ 。

为了提高训练效果,本文采用对训练学习率进行预热处理的方法,即把初始学习率设为 3.5×10^{-5} ,随后不断增加,直到 10 个 epoch 时增大到 3.5×10^{-4} ,并在迭代次数为 40 时减小为 3.5×10^{-5} ,再在迭代次数为 70 时减小为 3.5×10^{-6} ,其它时间保持不变。每次训练的迭代次数均设为 120,批量大小为 64。训练时优化器为 ADAM,进行梯度参数更新,其中两个超参数 β_1, β_2 均为 0.9。

2.2 数据集与评价指标

2.2.1 数据集

为验证本文模型的有效性,分别在 Market1501 与 DukeMTMC-reID 两个主流公共数据集上完成了相关消融实验和对比实验,并与一些近期提出的行人重识别算法进行了结果对比。数据集详尽信息见表 1。

表 1 数据集信息

Tab. 1 Information of datasets

数据集	训练集/张	训练集 人数/个	测试集/张	测试集 人数/个
Market-1501	12 936	751	19 732	750
DukeMTMC-reID	16 522	702	17 661	702

Market-1501 数据集是由 6 个不同的室外摄像头拍摄而来,摄像头为 5 个高分辨率,1 个低分辨率。数据集总共包含有 1 501 个行人的 32 217 张图像,使用 DPM 检测器来进行行人检测矩形框的标注。

DukeMTMC-reID 数据集是在杜克大学校园的冬季,由 8 个不同的室外高分辨率摄像机拍摄而来。数据集总共包含有 1 812 个行人的 36 441 张图像。其中行人图像存在不少遮挡的情况,具有较大的识别难度。

2.2.2 评估指标

实验中采用行人重识别中的首位匹配率 Rank-1 和平均精度均值(mAP)作为评价指标。Rank-k 为根据相似度由高到低排列后的前 k 个行人里与查询样本身份相同的准确率;Rank-1 即相似度最高的检索结果为正确匹配的概率,最能反映模型的识别能力。mAP 即平均精度均值,是指将分类问题中所有类别的平均精度(AP)相加再取平均值,可以较全面的衡量网络的识别能力和稳定性。计算步骤如下:

(1) 求准确率 P :

$$P = \frac{TP}{TP + FP}, \quad (7)$$

其中, TP 表示被正确预测的行人图片个数, FP 为被错误预测的行人图片个数。

(2) 求平均精度 AP(某类别行人检测结果的全部准确率的平均值):

$$AP = \frac{\sum P_i}{I_c}, \quad (8)$$

其中, i 表示检测出行人图片的序号,分母表示含有第 C 个类别行人图片的数量。

(3) 求任务中各个类别平均精度的均值(即为平均精度均值 mAP):

$$mAP = \frac{\sum_{k=0}^C AP_k}{C}, \quad (9)$$

其中, k 表示类之间平均精度的序号, C 为总类别数。

2.3 实验结果分析

2.3.1 消融实验

为了充分体现本文采用的训练技巧对模型骨干网络的提升效果,分别在两个数据集上完成了神经网络研究中常见的消融实验。即通过叠加部分网络结构,分别检测叠加部分对模型性能的影响。其中骨干网络为 ResNet50,通过在骨干网络上依次叠加特定训练技巧来进行各结构对性能影响的验证。

表 2 中,RE 为随机擦除,Auto 为自动增强,LS 为标签平滑,Stride=1 为下采样步长设置为 1,BN 为批归一化处理。由表中实验数据可知,不同训练技巧的添加对原始骨干网络的准确率均有不同程度的提高。最终形成的 baseline 较 ResNet50 网络,在 Market-1501 数据集上,Rank-1 和 mAP 分别提升了 6.7% 和 6.6%,达到了 90.1% 和 76.2%。在 DukeMTMC-ReID 数据集上,Rank-1 和 mAP 分别提升了 6.8% 和

12.8%,达到了85.1%和74.0%。

表2 消融实验

Tab. 2 Ablation Experiment

Model	Market1501		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP
Backbone	83.4	69.6	78.3	61.2
+RE	86.7	71.9	79.2	64.9
+Auto	87.3	72.3	79.7	66.2
+LS	87.5	72.6	80.6	67.5
+Stride=1	88.3	73.5	81.3	68.8
+BN	90.1	76.2	85.1	74.0

2.3.2 对比实验

为验证本文提出的融合损失优化方法对模型改进的效果,基于本文基准网络进行了不同损失优化的对比实验。

表3中, L_{id} 为原始的身份分类损失,作为基准网络(baseline),其表示只训练身份分类损失函数(softmax+cross entropy)的网络; $L_{id} + L_{center}$ 为在softmax上添加中心损失后的网络; $L_{id} + L_{triplet}$ 为联合分类损失与三元组损失联合训练的网络; $L_{id} + L_{triplet} + L_{center}$ 则表示基准网络加入了中心损失与三元组损失的训练。由表3数据可知,通过对原始基准网络的损失优化,模型性能不断的提升。

表3 与基准网络的对比实验

Tab. 3 Experiments compared with baseline

损失函数	Market1501		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP
L_{id}	90.1	76.2	85.1	74.0
$L_{id} + L_{center}$	90.5	79.5	85.8	75.2
$L_{id} + L_{triplet}$	91.3	81.2	86.3	76.1
$L_{id} + L_{triplet} + L_{center}$	93.3	84.0	88.3	77.9

表4中, L_{circle} 为在基准网络中的交叉熵损失里增加类内相似和类间相似的加权因子后的网络。作为初步优化后的baseline, $L_{circle} + L_{center}$ 和 $L_{circle} + L_{triplet}$ 分别为在此基础上增加中心损失和三元组损失; $L_{circle} + L_{center} + L_{triplet}$ 为本文所提出的网络结构,即融合加权因子与中心损失再与三元组损失联合训练的网络。由表4数据可知,在添加损失优化函数之后,模型基准网络性能不断得到提升。模型在Market-1501数据集上,Rank-1和mAP值分别为95.2%和86.1%;在DukeMTMC-ReID数据集上,Rank-1和mAP值分别为89.6%和79.8%。由此证明,加入本文的损失优化模块,能有效的让网络具有更强判别力与泛化能力,相较原始基准网络在准确率上有明

显的提高。

表4 与优化后的基准网络的对比实验

Tab. 4 Experiments comparison with optimized baseline

损失函数	Market1501		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP
L_{circle}	90.8	77.1	86.0	75.1
$L_{circle} + L_{center}$	92.2	81.7	87.4	76.5
$L_{circle} + L_{triplet}$	92.6	82.7	87.6	76.6
$L_{circle} + L_{triplet} + L_{center}$	95.2	86.1	89.6	79.8

2.3.3 与其它主流方法比较

本文提出网络与一些主流网络性能的比较结果见表5。由表中数据显示,本文所提网络性能指标在两个不同数据集上得到的Rank-1和mAP,比这些主流网络性能上存在不同程度的优势。

表5 与其它行人重识别方法比较

Tab. 5 Comparison with other Person Re-identification methods

方法	Market1501		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP
SVDNet ^[14]	82.3	62.1	76.7	56.8
TriNet ^[4]	84.9	69.1	-	-
PCB ^[15]	92.3	77.4	81.8	66.1
PCB+RPP ^[15]	93.8	81.6	83.3	69.2
AANet ^[16]	93.9	83.4	87.7	74.3
IANet ^[17]	94.4	83.1	87.1	73.4
DGNet ^[18]	94.8	86.0	86.6	74.8
OSNet ^[19]	94.8	84.9	88.6	73.5
MHN ^[20]	95.1	85.0	89.1	77.2
Ours	95.2	86.1	89.6	79.8
Ours+Re-rank	95.7	90.5	91.6	81.6

注:表中“-”意为该网络在数据集上无实验数据。

由于Re-rank是行人重识别网络提高精度的重要步骤,其原理是通过计算马氏距离和杰卡德距离,对网络训练结果重新排序的过程。因此,本文采用Re-Rank^[21]对网络结果进一步提升。在Re-rank优化后,Market-1501数据集上,Rank-1和mAP值分别达到95.7%和90.5%;在DukeMTMC-ReID数据集上,Rank-1和mAP值分别为91.6%和81.6%。

2.3.4 附加实验

为检测行人重识别中不同训练批次大小对结果的影响,本文采用不同 $P \times K$ 值对网络进行训练,见表6。其中, P 为行人身份的数量, K 为同一身份行人的图片数。

由图2中a,b曲线可以看出,模型在 $P \times K$ 值设置为 4×16 时,达到最佳性能。通过对比可初步

推断,模型的性能随 batch-size 的增大而提升,而当到达一定大小则结果趋于稳定且不再增加。

表 6 模型设置不同 batch sizes 数量大小的实验

Tab. 6 The experiment of setting different batch size in the model

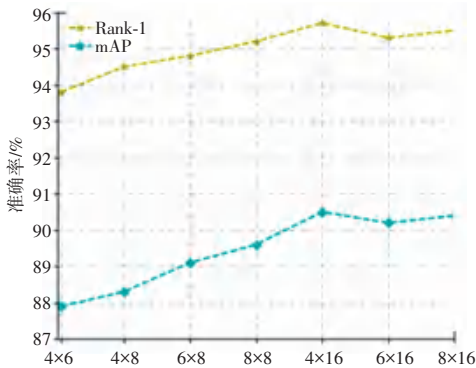
Batch Size $P \times K$	Market1501		DukeMTMC-reID	
	Rank-1	mAP	Rank-1	mAP
4×6	93.8	87.9	89.7	77.3
4×8	94.5	88.3	90.4	78.8
6×8	94.8	89.1	90.7	79.6
8×8	95.2	89.6	91.0	80.7
4×16	95.7	90.5	91.6	81.6
6×16	95.3	90.2	90.3	80.9
8×16	95.5	90.4	91.5	81.2

2.4 可视化分析

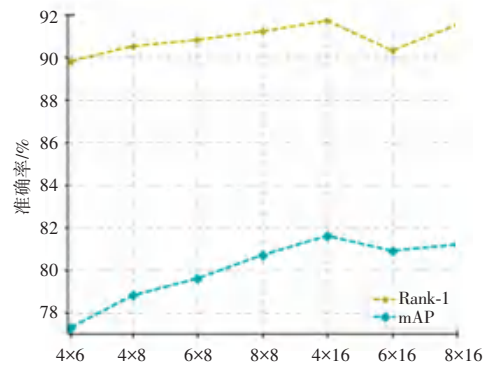
由于可视化结果可以更直观的验证模型对识别

任务的改进效果,本文基于 Market-1501 数据集的可视化结果比较如图 3 所示。图 3(a)和图 3(c)为基于基准网络的重识别结果。即只结合本文所提训练技巧,并未融合损失优化的 ResNet50 网络;图 3(b)和图 3(d)表示融合损失优化后网络的结果。

图中每行第一张图像为待查询的目标行人,之后 10 张图像表示为按准确率逐渐减小排列的前 10 张查询结果,且红色线框图像为正确查询的结果,蓝色线框图像为错误查询的结果。在图 3(a)、图 3(b)的对比中,基准网络的前 6 张为正确,后 4 张为错误;本文网络则前 9 张为正确,只有最后一张为错误。图 3(c)、图 3(d)的对比中,基准网络同样是前 6 张为正确,后 4 张为错误,本文网络为前 8 张为正确,后两张为错误。由此证明,本文网络有效的提升了识别的准确度。



(a) Market-1501 数据集上的对比



(b) DukeMTMC-reID 数据集上的对比

图 2 不同 batch sizes 在数据集上的性能对比

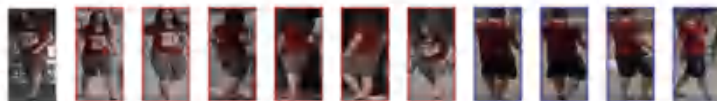
Fig. 2 Performance comparison of different batch sizes on datasets



(a) 基于基准网络的重识别结果



(b) 融合损失优化网络的重识别结果 1



(c) 基于基准网络的重识别结果 2



(d) 融合损失优化网络的重识别结果 2

图 3 Market-1501 数据集可视化结果对比

Fig. 3 Visual results comparison of Market-1501 dataset

3 结束语

针对改善行人重识别中类内差异大、类间差异小的问题,本文提出了融合损失优化的一种行人重识别方法。通过对交叉熵损失函数进行添加类内相似度和类间相似度的加权因子,使优化方式更加灵活,优化目的更明确。并提出添加中心损失来配合 softmax 训练,约束了类内紧凑性。并与三元组损失联合训练,一起作用于识别任务中。通过本文方法的改进,使得网络对行人的判别能力提升,提高了行人重识别的精度。并在主流公开数据集上对模型进行了验证。实验表明,改进后的网络模型在具有更好的鲁棒性和泛化能力的同时,对识别任务的准确率取得了有效提高。

参考文献

- [1] [1] WANRU S, QINGQING Z, CHANGHONG C, et al. Survey on pedestrian re-identification research[J]. CAAI Transactions on Intelligent Systems, 2017 (6): 770-780.
- [2] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, arXiv:1512.03385, 2016.
- [3] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition [J]. Computer Science, arXiv Preprint arXiv:1409.1556, 2014.
- [4] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: A unified embedding for face recognition and clustering [C] // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 7-12, Boston, USA. New York: IEEE, 2015: 815-823.
- [5] CHEN W, CHEN X, ZHANG J, et al. Beyond Triplet Loss: A Deep Quadruplet Network for Person Re-identification [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), arXiv Preprint arXiv:1704.01719, 2017.
- [6] Ustinova E, Lempitsky V. Learning deep embeddings with histogram loss [C] // Proc of the 2016 Annual Conference on Neural Information Processing Systems (NIPS), 2016: 4177-4185.
- [7] CHEN D, XU D, LI H, et al. Group Consistent Similarity Learning via Deep CRF for Person Re-identification [C] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA. 2018: 8649-8658.
- [8] IOFFE S, SZEGEDY C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift [J]. arXiv Preprint arXiv:1502.03167, 2015.
- [9] ZHONG Z, ZHENG L, KANG G, et al. Random Erasing Data Augmentation [J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2017, 34(7).
- [10] CUBUK E D, ZOPH B, DANDELION MANÉ, et al. AutoAugment: Learning Augmentation Strategies From Data [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019: 113-123.
- [11] WEN Y, ZHANG K, LI Z, et al. A Discriminative Feature Learning Approach for Deep Face Recognition [M] // European Conference on Computer Vision (ECCV). Springer, Cham, 2016, 10.1007.
- [12] SUN Y. Circle Loss: A Unified Perspective of Pair Similarity Optimization [C] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 2020: 6397-6406.
- [13] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision [C] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 2818-2826.
- [14] SUN Y, ZHENG L, DENG W, et al. SVDNet for Pedestrian Retrieval [C] // 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017: 3820-3828.
- [15] SUN Y, ZHENG L, YANG Y, et al. Beyond Part Models: Person Retrieval with Refined Part Pooling (and A Strong Convolutional Baseline) [J]. Springer, Cham, arXiv Preprint arXiv:1711.09349, 2017.
- [16] TAY C P, ROY S, YAP K H. AANet: Attribute Attention Network for Person Re-Identifications [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019: 7127-7136.
- [17] HOU R, MA B, CHANG H, et al. Interaction-And-Aggregation Network for Person Re-Identification [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, arXiv Preprint arXiv:1907.08435, 2019.
- [18] ZHENG Z, YANG X, YU Z, et al. Joint Discriminative and Generative Learning for Person Re-identification [C] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, 2019: 2133-2142.
- [19] ZHOU K, YANG Y, CAVALLARO A, et al. Omni-Scale Feature Learning for Person Re-Identification [C] // 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea (South), 2019: 3701-3711.
- [20] CHEN B, DENG W, HU J. Mixed High-Order Attention Network for Person Re-Identification [C] // International Conference on Computer Vision (ICCV). IEEE, arXiv Preprint arXiv:1908.05819, 2019.
- [21] ZHONG Z, ZHENG L, CAO Donglin et al. Re-ranking Person Re-identification with k-Reciprocal Encoding [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 3652-3661.