



计算机工程与应用
Computer Engineering and Applications
ISSN 1002-8331, CN 11-2127/TP

《计算机工程与应用》网络首发论文

题目: 多分支融合局部特征的行人重识别算法
作者: 肖雅妮, 范馨月, 陈文峰
网络首发日期: 2020-10-21
引用格式: 肖雅妮, 范馨月, 陈文峰. 多分支融合局部特征的行人重识别算法. 计算机工程与应用. <https://kns.cnki.net/kcms/detail/11.2127.TP.20201020.1654.008.html>



网络首发: 在编辑部工作流程中, 稿件从录用到出版要经历录用定稿、排版定稿、整期汇编定稿等阶段。录用定稿指内容已经确定, 且通过同行评议、主编终审同意刊用的稿件。排版定稿指录用定稿按照期刊特定版式 (包括网络呈现版式) 排版后的稿件, 可暂不确定出版年、卷、期和页码。整期汇编定稿指出版年、卷、期、页码均已确定的印刷或数字出版的整期汇编稿件。录用定稿网络首发稿件内容必须符合《出版管理条例》和《期刊出版管理规定》的有关规定; 学术研究成果具有创新性、科学性和先进性, 符合编辑部对刊文的录用要求, 不存在学术不端行为及其他侵权行为; 稿件内容应基本符合国家有关书刊编辑、出版的技术标准, 正确使用和统一规范语言文字、符号、数字、外文字母、法定计量单位及地图标注等。为确保录用定稿网络首发的严肃性, 录用定稿一经发布, 不得修改论文题目、作者、机构名称和学术内容, 只可基于编辑规范进行少量文字的修改。

出版确认: 纸质期刊编辑部通过与《中国学术期刊 (光盘版)》电子杂志社有限公司签约, 在《中国学术期刊 (网络版)》出版传播平台上创办与纸质期刊内容一致的网络版, 以单篇或整期出版形式, 在印刷出版之前刊发论文的录用定稿、排版定稿、整期汇编定稿。因为《中国学术期刊 (网络版)》是国家新闻出版广电总局批准的网络连续型出版物 (ISSN 2096-4188, CN 11-6037/Z), 所以签约期刊的网络版上网络首发论文视为正式出版。

多分支融合局部特征的行人重识别算法

肖雅妮^{1,2}, 范馨月^{1,2}, 陈文峰^{1,2}

1.重庆邮电大学 通信与信息工程学院, 重庆 400065

2.光通信与网络重点实验室(重庆邮电大学), 重庆 400065

摘要: 大部分结合深度学习的行人重识别算法主要以单分支的网络结构为主, 且大多利用图片的全局特征信息, 这样易错失关键行人信息, 导致度量学习效果、算法精度下降。因此, 为使网络获取到更多的关键行人信息, 减少对行人局部、细节信息的错失, 加强网络对行人特征的学习。基于 ResNet-50 的骨干网络, 采取多分支的网络结构设计, 综合考虑训练难易、运算量, 选择融合三个独立分支的结构设计: 随机擦除分支、全局学习分支、局部学习分支, 并在此基础上根据实验数据进行调整优化, 最后再结合最小二乘法分配损失函数权重使模型更具鲁棒性, 实验结果表明, 三个分支具有互补性, 使用融合分支特征做算法测试时, 相比基础的单分支、多分支网络, 该算法使得行人重识别精度提升。

关键词: 行人重识别; 多分支网络; 随机擦除; 局部特征; 深度学习

文献标志码:A **中图分类号:**TP391 **doi:** 10.3778/j.issn.1002-8331.2006-0226

肖雅妮, 范馨月, 陈文峰. 多分支融合局部特征的行人重识别算法. 计算机工程与应用

XIAO Yani, FAN Xinyue, CHEN Wenfeng. Research on Person Re-identification based on integrating local features under multi branches. Computer Engineering and Applications

Research on Person Re-identification based on integrating local features under multi branches

XIAO Yani^{1,2}, FAN Xinyue^{1,2}, CHEN Wenfeng^{1,2}

1. School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

2. Key Laboratory of Optical Communication and Networks (Chongqing University of Posts and Telecommunications), Chongqing 400065, China

Abstract: Most of the Person Re-identification algorithms combined with deep learning are mainly based on single branch network structure, and most of them use the global feature information of the picture, so it is easy to miss the key pedestrian information, which leads to the decline of metric learning effect and algorithm accuracy. Therefore, in order to make the network obtain more key pedestrian information and reduce the loss of obtaining pedestrian and detail information. Strengthen the network learning of pedestrian characteristics. Based on the network structure of resnet-50, the multi branches network structure design is adopted. Considering the training difficulty and calculated amount, this paper chooses the structure which merging three independent branches that are consisted of random erasing branch, global learning branch and local learning branch. On this basis, the adjustment and optimization are carried out according to the experimental data. Finally, the weight of loss function is allocated by combining the least square method to make the model more accurate. The experimental results show that the three branches struc-

作者简介: 肖雅妮(1996-),女,硕士,研究领域为结合深度学习的行人重识别算法,E-mail: s180101120@stu.cqupt.edu.cn; 范馨月(1979-),女,副教授,硕士,主要研究方向为认知无线电,信号处理,图像处理; 陈文峰(1995-),男,硕士,研究领域为结合深度学习的行人重识别算法。

ture are complementary. Compared with the basic single branch and multi branch network, the algorithm improves the Person Re-identification algorithms accuracy.

Key words: pedestrian re-identification; multi branch network; random erasing; local feature; deep learning

1 引言

在当今社会, 各地政府在公共场所关键点、交通路口、生活小区、停车场等区域安装大量摄像头, 因此获得了海量的行人数据, 而针对如何在复杂场景下获取人脸、步态等信息, 并进行行人检索的问题, 就需要行人重识别技术。行人重识别是指在已知来源与非重叠摄像机拍摄范围的视频序列中识别出特定行人的技术, 即指给定某摄像头拍摄到的某行人图片, 在资料库中检索该行人被其他摄像头拍到的图片^[1]。由于摄像头位置、角度和参数设置不同, 行人姿态频繁变化, 以及背景干扰、遮挡和成像质量不稳定等原因, 同一行人在不同时刻不同摄像头拍摄的图片中存在较大差异^[2], 这使得行人重识别的研究具有很大的挑战性。而深层网络可以提取到的行人细节相比传统算法效果更显著, 因此目前的行人重识别研究主要以深度学习的方向为主。

由于人体结构具有强分辨^[3], 所以大部分算法的研究重心放在如何使模型获得更多的行人表征特征上。目前大部分算法是从整体图片获取到行人的特征, 具有代表性的如论文^[4]提出的是行人重识别最基础的网络框架, 相比传统算法效果得到了一定提升但由于特征学习部分未引入增强机制, 错失对局部的学习、算法效果不显著。论文^[3]在行人身份 ID 的基础上, 额外增加行人属性, 算法效果提升一些的同时增加了复杂度, 由于忽略了部分不显著或不经常出现的详细信息, 相比论文^[4]算法效果增益小, 且仅使用全局特征在算法的测试部分不利于分辨不同行人和同一个行人不同摄像头的照片。为了进一步提高算法效果, 行人重识别的研究针对特征学习, 将重点从全局转移到

局部。因为加强网络对局部特征的学习, 能更多地获取行人细节信息。如论文^[5]通过对定位人体骨骼来学习身体区域的特征, 相比仅使用全局特征的方法提到了提升, 但由于对骨骼定点大大增加网络复杂度, 同时关注行人姿势大于对局部的学习, 因此最后的效果增益较小。论文^[6]先将行人分开再对齐学习特征、最后测试部分使用融合特征, 虽然该方法有对局部特征的学习, 但缺乏对局部注意力特征的学习和多分支网络更好的融合设计。而论文^[7]是加强网络对于人体姿势的学习, 因此引入行人姿势估计的算法, 需要提前预训练, 这大大提升了算法难度。综上所述, 为强化网络学习行人特征信息和保证网络的复杂度、训练难度, 本文设计了一种端到端的多分支网络机构, 一共三个分支: 随机擦除的分支, 全局学习分支, 局部学习分支。

在算法结构中全局学习分支的引入有两个目的, 一是联合训练确保算法精度, 二是获取人体的躯干信息。局部学习分支在本文的网络结构中是对被均分成三等分后的特征图部分分别进行学习, 其分支能获得行人局部细节信息, 随机擦除分支通过固定面积的掩膜随机地对特征图进行遮挡, 加强网络对未遮挡部分的注意力特征学习, 可看做是一种加强特征学习的策略, 因为遮挡使用的掩膜无网络参数, 所以不会增加网络规模。考虑到会有噪音干扰, 本文的三个分支在结构上仅共享骨干网络部分, 其余部分独立, 最后本文结合最小二乘法, 根据实验数据, 做损失函数的权重分配。最后本文使用三个分支得到的融合特征进行测试, 实验结果表明, 本文提出的方法在 Market-1501 数据集, DukeMTMC-reID 数据集,

CUHK03-Label、CUHK03-Detect 数据集上都取得了更好的准确率。

2 相关工作与多分支网络设计

虽然深层的网络仍可根据人体不同的身体部位语义信息来初步区分,但缺失关键和细微的行人特征信息和学习到额外的背景信息都会导致算法增益小。且论文^[8]和表 1 的结果数据显示,当网络结构分支增加、图片细分增多,将增强网络单个分支对图片的注意力,提高算法效果。因此本文在网络结构的设计上,选择的不是单分支的结构而是多分支的结构。而本文在分支的选择上考虑单个分支的算法效果与运算量以及训练难度,同时结合上述理论。

根据论文^[8,9],本文首先选择了三个分支:全局学习分支、局部学习分支、随机擦除分支。其中,全局学习分支强调对行人躯干和行人图片全局的学习,局部学习分支强调的是对行人局部以及图片局部的学习,而随机擦除分支的引入目的是在前两个特征学习分支上,增强网络对局部注意力特征和弱特征部分的学习,同时有助于缓解因行人姿势改变和重叠导致局部特征在训练卷积神经网络时被抑制的问题。全局学习分支的结构设计参照最基础的单分支网络,未引入任何加强特征学习的机制。而局部分支的设计,考虑网络随着特征分支数量的增加,提升对局部注意力的同时也会增加复杂度。经本文实验获知,单个分支相比,均分四分支的算法运算量比均分三等分的分支大,且尝试与其它分支组合的训练的过程中,模型收敛慢、训练不稳定,同时且算法效果也仅比三等分高零点几点精度。所以本文选择的是均分三等分的局部学习分支,在此基础上,为保证训练精度、缓解各个分支间训练时产生的干扰,在局部学习分支上又引入一个全局学习分支作为联合训练。第三个随机擦除分支参照局部学

习分支也引入一个全局学习分支作为联合训练,包含掩膜用于对未遮挡部分学习的分支将对网络输出的一个 batch 的特征图进行操作。

确定分支类型后,在选择分支数量和类型的问题上,本文先在数据集 DukeMTMC-reID 上做了网络结构的对比实验,一是仅有单分支的网络结构:随机擦除分支、局部学习分支、全局学习分支。二是两个分支的组合:全局学习分支和随机擦除分支,全局学习分支和局部学习分支,随机擦除分支和局部学习分支。三是三个分支:全局学习分支和随机擦除分支、局部学习分支,四是在三分支的基础上增加全局学习分支的四分支网络结构,表 1 中未包含随机擦除分支和局部分支分别加在三分支的实验数据,一是因为开始训练时服务器因为内存报错,二是降低了训练的 batchsize 和 epoch 后与其它实验数据对比误差大。

表 1 在数据集 DukeMTMC-reID 上未调权重的对比实验
Table 1 Comparative experiment of unadjusted weights on the dataset DukeMTMC-reID

对比网络结构	MAP/%	Rank-1/%
随机擦除分支	54.52	69.87
三等分局部学习分支	60.77	73.84
全局学习分支	52.61	69.57
全局学习分支+随机擦除分支	59.97	73.35
全局学习分支+局部学习分支	62.63	75.23
随机擦除分支+局部学习分支	62.75	75.28
全局学习分支+随机擦除分支+局部学习分支	73.13	84.64
2*全局学习分支+随机擦除分支+三等分局部学习分支	73.24	85.02

从表 1 获悉,两分支和三分支的网络结构算法结果相比单分支的网络结构,其精度提到了至少 10 点以上,并且两分支放在同一网络都会在原有基础上提升算法效果,这侧面反映了分支在网络结构中具有互补的作用,随机擦除分支与全局学习分支学习到全局信息,并增加对局部和弱

特征注意力。全局学习分支与局部学习分支的组合学习到了全局信息与局部信息。随机擦除分支与局部学习分支的组合在二分支的网络结构里效果最好,分析原因可能与随机擦除分支能缓解局部特征在训练时被抑制的问题,同时随机擦除分支也有助于网络学习弱特征和注意力的区域。因此两两分支放在同一个网络中,能互相增强彼此对特征的学习。虽然在三支基础上增加全局学习分支的四分支结构的算法结果最高,但训练过程中,由于分支过多,导致网络收敛慢,测试部分耗时,最后的结果也仅比三支的效果高0.11。综上所述,根据算法结果和训练难易,本文最后使用的是由全局学习分支、随机擦除分支、局部学习分支构成的三支网络而不是四分支的网络结构。三个分支仅共享骨干网络部分,在训练过程中分别学习其他分支未注意到的特征部分,通过这样的网络设计,使得模型的注意力不仅仅关注行人的躯干部分,而是融合三个分支的注意力,增强网络提取特征的能力。在单个分

支的对比中,局部学习分支获得的效果最好,这可能与网络越深越会注重区域的局部学习和监督学习本身会强迫网络不断地去获取具有强辨别力的特征有关。

3 多分支算法介绍

3.1 三个分支的介绍

由于 ResNet-50 在其他行人重识别算法中取得了良好的结果,所以在本文中采取的骨干网络是 ResNet-50,其框架相比于其他网络,具有网络层数深、获取到的特征信息更为丰富,引入残差块、避免梯度爆炸和梯度消失的优点。在本文图 1 所示的算法结构中,与 ResNet-50 有所区别的是后面的输出分支被分为三个分支。在图 1 中,方块表示的是特征图,为了方便显示采取的是三维的立方体,从 ResNet-50 stage1,2,3 开始分开的三支从上往下,分别是随机擦除分支 (RE branch),全局学习分支 (Global branch),局部学习分支 (Part branch),在每个分支上继续编号,从上往下分别是 R_0、R_1、G_0、P_0、P_1。

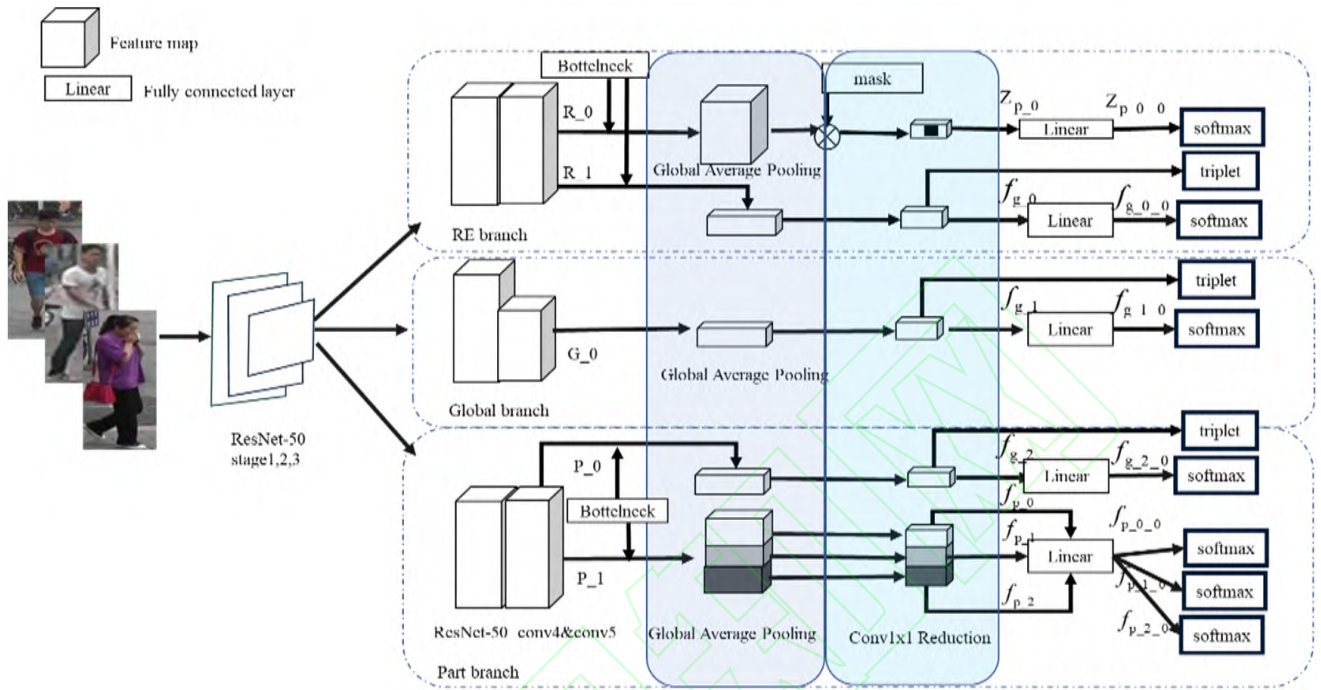


Fig.1 Algorithm network framework

图 1 算法网络框架

随机擦除分支 (RE branch) 包含 R_0 、 R_1 分支,在 R_0 分支上,数据通过 ResNet-50 stage1,2,3 后,进入到 resnet 的 conv4 和 conv5 后,为了保留足够的信息,没有在 res_conv5_1 区块使用下采样,之后为防止过拟合会通过一个 bottleneck,之后特征图会经过平均池化层,核大小为 24×8 ,经实验验证,本文网络使用平均池化层优于最大池化,分析原因,平均池化层比最大池化层保留的特征信息更多,这有助于各个分支学习强调的特征内容。经过池化层,特征图再通过掩膜 (mask),其掩膜比例经过多次试验调试确定为 (1,0.33)。其中,随机擦除分支掩膜部分会对特征图像按照比例掩膜进行随机擦除,最后得到特征图像,再经过正则化、 1×1 卷积层后特征图维数会降维到 256 得到特征 z_{p_0} ,之后通过全连接层得到特征 $z_{p_0_0}$,在整个过程中,全连接层不共享参数,每个分支到损失函数的路径也独立。

全局学习分支 (Global branch) 仅包含 G_0 分支, G_0 与随机擦除分支的 R_1 、局部学习的 P_0 在结构和设置的区别仅在于 res_conv5_1 区块使用下采样和没有引入 bottleneck、池化层设置的核大小为 12×4 ,而 R_1 与 P_0 的池化层核大小为 24×8 。在该分支,数据经历骨干网络后,通过平均池化层,从 1×1 卷积层输出特征 f_{g_1} ,从全连接层输出特征 $f_{g_1_0}$ 。而 R_1 分支是从 1×1 卷积层输出特征 f_{g_0} ,从全连接层输出特征 $f_{g_0_0}$ 。与 R_1 的构造一致的 P_0 分支,从 1×1 卷积层输出特征 f_{g_2} ,从全连接层输出特征 $f_{g_2_0}$,

局部学习分支 (Part branch) 包含 P_0 、 P_1 分支,而 P_1 分支的数据通过 ResNet-50 stage1,2,3,进入到 resnet 的 conv4 和 conv5 后,也没有在 res_conv5_1 区块使用下采样,同随机擦除 R_0 的分支一样,也引入了 bottleneck,通过核的大小为 24×8 的平均池化层后的特征图会被均分成三等然后分别送往 1×1 卷积层分别获得特征 f_{p_0} 、

f_{p-1} 、 f_{p-2} ，从全连接层输出特征 f_{p-0-0} 、 f_{p-1-0} 、 f_{p-2-0} 。在算法测试部分，本文使用的是每个分支经过降维后的输出特征融合在一起的特征，如公式 (1) 所示：

$$f = z_{p-0} + f_{g-0} + f_{g-1} + f_{g-2} + f_{p-0} + f_{p-1} + f_{p-2} \quad (1)$$

在图 1 的算法网络框架中，一共包含七个小分支的三个大分支的融合特征的过程如下图 2 所示。在图 2 中，左图为数据集 Market-1501 的训练图片，分支上的符号为上文的特征图名称，分别相加后获得测试使用的融合特征 f 。

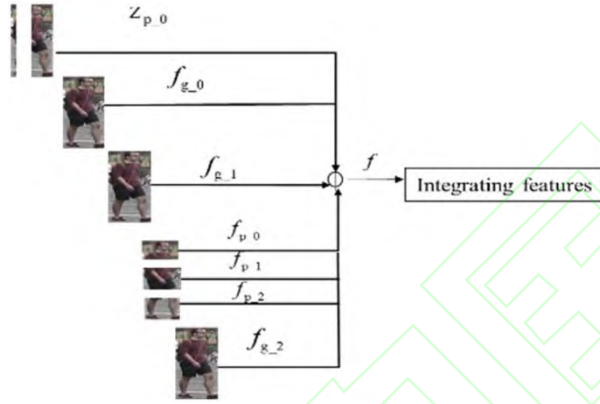


图 2 融合展示

Fig.2 Fusion display

表 2 分支组合维数对比

Table 2 Comparison of branch combination dimensions

分支组合	数量/ (条)	总输出维度/ (维)	测试帧数/(帧/ 秒)
三支	5	1792	2.8
三支+全局学习	6	2048	1.5
三支+局部学习	7	2816	-
三支+随机擦除	7	2304	-

表 2 为三支与四支网络结构输出的融合特征图维数对比。结合数据计算可得三支网络结构 (RE branch、Global branch、Part branch) 加上全局学习分支的四支网络结构的融合特征维数为 2048，同理在三支网络结构的基础上加上随机擦除分支和局部学习分支，其融合特征维数分别在三支融合特征维数的基础上增多 512 和

1024，由于大分支里又包含小分支而在本文设计里骨干网络后分支独立，因此增加一个大分支，除了增大特征图的维数还会增加池化层、卷积层、全连接层的数量，所以三支结构的算法训练更快，测试时大约每秒处理 2.8 帧。而四支网络结构算法在前期训练除了加上全局学习分支的四支结构，其余的四支算法如果不降低 epoch 和 batchsize，就会在训练时出现问题。所以在实验的对比上，四支只选了维数增加最少的三支基础上加入全局学习分支的四支网络组合。其在训练测试时大约每秒处理 1.5 帧。这从侧面反应了三支的网络相比四支测试速度更快，而网络结构也比四支的更为简单，通过实验，也获悉训练时，三支的网络比四支更加容易收敛。

3.2 分支与损失函数

在全局学习分支 (global branch) 中使用同时使用 $L_{triplet}^{[10]}$ 和 $L_{softmax}$ ，网络结构中剩余的两个大分支里作为联合训练的全局学习分支 R_1、P_0 和其设置一样的损失函数。分别是 $L_{triplet}$ 和 $L_{softmax}$ ，其损失函数如公式 (2) (3) 所示。在公式 (2) 中， N 代表训练的过程， C 代表数据集， w_k^T 代表数据集中行人类别对应的权重向量， f_i 代表各个分支学习到的特征，包含图 1 中通过全连接层的特征。在公式 (3) 中， f_a 、 f_p 、 f_n 分别对应 anchor 的特征、挖掘到的正样本特征和负样本特征，都是从训练的 mini-batch 获取，包含图 1 所示全局学习分支降维后输出的特征。 p 为每个 batch 训练时，抽取的行人数目， k 对应训练中抽取每个行人抽取的图片数量。P_1 分支的三个小分支和随机擦除分支的 R_0 分支失函数为公式 (2) 中的 $L_{softmax}$ 。

$$L_{softmax} = - \sum_{i=1}^N \ln \frac{e^{w_{y_i}^T f_i}}{\sum_{k=1}^C e^{w_k^T f_i}} \quad (2)$$

$$L_{triplet} = - \sum_{i=1}^p \sum_{a=1}^k \left[\alpha + \max_{p=1 \dots K} \|f_a^{(i)} - f_p^{(i)}\|_2 - \min_{\substack{n=1 \dots K \\ j=1 \dots P \\ j \neq i}} \|f_a^{(i)} - f_n^{(j)}\|_2 \right] \quad (3)$$

3.3 最小二乘法的权重分配

多任务学习分支在训练过程中存在分支由于复杂度和学习率不同,从而导致分支间训练时有干扰现象。为缓解上述问题,本文在设计网络结构时,分别在随机擦除分支和局部学习分支额外引入全局学习分支进行联合训练。但还是存在一定的干扰性,因此本文又在损失函数引入权重分配。有算法^[11]可对多任务学习网络进行损失函数权值调整,但本身具有不稳定性,同时也会大大提升网络复杂度。由于本文仅使用了两个 loss,且公式(2) $L_{soft \max}$ 的使用率远大于公式(3) $L_{triplet}$ 的使用率,因此在调整权重时,为降低复杂度,本文仅关注 $L_{soft \max}$ 的权重值,而将 $L_{triplet}$ 的权重设为 1,最后将总损失 L 定义为公式(4),其中 λ 为损失函数对应的权重, j 对应本文的网络分支。

$$L = \sum_{j=1}^3 L_{triplet} + \lambda \sum_{j=1}^7 L_{soft \max} \quad (4)$$

最小二乘法是在曲线预测中被使用广泛,同时高阶的非线性函数曲线易出现波峰。而本文选择的算法评判标准之一的 mAP (Mean Average Precision),其定义是将图片的相似度从高到低,统计从第一项到最后项相同行人图片之间正确识别的平均概率。因此本文将 mAP 设为纵轴、将 λ 设为横轴,利用实验数据结合最小二乘法预测从低阶到高阶的 mAP 和 λ 的曲线,借此找到 mAP 的最高波峰对应的 λ 区域,然后进行 loss 权重 λ 的调节,通过这样的方式可减少不必要的工作量,其中误差公式如公式(5)所示。在公式(5)中, mAP 为本文通过实验获取的实际平均精度均值 y_i 为本文结合最小二乘法获得的预测值。

$$\min \left[\sum_{i=1}^m ((m a p_i - y_i)^2) \right] \quad (5)$$

4 实验与结果分析

本文的实验使用的数据集是在行人重识别中具有代表性的 Market-1501 和 DukeMTMC-reID, CUHK03-Label 和 CUHK03-detect。在实验指标上选择 Rank- n 和 mAP, Rank- n 指将图片的相似度从高到低排序,前 n 项包含被查询行人图片的概率。实验服务器 GPU 配置是英伟达 1080Ti, 框架为 pytorch, 为显示清楚,下文表格和正文中的数据集名称使用略写: DukeMTMC-reID (Duke), Market-1501 (Market), CUHK03-Label (Label), CUHK03-Detect (Detect) 表 3 为数据集详细信息。

表 3 数据集信息

数据集	训练集图片 数/(张)	训练集人数 /(个)	测试集图 片数/(张)	测试集人数 /(个)
Duke	16522	702	17661	702
Market	12936	751	19732	750
Label	7368	767	5328	700
Detect	7365	767	5332	700

根据图 3 可知,数据集 Market 拍摄于夏日白天,因此数据集的光线比冬天拍摄的 Duke 和在过道拍摄的 Label、Detect 要亮,且场景简单,同时 Duke 的行人被遮挡现象更为严重,背景时而清晰时而与行人混杂,而 Label 和 Detect 存在较多行人侧面和背面的图像,同时根据表 3,也可知, Detect 的训练行人类型少于 Label。



图 3 Market, Duke, Label, Detect 数据集对比

Fig.3 Comparison of Market, Duke, Label, Detect dataset

因此相比之下,四个数据集的难度从易到难的排序分别为 Market、Duke、Label、Detect。然后本文基于数据集 DukeMTMC-reID 随机地调公式(3)中的 λ , 分别 2, 5, 10, 12, 15, 20, 30,

40, 然后运行算法获得八组 mAP 结果数据 :75.24 , 75.96 , 75.58 , 75.64 , 74.87 , 75.02 , 75.66 , 74.33 , 73.12 , 因 λ 大于 20 时 , mAP 下降显著 , 所以将横轴 λ 限定在 (0 , 20) 间 , 同时为清晰显示 , 将 mAP 与 λ 的预测曲线的纵轴显示区域限定在 (74 , 100)

之间。因为曲线阶数越高, 波动越明显, 因此本文从低阶到高阶做曲线预测, 即从 1 阶开始实验, 直到预测的 mAP 曲线出现波峰为止, 最后结果如图 4 所示。

不同次数方程曲线拟合结果, 从1到9阶

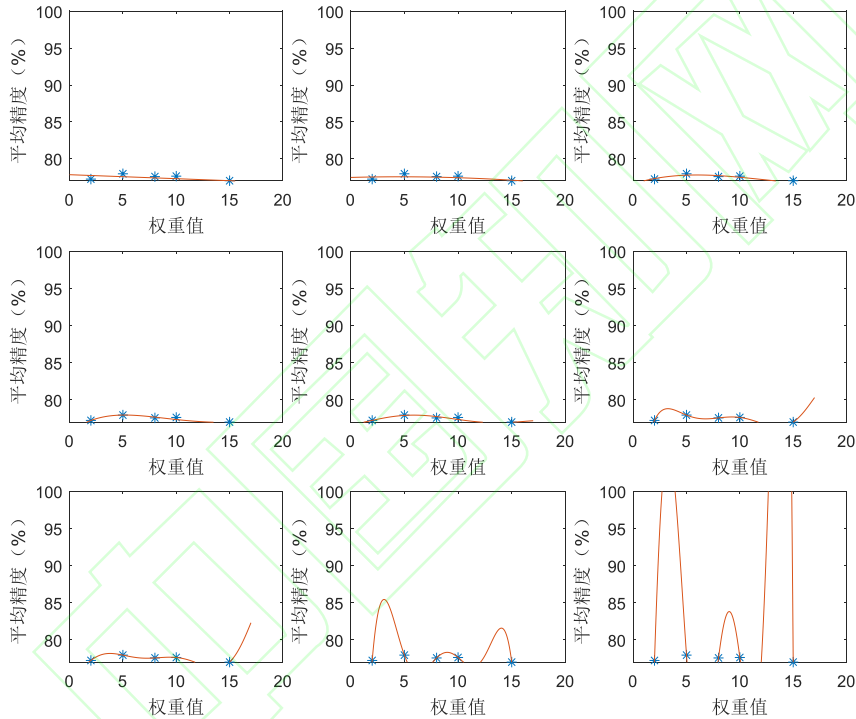


图 4 最小二乘法预测曲线

Fig.4 Least Squares Forecast Curve

当预测曲线为 8 阶时, 得到调试权重的建议即 λ 在 2 到 5 之间有最高的 mAP 出现, 最后经过在两个数据集上的实验, 最后使算法结果到达最佳的 λ 为 3。综上所述, 本文又分别在四个数据集进行上实验, 为了提升, 本文在数据处理部分加入了常用的加强算法 Random Erasing^[12], 但精度不升反降。分析原因, 可能是因为随机擦除了行人的关键特征, 导致模型学习到的行人特征变少, 影响正负样本的挖掘。最后本文加入 Re-Rank^[13]算法, 精度再此提升, Re_Rank 算法原理是在利用马氏距离和 Jaccard 距离, 在模型训练完成后再进行挖掘正负样本的工

作, 以此更好地提升精度^[13]。

表 4 Duke、Market 数据集结果对比

Table 4 Comparison of Duke and Market data set results

数据集	Duke mAP/%	Duke Rank—1/%	Matket mAP/%	Market Rank—1/%
算法名称				
PAN[14]	51.5	71.6	63.4	82.8
SVDNet[15]	56.8	76.7	62.1	82.3
DPFL[16]	60.3	79.2	73.1	88.9
BDB[9]	72.1	86.8	84.3	94.2
本文算法	76.42	87.3	85.08	94.17
本文算法 +Re-Rank-	88.11	91.14	93.19	95.5

为了进一步验证算法的有效性, 本文又分别选取了四个算法作对比, 分别是在基础网络结构上引入行人对齐的 PAN^[14], 一个在全连接层引入奇异值分解的 SVDNet^[15], 一个提出深层金字塔特征学习的 DPEL^[16], 以及本文参考仅使用全局学习分支和随机擦除分支未加入局部学习分支的论文方法 BDB^[9], 在数据集 Duke 和 Market 的对比结果如表 4 所示, 在数据集 Label 和 Detect 的对比结果如表 5 所示。

从表 4 获悉, 本文的算法未加强也比前三个算法在 mAP 和 Rank-1 分别高出十多个点, 相比参考方法^[9], 在两个数据集上, mAP 分别高出了 4 个点和 1 个点, 而经过加强的本文算法结果在数据集 Duke、Market 上 mAP 更是到达了 88.11 和 93.19, Rank-1 分别能达到 93.14、95.5。

从表 5 获悉, 在数据集 Label 与 Detect 上, 相比前三个算法精度提高了至少 10 点及以上, 而本文加入 Re-Rank 的算法精度更是提高了二十多点, 在数据集 Label 和 Detect 上 mAP 分别到达了 76.26 和 74.86, Rank-1 分别到达了 74.07 和 72.57。表 7 中, BDB^[9]算法虽然在 Label、Detect 数据集上的结果比本文未加强的算法稍好, 但本文的未加强的算法精度在 Market 和 Duke 数据集上的精度分别领先 1 一个点和 4 个点, 且 Label、Detect 数据集上, 本文加强后的算法结果比其 mAP, 分别领先 4.56 个点和 5.56 个点。

表 5 Label、Detect 数据集结果对比

数据集	Duke	Duke	Market	Market
	mAP/%	Rank-1/%	mAP/%	Rank-1/%
算法名称				
PAN ^[14]	51.5	71.6	63.4	82.8
SVDNet ^[15]	56.8	76.7	62.1	82.3
DPFL ^[16]	60.3	79.2	73.1	88.9
BDB ^[9]	72.1	86.8	84.3	94.2
本文算法	76.42	87.3	85.08	94.17
本文算法	88.11	91.14	93.19	95.5
+Re-Rank				

5 结束语

本文提出了一种融合全局学习分支、随机擦除分支、局部学习分支的算法, 并结合最小二乘法进行权重分配, 相比对行人进行骨骼定点和增加行人属性的算法, 无需预训练也不需对数据集进行额外的属性标注。最后在 pytorch 框架下, 分别在三个数据集上, 做了对比实验, 其结果显示, 本文融合全局学习分支、随机擦除分支、局部学习分支的算法具有有效性, 在结果上相比参考算法, 得到一定提升。在接下来的工作中, 本文就加强网络对行人特征学习, 将结合对齐框架进行下一步的研究。

参考文献:

- [1] 徐家臻, 李婷, 杨巍. 多尺度局部特征选择的行人重识别算法[J]. 计算机工程与应用, 2020, 56(2): 141-145.
Xu Jiazhen, Li Ting, Yang Wei. Person Re-identification algorithm based on multi-scale local feature selection [J]. Computer engineering and applications, 2020, 56 (2): 141-145.
- [2] 崔鹏, 赵莎莎, 范志旭. 行人检测的深度稀疏自编码网络[J]. 计算机科学与探索, 2019, 13(6): 1027-1037.
Cui Peng, Zhao Shasha, fan Zhixu. Deep sparse self coding network for Person Re-identification[J]. Computer science and exploration, 2019, 13 (6): 1027-1037
- [3] 徐思敏, 胡士强. 多属性融合网络的行人重识别方法[J]. 计算机工程与应用, 2020, 56(6): 126-132.
Xu Simin, Hu Shiqiang. Person Re-identification method based on multi-attribute fusion network[J]. Computer engineering and application, 2020, 56 (6): 126- 132
- [4] Zheng Liang, Yang Yi, Hauptmann Alexander G. Person Re-identification: Past, present and future [EB/OL]. 2016. <https://arxiv.org/pdf/1610.02984.pdf>
- [5] Zhao H, Tian M, Sun S, et al. Spindle Net: Person Re-identification with Human Body Region Guided Feature Decomposition and Fusion[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Waston D.C, U.S.A, September 6-9, 2017, IEEE, 2017:1077-1085.
- [6] Wei L, Zhang S, Yao H, et al. GLAD: Global-Local-Alignment Descriptor for Pedestrian Retrieval[C]//ACM Multimedia Conference, Dallas, USA, Oct 30th-Nov 3rd, 2017, ACM, 2017:420-428.
- [7] Su C, Li J, Zhang S, et al. Pose-driven Deep Convolutional Model for Person Re-identification[C]//IEEE In-

- ternational Conference on Computer Vision , Waston D.C,U.S.A, September 6-9,2017, IEEE, 2017: 3960-3969.
- [8] Wang G , Yuan Y , Chen X , et al. Learning Discriminative Features with Multiple Granularities for Person Re-identification[C]//2018 ACM Multimedia Conference., Beijing, China, June 12-15, 2018, ACM, 2018: 274-282.
 - [9] Dai Z , Chen M , Gu X , et al. Batch DropBlock Network for Person Re-identification and Beyond[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea, October 27 to November 2, 2019, IEEE, 2019: 3691-3701.
 - [10] Alexander Hermans, Lucas Beyer, Bastian Leibe, In Defense of the Triplet Loss for Person Re-identification[EB/OL].2017.<https://arxiv.org/abs/1703.07737>.
 - [11] Kendall A,Gal Y,Cipolla R.Multi-Task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics[C]//IEEE Conference on Computer Vision and Pattern Recognition,Salt Lake City, USA,June 18-22,2018,IEEE, 2018:7482-7491.
 - [12] Zhong Z, Zheng L, Kang G, et al. Random Erasing Data Augmentation[C]//AAAI, San Francisco,U.S.A., February 4 -9, 2017, AAAI, 2017:13001-13008.
 - [13] Zhong Z, Zheng L, Cao D, et al. Re-Rank-ing Person Re-identification with k-reciprocal Encoding[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR).Waston D.C U.S.A,September 6-9, 2017, IEEE, 2017: 1318-1327.
 - [14] Zheng Z, Zheng L, Yang Y. Pedestrian Alignment Network for Large-scale Person Re-identification[J]. IEEE Transactions on Circuits & Systems for Video Technology, 2017, 10(6): 1005-1015.
 - [15] Sun Y, Zheng L, Deng W, et al. SVDNet for Pedestrian Retrieval[C]. IEEE International Conference on Computer Vision, Venice, Italy, October 22-29, 2017, IEEE, 2017: 2590-2600.
 - [16] Yanbei Chen, Xiaotang Zhu, Shaogang Gong, et al. Person Re-identification by deep learning multi-scale representations[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Waston D.C,U.S.A, September 6-9,2017,IEEE,2017:2590-2600.