

DOI: 10.13382/j.jemi.B2104857

利用全局与局部关联特征的行人重识别方法^{*}

张勃兴¹ 马敬奇² 张寿明¹ 李辰潼² 钟震宇²

(1. 昆明理工大学信息工程与自动化学院 昆明 650504; 2. 广东省科学院智能制造研究所
广东省现代控制技术重点实验室 广州 510095)

摘 要:针对因行人图像背景差异大、人体外观相似导致的行人再识别准确率低的问题,提出了一种利用特征融合与多尺度信息的行人重识别方法。首先,通过 ResNet50_IBN 提取人体图像全局特征图。其次,设计分支结构,第 1 分支利用空间变换网络对全局特征图进行自适应的空间特征对齐,水平切分全局特征图得到局部特征,采用全局特征与每个局部特征分别融合的方式来挖掘特征之间的关联关系。第 2 分支增加了 4 种不同尺度的卷积层提取全局图像的多尺度特征。最后,在推理阶段将第 1 分支和第 2 分支的特征进行通道维度的串联,作为行人的对比特征。通过在 Market-1501、DukeMTMC 数据集上的实验表明,所提方法与 AlignedReID 和 EA-Net 等特征对齐和局部特征提取方法相比具备更强的性能,在 Market-1501 上,mAP 和 Rank-1 分别达到了 86.77%和 94.83%。

关键词: 行人重识别;ResNet50;空间变换网络;特征融合;多尺度特征

中图分类号: TP391;TN911.7 **文献标识码:** A **国家标准学科分类代码:** 510.40

Person re-identification method based on global and local relation features

Zhang Boxing¹ Ma Jingqi² Zhang Shouming¹ Li Chentong² Zhong Zhenyu²

(1. School of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650504, China;
2. Institute of Intelligent Manufacturing, Guangdong Academy of Sciences, Guangzhou 510095, China)

Abstract: A person re-identification method based on feature fusion and multi-scale information was proposed to solve the problem of low accuracy of person re-identification due to the large difference of human image background and similar global appearance of human body. Firstly, the global feature map of human body image is extracted by ResNet50. Secondly, the branch structure is designed. In the first branch, the spatial transformation network is used to align the global feature images adaptively, and the local feature images are obtained by horizontal segmentation of the global feature images. The correlation between the global feature and each local feature is mined by fusing the global feature and each local feature separately. The second branch adds four convolution layers of different scales to extract multi-scale features from global images. Finally, in the reasoning stage, the features of the first branch and the second branch are connected in series as the comparative features of person. Experiments on the Market-1501 and DukeMTMC datasets show that the proposed method has better performance than the AlignedReID and EA-NET feature alignment and local feature extraction methods. In the Market-1501 dataset, mAP and Rank-1 reach 86.77% and 94.83%, respectively.

Keywords: person re-identification; ResNet50; spatial transformation network; feature fusion; multiscale feature

0 引言

近年来随着深度学习技术的快速发展,行人重识别(Re-ID)算法的准确率不断提高,成为处理监控视频大数据的有效手段,在智能监控、智慧安防等社会公共安全领域有着广阔的应用前景,为相关部门通过监控系统快速定位嫌疑人提供了非常高效的途径。行人重识别技术的应用场景一般是跨摄像机的,图像具有大视场、远距离、背景复杂的特点,有不少挑战性,近年来一直是热点研究课题。

目前,基于深度学习的行人重识别方法中结合全局特征和局部特征来识别的策略是应用最广泛且相对高效、稳定的方法。其中,全局特征提取方式简单、推理速度快,但特征不够丰富,容易忽略局部细节信息,不适用于对复杂样本的识别。而提取局部特征可以让网络的关注范围更加广泛,提取到的特征更加多样,极大改善模型的识别效果,同时提高模型的鲁棒性和泛化性^[1]。因此,局部特征提取方法是目前的研究重点。

当前主流的行人局部特征提取方法包括:1)按照人的形体结构等先验知识得到行人图像不同语义部件的特征,作为行人图像的局部特征表示^[2-4];2)将深度特征图水平均匀切分,得到覆盖图像不同区域的局部特征^[5-6];3)利用注意力机制使网络关注显著的细节信息^[7-8]。这些方法能使网络提取更有区分度的局部细节特征,已有研究也证明了该类方法的有效性。

在特征提取网络中,残差网络凭借性能和速度的优势,已成为行人重识别方法最常用的主干网络,结合深度特征图水平切分策略,获得了大量应用。其中,Sun 等^[5]提出了行人局部特征提取网络(PCB),将图像深度特征图水平均匀切分为 6 份,单独用每个局部特征进行行人身份预测,有效提升了识别效果。Wang 等^[6]提出了基于多粒度判别特征的行人重识别方法(MGN),通过改变不同局部分支的水平条带数量来得到不同粒度的局部信息。但是,该类方法没有考虑局部特征未对齐以及局部特征之间语义相关性的影响。因此,Fan 等^[9]提出了联合整体和局部行人重识别的空间通道并行网络(SCPNet),将水平切分得到的局部特征进行通道维度的串联来获得全局特征。Jaderberg 等^[10]和 He 等^[11]都提出了局部特征融合策略,挖掘不同局部特征之间的相关性,获得全局范围的相关性表示。这些方法都是基于局部特征的融合,没有将全局特征有效利用起来。除此之外,注意力机制也被广泛应用在行人重识别网络中,给显著特征赋予更高的权重,如 CBAM 网络^[12]将通道注意力模块和空间注意力模块进行串联融合,达到对重要特征的增强,对无关特征的抑制。Chen 等基于通道和空间注意力提出了显

著特征级联抑制网络,通过下层级联阶段对上层级联阶段的抑制来提取潜在的局部特征,还设计了残差双注意力模块和多阶段特征融合模块,挖掘出覆盖整张特征图的潜在特征。但是,注意力机制可能会造成一部分有用信息的损失。

目前的大多数研究方法聚焦在行人全局信息或局部信息的提取上,忽略了全局信息与局部细节之间的关联关系。此外,监控场景多样,环境复杂,行人目标尺度变化大,行人多尺度信息未得到合理的利用。因此,研究行人的全局特征图与局部细节特征的融合方法,同时结合行人的多尺度信息,来提高行人重识别算法的准确率,是十分必要的。

为此,提出了一种利用特征融合与多尺度信息的行人重识别方法,主要工作和创新如下:

1)将 ResNet50_IBN 作为主干网络提取行人图像全局特征图,引入了 STN^[13]实现了全局特征的对齐,通过全局和局部特征融合的方式,实现整体语义信息和显著性局部信息的聚合;

2)采用多种尺度的卷积层来提取深度特征图的多尺度信息,获取更加丰富的全局特征表示,提高模型的甄别能力。

1 方法

为同时兼顾行人重识别的准确率与识别速度,本文采用在 ImageNet 上预训练的 ResNet50_IBN^[14]为主干网络,删除最后的池化层和全连接层,将网络第 4 阶段的卷积步长设置为 1,以得到更大、信息更丰富的特征图^[15]。其中的 IBN 是一个通用的风格信息过滤网络,通过在 ResNet50 前 3 个阶段共同使用 IN 和 BN 进行数据归一化的方式来提升模型对图像色彩变换的鲁棒性。行人图像数据经过 ResNet50_IBN 网络处理后得到了全局的行人图像特征图。接着网络分为两个分支,第 1 分支中引入了 STN 模块,对行人图像特征图进行了自适应对齐,消除了行人图像中的多余背景区域特征。对齐后的特征采用水平切分策略,将特征图水平均匀切分为 6 份,得到 6 个局部特征,并将每个局部特征图分别与全局特征图融合,挖掘出全局与局部特征之间的关联关系。第 2 分支增加了 1×1 、 3×3 、 5×5 、 7×7 这 4 种不同尺度的卷积核提取全局特征的多尺度信息,有效提高了对全局多尺度特征的表达能力。在行人图像对比检索时,将局部特征、融合特征、多尺度特征进行通道维度的串联,作为最终的特征表示,网络结构如图 1 所示。

1.1 空间特征对齐

由于实际监控场景的复杂性,目标检测模型得到的行人图像其空间分布并不一致,一些行人图像中的人体

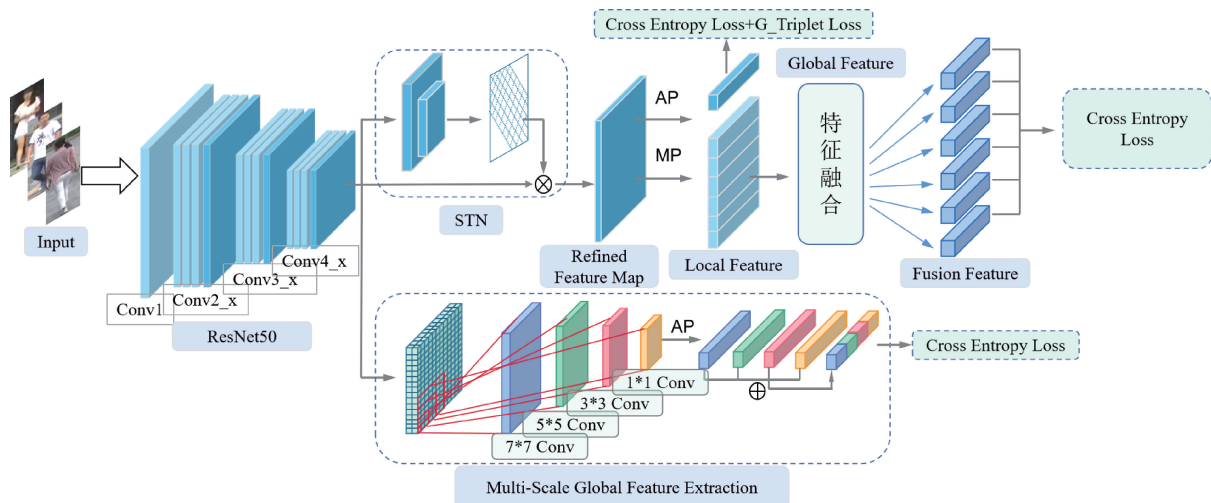


图1 网络结构

Fig. 1 Network structure

区域仅占一部分,其余为多余的背景区域,影响行人图像特征的提取。特别是在切分得到局部特征时,空间错位会造成局部特征的显著差异,干扰模型的判别。因此本文引入了空间变换网络 STN,自适应地对图像特征做仿射变换,通过对特征图的放缩、平移,将特征图中的人体区域放大,过滤掉无用的背景信息,空间对齐示意如图 2 所示。

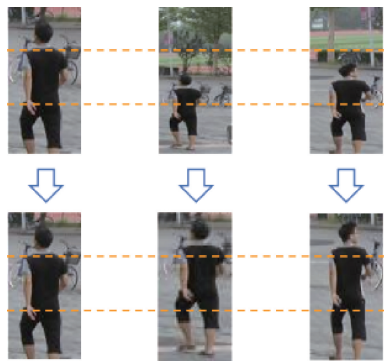


图2 行人空间特征对齐

Fig. 2 Alignment of person spatial features

STN 包含 3 个部分,分别为定位网络、网格生成器和采样器,通过仿射变换系数矩阵采样来得到变换结果,变换过程如下所示:

$$\begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{Source} \\ y_{Source} \\ 1 \end{bmatrix} = \begin{bmatrix} x_{Target} \\ y_{Target} \\ 1 \end{bmatrix} \quad (1)$$

式中: (x_{Source}, y_{Source}) 表示原图像像素点, (x_{Target}, y_{Target}) 表示仿射变换后的图像像素点,系数矩阵 θ 即为仿射变换系数, $(\theta_{12}, \theta_{21})$ 为旋转因子, $(\theta_{11}, \theta_{22})$ 为放缩因子, $(\theta_{13}, \theta_{23})$ 为平移因子,网络通过训练来学习 θ 矩阵。由

于旋转对特征对齐是无用的,故本文将 STN 系数矩阵的旋转因子置为 0,定位网络由一个 1×1 卷积层、批归一化层、激活层和一个最大池化层组成,全连接层回归得到仿射变换系数矩阵,即可对特征图进行变换^[16]。

1.2 特征融合

行人特征图水平划分获取局部特征的策略符合人体垂直视点分布的特点,有利于提高模型的稳定性。将主干网提取的 $H \times W \times C$ 的特征图池化为 $6 \times 1 \times C$,水平切分得到 6 个 $1 \times 1 \times C$ 的特征向量代表 6 个不同部位的局部特征。但是,这种切分方式破坏了局部特征之间的语义相关性,而且特征彼此之间是独立的。但在现实情况下,相似样本是很常见的,例如两个穿衣风格很相似的行人,他们的某些局部特征可能高度相似,单独利用这些局部特征很容易将他们判定为同一个行人。因此,本文提出了将全局特征图分别与每个局部特征图融合的方式,挖掘出全局与局部特征之间的关联关系,实现了整体语义信息和显著性局部信息的聚合,可以提高模型对相似样本的识别能力,融合方法如图 3 所示。

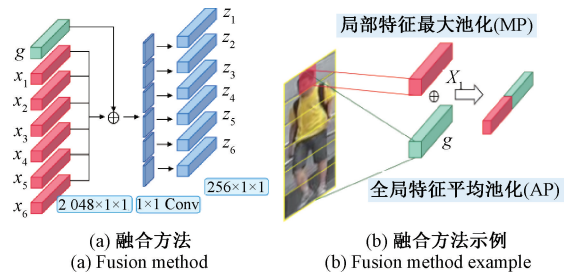


图3 特征融合

Fig. 3 Feature fusion

特征图在经过空间变换网络对齐之后分两路分别进

行全局平均池化和最大池化,得到两个 $1 \times 1 \times C$ 和 $6 \times 1 \times C$ 的特征向量。 $1 \times 1 \times C$ 的特征向量代表全局特征,保留了整体的特征信息。 $6 \times 1 \times C$ 的特征向量水平切分得到 6 个 $1 \times 1 \times C$ 的特征向量,分别代表人体的 6 个局部特征。接着,分别将每个局部特征和全局特征叠加来获得关联特征,融合过程可以表示为:

$$z_i = C\left(\frac{1}{5}g + x_i\right) \quad (i = 1, \dots, 6) \quad (2)$$

其中, g 表示全局特征, $x_i (i = 1, \dots, 6)$ 表示局部特征, C 代表通道维度的降维操作,通过 1×1 卷积层将融合后的特征向量从 2 048 维降到 256 维, z_i 代表最终的关联特征。

1.3 多尺度特征提取

为增强网络提取全局特征图多尺度信息的能力,让网络捕获更多有用的信息,本文在对未经过空间变换网络的行人图像全局特征引入了 4 种不同尺度的卷积层,并将卷积核为 $n \times n$ 的卷积层用 $n \times 1$ 和 $1 \times n$ 卷积层串联的方式代替,有效节约了参数量,垂直和水平卷积核还有助于提升模型的场景解析能力。

如图 4 所示,增加了 $1 \times 1, 3 \times 3, 5 \times 5, 7 \times 7$ 这 4 种不同尺度的卷积核,提取全局图像的多尺度特征,输出 4 个不同尺度的全局特征。然后分别对 4 个不同尺度的全局特征进行全局平均池化,之后叠加得到一个融合的多尺度特征。每个尺度的特征都被用来做为该行人的特征表示。在特征图和分类层增加一个归一化层,使同一个人的特征更加紧密,网络更容易收敛。

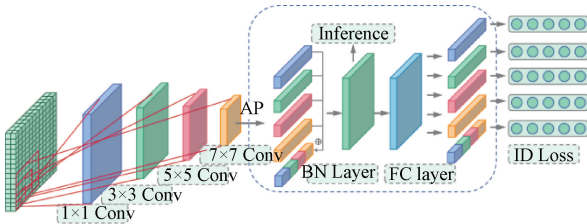


图 4 多尺度特征提取

Fig. 4 Multi-scale feature extraction

1.4 损失函数

损失函数采用了交叉熵损失 l_{cross} 和三元组损失 $l_{triplet}$ 结合的方式,三元组损失用来增加类间距离,减小类内距离,以此来提升检索的准确性,交叉熵损失用来衡量真实概率与预测概率之间的分布差异,损失函数如下所示:

$$l = l_{triplet}^g + l_{cross}^g + l_{cross}^f + l_{cross}^p + l_{cross}^m \quad (3)$$

其中, $l_{triplet}^g$ 和 l_{cross}^g 分别代表对全局特征计算三元组损失和交叉熵损失, l_{cross}^f 代表融合特征的交叉熵损失, l_{cross}^p 代表局部特征的交叉熵损失, l_{cross}^m 代表多尺度特征的交叉熵损失。

交叉熵损失的定义为:

$$l_{cross} = - \sum_{i=0}^n p(x_i) \log(q(x_i)) \quad (4)$$

其中, n 表示一个批次行人图片总数, $p(x_i)$ 表示真实的概率分布, $q(x_i)$ 表示预测的概率分布。

三元组损失定义为:

$$l_{triplet} = - \sum_{i=1}^p \sum_{a=1}^R [\tau + \max_{p=1, \dots, r} \|F_a^i - F_p^i\|_2 - \min_{\substack{n=1, \dots, r \\ j=1, \dots, p \\ j \neq i}} \|F_a^i - F_n^j\|_2]_+ \quad (5)$$

其中, P 表示一个最小批次行人图片总数, R 表示一个最小批次中相同身份标签的图片数量。 τ 是一个阈值参数用来约束特征空间中正样本对和负样本对之间的距离, F_a^i, F_p^i, F_n^j 分别代表采样得到的锚、正样本和负样本的特征,正样本特征指与锚具有相同的标签,负样本特征指与锚的标签不同。

2 实验

2.1 实验环境和数据

所有实验在 64 bit 的 Ubuntu18.04 环境下进行,采用 Python 编程语言和 Pytorch 深度学习框架,硬件环境为两张 NVIDIA_Tesla_P100 显卡。

实验数据包括了 Market-1501^[17]、DukeMTMC-reID^[18] 两个常用的行人重识别数据集。Market-1501 数据集包括由 6 个摄像头拍摄到的 1 501 个不同身份的行人,包含 32 668 幅行人图像,DukeMTMC-reID 数据集使用 8 个摄像头采集得到,其中包含 1 812 个行人的 36 411 幅图像。

2.2 参数设置

对输入图像采用水平翻转和随机遮挡进行数据增强,输入图像的尺寸被调整到 384×128 , batch_size 设置为 64,采用 Adam 优化器来训练模型,初始学习率为 2×10^{-4} ,训练 400 个 epoch,在第 320 和第 380 个 epoch 学习率分别乘以 0.1,训练大约需要 4 h。

分别采用首次(Rank-1)、前 5 次(Rank-5)、前 10 次(Rank-10)和平均识别准确率(mAP)来衡量模型的性能。

2.3 消融实验

为了验证所提网络和模块的有效性,进行了消融实验,实验过程中 Baseline 采用 PCB 网络结合全局特征,使用交叉熵和三元组两种损失函数,数据集采用 Market-1501,分别在 Baseline 的基础上增加了 IBN(风格归一化)、MGF(多尺度全局特征)、STN+LF(融合特征)、STN+LF+MGF,实验结果如表 1 所示。原始 PCB 网络的 mAP 为 77.4%, Rank-1 为 92.3%, 本文方法 mAP 和

Rank-1 可以分别达到 86.77% 和 94.63%, 证明了所提方法的有效性。而在 Baseline 的基础上分别融合 MGF 或 STN+LF 也起到了积极作用, 体现出了一定的效果。

表 1 消融实验
Table 1 Ablation experiments

方法	Rank-1	mAP
Baseline	94.09%	84.19%
Baseline + IBN	94.41%	85.71%
Baseline + IBN + LF	94.60%	86.07%
Baseline + IBN + STN + LF	93.94%	86.56%
Baseline + IBN + STN + LF + MGF	94.83%	86.77%

2.4 对比实验

为验证本文方法的有效性和性能, 分别在 Market-1501 和 DukeMTMC-reID 数据集上与 PCB、DATRL-ReID、EA-Net 等近 3 年的行人重识别方法开展对比实验与分析。在 Market-1501 数据集实验结果对比如表 2 所示, 本文方法的 Rank-1 与 mAP 与对比方法相比均为最优, 分别达到了 94.8% 和 86.7%, Rank-5 与性能最优的 DATRL-ReID 的 98.1% 相比差 0.2%, Rank-10 与性能最优的 DATRL-ReID 的 98.8% 相比差 0.1%。在 DukeMTMC-reID 数据集实验结果对比如表 3 所示, 本文方法的 Rank-1 为 88.4%、Rank-5 为 94.5%、Rank-10 为 96.1% 及 mAP 为 77.5%, 对比性能指标均为最优。相比而言, 本文方法同时提取了覆盖整幅行人图像的局部特征和多尺度全局特征, 可以应对复杂开放环境下行人背景差异大、人体全局外观相似的重识别任务, 具有一定的优势。

表 2 Market-1501 数据集实验结果对比
Table 2 Comparison of experimental results of Market-1501 dataset

方法	Market-1501			
	r = 1	r = 5	r = 10	mAP
PCB	92.3%	97.2%	98.2%	77.4%
PCB+RPP	93.8%	97.5%	98.5%	81.6%
DATRL-ReID ^[19]	94.4%	98.1%	98.8%	81.5%
AlignedReID ^[20]	92.6%	—	—	82.3%
EA-Net ^[21]	94.6%	—	—	85.6%
HOReID	94.2%	—	—	84.9%
BoT	94.5%	—	—	85.9%
Ours	94.8%	97.9%	98.7%	86.7%

所对比方法包括水平切分提取部件特征、水平特征对齐、注意力机制、人体语义分割。该类方法以各种先验信息来制定人体部件特征提取策略。其中, PCB 网络首次提出了单独利用每块特征进行行人身份度量, PCB+RPP 是在 PCB 方法上对局部特征划分方法的改进, 但是

表 3 DukeMTMC-reID 数据集实验结果对比
Table 3 Comparison of experimental results of DukeMTMC-reID dataset

方法	DukeMTMC-reID			
	r = 1	r = 5	r = 10	mAP
PCB	81.9%	—	—	65.3%
PCB+RPP	83.3%	—	—	69.2%
DATRL-ReID ^[19]	86.3%	93.1%	95.1%	72.9%
AlignedReID ^[20]	82.1%	—	—	69.7%
EA-Net ^[21]	87.5%	—	—	74.6%
HOReID	86.9%	—	—	75.6%
BoT	86.4%	—	—	76.7%
Ours	88.4%	94.5%	96.1%	77.5%

该方法未考虑特征对齐, 也没有融合全局特征, 难以识别外观相似度较高的行人。AlignedReID 和 PPS 分别提出了基于最短路径和特征关联的方法来改善人体特征未对齐的影响, 但是两种方法并没有单独提取局部特征, 制约了模型的识别能力, 所有指标均低于本文方法。EA-Net 和 HOReID 则分别提出了基于人体语义分割和人体骨架关键节点的局部特征提取方法。该类方法依据人体结构构造来提取局部特征, 需要引入新的模型, 造成网络参数量庞大, 而且这种过于细粒度的划分并未取得可观的结果。相比之下, 水平切分获取局部特征已经被证明是该领域目前最稳定且最高效的识别方法。BoT 方法是行人重识别领域高效的 Baseline, 融合了人体特征提取方面比较有效的数据增强方案、损失函数设置方法以及训练策略。通过对比发现, 所提方法在主流数据集上可以达到当前大多数方法的识别效果。具体准确率与其他方法的对比结果如表 2 和 3 所示。

图 6 通过可视化展示了 PCB 方法与本文方法在 Market-1501 数据集上进行行人重识别的 Top-10 检索结果。图 5(a) 为挑选的部分行人的输入 query 查询图像, 即模板图像, 图 5(b) 为 PCB 方法的行人重识别结果, 图 5(c) 为本文所提方法的行人重识别结果。图中粗线框为检索错误的结果。行人①的 PCB 方法检索有 1 个错误结果, 错误结果图像中的行人与行人①高度相似, 但错误结果图像中的行人未佩戴帽子, 即未能将局部特征和全局特征结合使用, 本文方法可以正确检索。在行人②的检索中出现了多个错误结果, 主要来源于模型对一些颜色特征的混淆, 可以看出 PCB 方法过于关注细节而忽略了整体信息。行人③和行人④的检索结果当中, 检索到的错误样本和模板样本虽然有一些细节上的差异, 但是整体外观相似度还是很高的, 两种方法的检索结果可以看出, 本文方法具备更强的特征提取能力, 证明了所提方法在一定程度上改善了模型对相似样本的识别能力。

2.5 系统测试

为验证本文算法在实际监控系统中的性能表现, 依

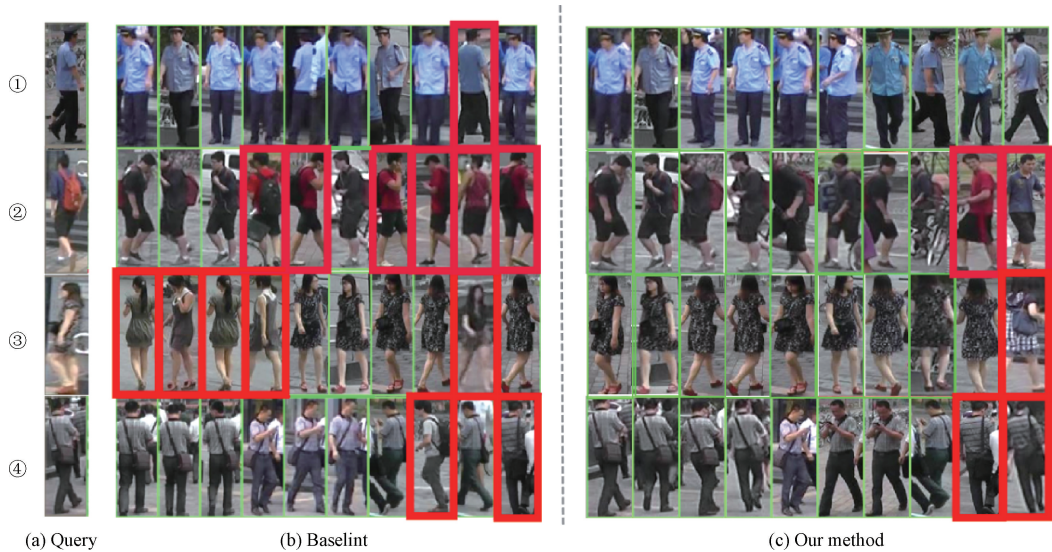


图 5 识别效果可视化

Fig. 5 Recognition effect visualization

托实习单位广东省科学院智能制造研究所搭建的室内场景行人重识别实验平台(包括 7 个监控相机)。7 个相机分别监控不同的场景,场景光照条件、背景信息、视场范围、拍摄角度等都不相同,参与的测试人员 12 人(包括 11 名男性和 1 名女性),参与实验的人员所穿衣物均为日常服饰。实验过程中以图 6 视场 1 中箭头所指检测框行人为识别目标,该行人穿白色衣服+黑色裤子,是日常生活中较为常见的搭配风格。测试人员中有多个与目标行人穿搭风格类似,如视场 1 中其余检测框所示两个测试人员,均穿白色衣服+黑色裤子。如图 6 所示,测试过程中,测试人员与识别目标行人均在 7 个视场内随机走

动,行人目标在经过视场 1~5 时能被准确实时的识别出来,在经过 6、7 视场时未能被成功识别,原因主要是,视场 6 的光照条件严重影响了监控图像质量,行人图像失真严重。视场 7 摄像机视角倾斜角度较大,行人图像发生了明显畸形,导致识别的失败。因此,为改进算法的性能以应对更复杂的真实场景,将开展下一阶段的研究工作:

- 1) 研究光照自适应补偿算法,经过融合,提高网络对过亮、非均匀光照的应对能力;
- 2) 研究行人图像纵向畸形修复方法,解决真实场景监控相机大俯仰角导致行人图像上下半身比例失衡的问题。

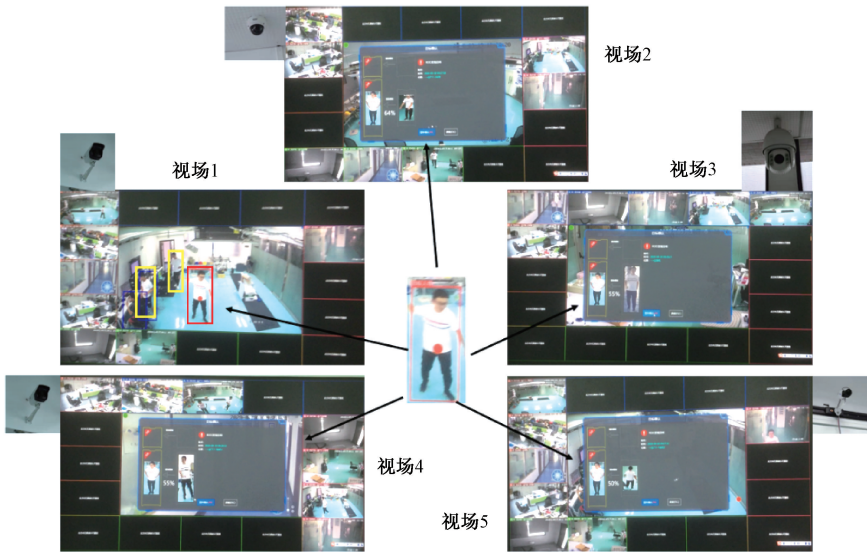


图 6 真实场景测试验证

Fig. 6 Test in real scenarios

3 结 论

为解决因行人图像背景差异大、人体外观相似导致的行人再识别准确率低的问题,提出了一种利用特征融合与多尺度信息的行人重识别方法,同时利用了行人全局特征和局部细节特征,使两者互相补充,还提取了行人全局图像的多尺度特征,提高模型的甄别能力。实验结果表明,该方法在 Rank-1 与 mAP 上都优于对比方法,同时还在真实的室内监控平台测试验证了算法的实时性和有效性,为行人重识别算法的应用提供了一定的参考。但还需解决例如非均匀光照、畸形行人图像校正算法,才能进行实际室外场景的测试验证。

参考文献

[1] 朱阳光,刘瑞敏,黄琼桃. 基于深度神经网络的弱监督信息细粒度图像识别[J]. 电子测量与仪器学报, 2020, 34(2): 115-122.

ZHU Y G, LIU R M, HUANG Q T. Fine-grained image recognition with weakly supervised information based on deep neural network [J]. Journal of Electronic Measurement and Instrumentation, 2020, 34 (2): 115-122.

[2] SUH Y, WANG J, TANG S, et al. Part-aligned bilinear representations for person re-identification [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 402-419.

[3] ZHAO H, TIAN M, SUN S, et al. Spindle net: Person re-identification with human body region guided feature decomposition and fusion[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1077-1085.

[4] WANG G, YANG S, LIU H, et al. High-order information matters: Learning relation and topology for occluded person re-identification[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 6449-6458.

[5] SUN Y, ZHENG L, YANG Y, et al. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline) [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 480-496.

[6] WANG G, YUAN Y, CHEN X, et al. Learning discriminative features with multiple granularities for person re-identification [C]. Proceedings of the 26th ACM International Conference on Multimedia, 2018: 274-282.

[7] ZHANG S, ZHANG L, WANG W, et al. AsNet:

Asymmetrical network for learning rich features in person re-identification [J]. IEEE Signal Processing Letters, 2020, 27: 850-854.

[8] CHEN X, FU C, ZHAO Y, et al. Saliency-guided cascaded suppression network for person re-identification [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 3300-3310.

[9] FAN X, LUO H, ZHANG X, et al. Scpnet: Spatial-channel parallelism network for joint holistic and partial person re-identification [C]. Asian Conference on Computer Vision. Springer, Cham, 2018: 19-34.

[10] PARK H, HAM B. Relation network for person re-identification [C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 11839-11847.

[11] HOU Y, LIAN S, HU H, et al. Part-relation-aware feature fusion network for person re-identification [J]. IEEE Signal Processing Letters, 2021, 28: 743-747.

[12] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.

[13] JADERBERG M, SIMONYAN K, ZISSERMAN A. Spatial transformer networks [J]. Advances in Neural Information Processing Systems, 2015, 28.

[14] PAN X, LUO P, SHI J, et al. Two at once: Enhancing learning and generalization capacities via ibn-net [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 464-479.

[15] 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, 2019.

[16] PENG P, TIAN Y, HUANG Y, et al. Discriminative spatial feature learning for person re-identification [C]. Proceedings of the 28th ACM International Conference on Multimedia, 2020: 274-283.

[17] ZHENG L, SHEN L, TIAN L, et al. Scalable person re-identification: A benchmark [C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1116-1124.

[18] RISTANI E, SOLERA F, ZOU R, et al. Performance measures and a data set for multi-target, multi-camera tracking [C]. European Conference on Computer Vision. Springer, Cham, 2016: 17-35.

[19] HUANG Y, HUANG Y, HU H, et al. Deeply associative two-stage representations learning based on labels interval extension loss and group loss for person re-identification [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2019, 30 (12):

4526-4539.

- [20] LUO H, JIANG W, ZHANG X, et al. Alignedreid++: Dynamically matching local information for person re-identification [J]. Pattern Recognition, 2019, 94: 53-61.
- [21] HUANG H, YANG W, CHEN X, et al. Eanet: Enhancing alignment for cross-domain person re-identification [J]. arXiv preprint arXiv: 1812.11369, 2018.

作者简介



张勃兴, 2019 年于陇东学院获得学士学位, 现为昆明理工大学在读硕士研究生, 主要研究方向为计算机视觉。

E-mail: zhangboxing@stu.kust.edu.cn

Zhang Boxing received his B. Sc. degree from Longdong University in 2019. He is currently a M. Sc. candidate at Kunming University of Science and Technology. His main research interest includes computer vision.



马敬奇, 2016 年于广东工业大学获得硕士学位, 现为广东省科学院智能制造研究所工程师, 主要研究方向为计算机视觉、人机协作。

E-mail: majingqiauto123@126.com

Ma Jingqi received his M. Sc. degree from Guangdong University of Technology in 2016. He is currently an engineer at the Intelligent Manufacturing Institute of Guangdong Academy of Sciences. His main research interests include computer vision and human-machine collaboration.



张寿明 (通信作者), 2009 年于昆明理工大学获得博士学位, 现为昆明理工大学教授, 硕士生导师, 主要研究方向为复杂工业过程控制。

E-mail: 1411834974@qq.com

Zhang Shouming (Corresponding author) received his Ph. D. degree from Kunming University of Science and Technology in 2009. His main research interest includes complex industrial process control.