

DOI:10.16136/j.joel.2022.09.0886

# 基于多粒度特征融合网络的行人重识别

张勃兴<sup>1,2</sup>, 张寿明<sup>1\*</sup>, 钟震宇<sup>2</sup>

(1. 昆明理工大学 信息工程与自动化学院, 云南 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, 广东 广州 510095)

**摘要:**针对复杂环境下行人细节特征不明显、姿态多变等情况造成的行人重识别(person-reidentification, ReID)算法精度不高的问题,提出了一种基于多粒度特征提取与特征融合的ReID网络。首先,在主干网络输入和输出端采用两种粒度的划分方式获取图像的局部特征。其次,引入空间变换网络(spatial transformation network, STN)对输入全局图像进行空间对齐,对局部图像进行特征增强。最后,采用局部特征融合的方式来挖掘特征之间的关联信息,提升模型对相似样本的识别能力。实验结果表明,所提方法在多个数据集上均取得了良好的识别效果。在Market-1501数据集上的平均查准率(mean average precision,  $mAP$ )和首次查准率(Rank-1)分别为84.87%和94.45%,通过和目前主流的ReID算法相比,本文所提方法具有更优的识别效果。

**关键词:**行人重识别(ReID); 姿态; 局部特征; 空间变换网络(STN); 特征融合  
**中图分类号:** TP751.1 **文献标识码:** A **文章编号:** 1005-0086(2022)09-0977-07

## Person re-identification based on multi-granularity feature fusion network

ZHANG Boxing<sup>1,2</sup>, ZHANG Shouming<sup>1\*</sup>, ZHONG Zhengyu<sup>2</sup>

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

**Abstract:** In order to solve the problem of low accuracy of person re-identification (ReID) algorithm in complex environment caused by unclear detail features and changeable attitudes of pedestrians, a ReID network based on multi-granularity feature extraction and feature fusion is proposed. Firstly, two granularity partitioning methods are used to obtain the local features of the image at the input and output ends of the backbone network. Secondly, spatial transformation network (STN) is introduced to align the global image and enhance the local image. Finally, local feature fusion is used to mine the correlation information between features to improve the model's ability to recognize similar samples. Experimental results show that the proposed method achieves good recognition performance on multiple datasets. The mean average precision ( $mAP$ ) and first accuracy (Rank-1) of the market-1501 dataset are 84.87% and 94.45%, respectively. Compared with the current mainstream ReID algorithms, the proposed method has better recognition effect.

**Key words:** person re-identification (ReID); posture; local features; spatial transformation network (STN); feature fusion

## 1 引言

行人重识别(person re-identification, ReID)的

目的是跨越不同的时间、地点、摄像头来匹配特定的行人。其挑战在于如何从背景杂乱、姿态多样、遮挡等情况下的人物图像中提取有区别的特征。

\* E-mail: 1411834974@qq.com

收稿日期: 2021-12-29 修订日期: 2022-01-31

基金项目: 广东省重点领域研发计划项目(2018B010108006, 2020B090925002)和广州市科技计划项目(202007040007)资助项目

当前在该领域,深度学习已经成为了主流的特征提取方式。利用深度网络提取的鉴别特征已基本上取代了手工设定的描述特征,其目的是通过设计更有效的特征提取方式来使网络学习到更有判别力的特征表达<sup>[1]</sup>。其中,早期的全局特征提取方式结构简单、推理速度快,但是其可用于识别的线索太少,不适用于复杂的外部环境。而同时提取全局、局部特征的方法取得了显著效果。由于图像整体结构复杂多变,单一的全局特征无法表示类间差异,识别性能非常有限,但是局部特征包含更多的细粒度信息,一方面,可以使网络关注到显著的细节特征,另一方面,不同部位的细粒度信息结合,让网络提取到的特征更丰富。

ReID局部特征提取方法主要有以下3类:1)将深度特征图横向切分,单独利用每个局部特征来进行距离度量<sup>[2,3]</sup>;2)利用人体语义分割模型将行人图像按照形体结构划分为不同的区域<sup>[4,5]</sup>;3)利用人体骨架关键点检测模型划分局部区域,提取局部特征<sup>[6,7]</sup>。

上述方法当中,水平切分策略是目前最稳定的方法,该类方法通过水平切分深度特征图来获取局部特征,简洁高效,相比于其他两类方法,不需要引入分割模型或关键点检测模型,极大减少了计算量。然而,该类方法往往会受到人体姿态变化以及检测误差的影响,开放环境下,随机抓取的运动行人图像之间姿态差异过大,空间分布并非严格一致,而且普遍存在的检测误差会导致检

测框与实际的人体区域不重合,如果单纯采用这种方式,不可避免会存在局部特征的空间错位,甚至某些局部特征当中会包含大量的背景信息。另一方面,深度特征图往往感受野很大,造成水平切分深度特征图获取的不同局部特征之间存在很强的特征冗余,细节信息不明显<sup>[8]</sup>。除此之外,直接使用人体的局部层次特征,而不考虑各部位之间的关系,容易干扰模型对具有局部相似属性的行人身份的识别<sup>[9]</sup>。

基于此,本文从空间特征对齐、局部特征提取、特征融合入手来优化 ReID 网络,提取更具有判别力的特征。

首先,基于水平切分提取局部特征的策略,在网络的输入和输出端采用不同粒度的划分方式进行局部特征的提取,缓解输出端局部特征的特征冗余,提取到特征更丰富的细节信息。然后,输入图像通过全局和局部空间变换网络(spatial transformation network, STN)<sup>[10]</sup>进行水平对齐和特征增强,给网络输入加入空间不变性,解决行人姿态多变以及检测误差造成的空间错位问题。最后,采用特征融合策略,对孤立的局部特征进行融合,挖掘特征之间的关联关系,提升网络对相似样本的识别能力。

## 2 基于多粒度融合网络的 ReID

本节中,提出了一个针对特征对齐和特征融合的 ReID 网络,网络结构如图 1 所示。下面将分为 5 部分对其进行阐述。

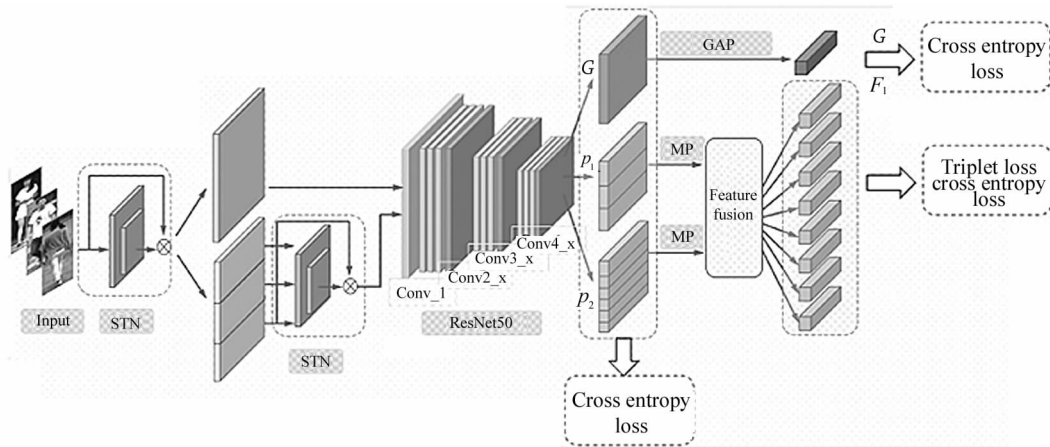


图 1 网络结构

Fig. 1 Network structure

### 2.1 网络结构

给定一个源域  $S=\{X_s, L_s\}$ , 包含  $P$  个行人的  $N_s$  张图像, 其中  $X_s$  代表行人图像,  $L_s \in \{1, 2, \dots, N_{id}\}$  代表图像的标签信息, 通过优化行人深度特征表示

来正确匹配行人。本文使用在 ImageNet 上预训练的 ResNet50<sup>[11]</sup> 网络, 删除最后的池化层和全连接层, 将网络第 4 阶段的卷积步长改为 1, 得到更大、细节信息更丰富的特征图<sup>[12]</sup>。

本文方法采用了两种粒度的局部特征提取方式。假设网络输入表示为 $(H \times W \times C)$ ,首先,输入图像经过STN进行特征对齐。然后,一方面将对齐后的全局图像输入主干网络得到全局特征,再对全局特征图进行水平均匀切分得到6个局部特征。另一方面,将对齐后的图像切分为3块,得到3个局部图像,输入即变为 $H/3 \times W \times C$ ,每个局部图像之后连接3个局部STN,再将局部图像依次输入主干网络得到3个局部特征。最后,将提取到的局部特征进行融合得到融合特征。在训练阶段,全局特征、局部特征以及融合特征分别与图像标签计算损失。在测试阶段,所有特征在通道维度上串联做为行人的特征表示。

## 2.2 局部特征提取

通过对主干网络提取的特征图进行切分获取局部特征是目前常用的方法,但是,主干网络最后一层卷积层覆盖的感受野很大,特征图水平切分以后不同局部特征之间会有很大的重复,造成特征冗余,细节信息不明显。图2为不同局部特征之间存在特征冗余的示例。如果在网络输入端直接对图像进行划分可以避免特征冗余,但是输入图像的空间结构不对齐会导致某些局部特征包含大量的背景信息,特征错位也会干扰网络的识别。故采用STN对行人图像进行特征对齐以后再进行水平切分获得局部图像。本文在局部特征提取时,采用了在主干网络输入和输出端分别进行水平切分获取局部特征的方式,切分方式包含两种粒度,由于输入图像包含更多的空间信息,对输入图像采用粗粒度的划分方式划分为3块,主干网络提取的特征图包含更多的高维语义信息,采用细粒度的划分方式划分为6块。在输入和输出端进行划分并结合两种粒度使网络提取到更加丰富的特征。

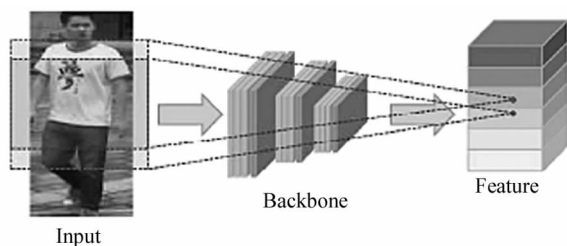


图2 特征冗余示例

Fig. 2 Example of feature redundancy

## 2.3 空间特征对齐与增强

STN会对图像做自适应仿射变换,通过对图像的放缩、平移、旋转3种变换来放大图像前景,达到

特征对齐的目的。其网络组成包含3个部分,分别为定位网络、网格生成器和采样器,网络结构如图3所示。STN通过学习一个仿射变换矩阵来执行逐点的变换:

$$\begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x^{\text{Source}} \\ y^{\text{Source}} \\ 1 \end{bmatrix} = \begin{bmatrix} x^{\text{Target}} \\ y^{\text{Target}} \\ 1 \end{bmatrix}, \quad (1)$$

式中, $(x^{\text{Source}}, y^{\text{Source}})$ 表示原图像像素点, $(x^{\text{Target}}, y^{\text{Target}})$ 表示仿射变换后的图像像素点,系数矩阵 $\theta$ 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 $\theta$ 矩阵。其中图像旋转对特征对齐是无用的,本文参考文献[13]的做法将全局STN系数矩阵的旋转因子置为0。定位网络由一个 $1 \times 1$ 卷积层、批归一化层、激活层和一个最大池化层组成,之后运用全连接层回归得到仿射变换系数矩阵。STN的具体细节参阅文献[10]。

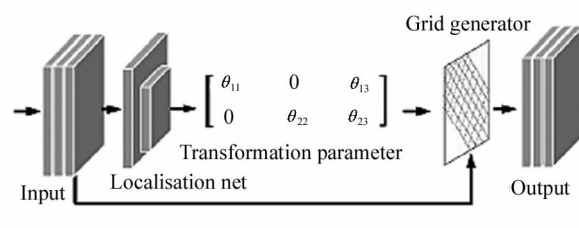


图3 空间变换网络

Fig. 3 STN

本文首先对全局图像采用STN进行特征对齐,然后对输入端水平切分得到的局部图像,采用STN进行特征增强。两方面的使用可以改善局部特征的空间错位,同时让网络对输入图像的空间变换保持鲁棒性,一定程度上缓解行人姿态变化对模型识别的影响。

## 2.4 局部特征融合

局部特征可以显著改善模型适应复杂环境的能力,而且全局范围内的局部特征使网络的关注点更广泛。但是采用水平切分获得的每个局部特征之间是相互孤立的,这种孤立性不利于模型对相似样本的识别。在ReID领域,存在大量的相似样本,例如两个穿衣风格很相似的行人,他们可能身体局部特征特别相似,单纯利用身体部位的孤立特征很可能将他们判定为同一个人,这时就需要联合头部、脚部或者整体的体态特征进行判别。故采用特征融合让网络输出部件之间的关联信息。

图4是本文采用的局部特征融合模块。首先,

将全局图像经过主干网络得到的特征向量水平切分,获得6个相同尺度的局部特征向量,采用最大池化的方式池化为 $1 \times 1 \times 2048$ ,选择其中一个局部特征向量 $p_i (i=1,2,\dots,6)$ ,将其余局部特征向量 $p_j (j \neq i)$ 相加获得包含其余特征所有信息的向量 $g_j$ 。然后,分别利用 $1 \times 1$ 卷积层减小 $p_i$ 和 $g_j$ 的通道数量,获得 $1 \times 1 \times 256$ 的特征向量 $\bar{p}_i$ 和 $\bar{g}_j$ 。将 $\bar{p}_i$ 和 $\bar{g}_j$ 进行通道维度的串联获得 $1 \times 1 \times 512$ 的特征向量 $q_i$ ,采用 $1 \times 1$ 卷积层减小 $q_i$ 的通道数量,获得 $1 \times 1 \times 256$ 的特征向量 $\bar{q}_i$ 。最后,将 $\bar{p}_i$ 和 $\bar{q}_i$ 相加获得最终的融合特征 $f_i$ 。融合过程可以表示为:

$$\bar{g}_j = T[\frac{1}{n} \sum_{j \neq i} p_j], (j=1, \dots, 6), \quad (2)$$

$$f_i = T[C(\bar{p}_i, \bar{g}_j)] + \bar{p}_i, (i, j=1, \dots, 6), \quad (3)$$

式中, $C$ 表示通道维度的串联, $T$ 表示一个包含 $1 \times 1$ 卷积层、BN层和ReLU层的子网络,用来对特征进行降通道操作。本文网络包含两种粒度的局部特征,故式(2)中的系数 $1/n$ 分别设置为 $1/5$ 和 $1/2$ 。

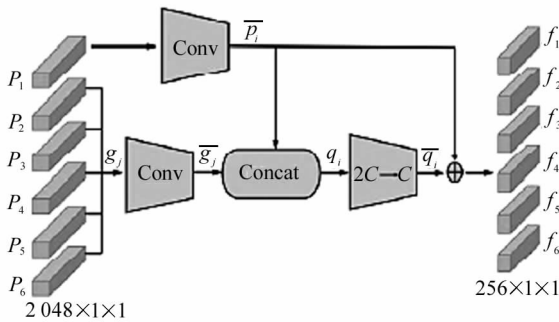


图4 特征融合

Fig. 4 Feature fusion

## 2.5 损失函数

采用联合交叉熵损失 $L_{\text{cross}}^{[14]}$ 和三元组损失 $L_{\text{triplet}}^{[15]}$ 来训练模型。三元组损失用来得到两个特征的差异性度量,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。完整的损失函数如下所示:

$$L = L_{\text{triplet}}^g + L_{\text{cross}}^g + L_{\text{cross}}^x + L_{\text{cross}}^f, \quad (4)$$

式中, $L_{\text{triplet}}^g$ 和 $L_{\text{cross}}^g$ 分别代表使用全局特征计算三元组损失和交叉熵损失, $L_{\text{cross}}^x$ 代表局部特征的交叉熵损失, $L_{\text{cross}}^f$ 代表融合特征的交叉熵损失,交叉熵损失的定义为:

$$L_{\text{cross}} = - \sum_{i=0}^n p(x_i) \lg(q(x_i)), \quad (5)$$

式中, $n$ 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。 $q(x_i)$ 的计

算方式为:

$$q(x_i) = \arg \max_{c \in K} \frac{\exp((w_c^f)^T f_i)}{\sum_{k=1}^K \exp((w_k^f)^T f_i)}, \quad (6)$$

式中, $K$ 是身份标签的数量, $w_c^f$ 代表用作分类的全连接层的权重, $f_i$ 代表模型提取到的行人特征。

采用三元组损失的目标是增加类间距离,减小类内距离,以此来提升检索的准确性。三元组损失定义为:

$$L_{\text{triplet}} = - \sum_{i=1}^P \sum_{a=1}^R [\tau + \max_{p=1, \dots, r} \|F_a^i - F_p^i\|_2 - \min_{\substack{n=1, \dots, r \\ j=1, \dots, p \\ j \neq i}} \|F_a^i - F_n^j\|_2], \quad (7)$$

式中, $P$ 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, $\tau$ 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。

在损失函数设置时,由于采用局部特征进行距离度量可能会造成网络的误判,故仅对全局特征采用三元组损失,如图5所示。

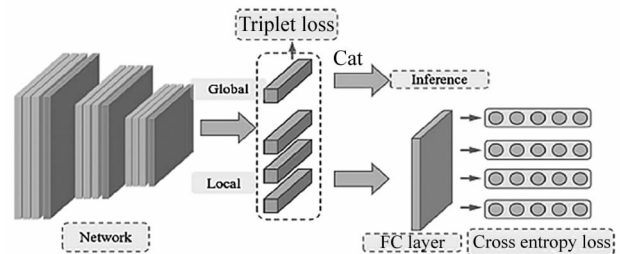


图5 损失函数设置

Fig. 5 Loss function setting

## 3 实验与分析

### 3.1 实验数据和评价指标

本文在ReID常用的两个数据集Market-1501<sup>[16]</sup>和DukeMTMC-ReID<sup>[17]</sup>上进行实验。实验采用首次(Rank-1)成功匹配概率和平均查准率(mean average precision, mAP)来衡量模型的性能。数据集的详细信息如表1所示。

表1 数据集

Tab. 1 Dataset

| Dataset     | Identities | Images | Training set | Test set |
|-------------|------------|--------|--------------|----------|
| Market-1501 | 1 501      | 32 217 | 751          | 750      |
| DukeMTMC    | 1 404      | 36 411 | 702          | 702      |

3.2 参数设置

对输入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  $384 \times 128$ ,每个批次大小为 64。采用随机梯度下降(stochastic gradient descent,SGD)优化器来训练模型,动量为 0.9,权重衰减为  $5 \times 10^{-4}$ ,初始学习率为 0.1,其中 STN 模块的学习率为初始学习率除以 100,每 20 个 epoch 学习率乘以 0.1,总共训练 60 个 epoch。在一个显存为 16 G 的 NVIDIA-Tesla-P100 上训练,大约需要 6 h。

3.3 消融实验

为了验证所提网络和模块的有效性,对各部分进行了消融实验。通过实验发现,当特征对齐、多粒度特征提取、特征融合共同使用时,模型达到了最佳性能。

本文采用 ReID\_PCB 为基线网络,其在 Market-1501 数据集上  $mAP$  为 78.30%,Rank-1 为 92.22%,本文方法在基线方法的基础上的  $mAP$  和 Rank-1 可以分别提升 6.57%和 2.23%。首先,将 STN 用于输入图像进行仿射变换达到特征对齐的目的,图 6 为输入图像经过 STN 之后的变换效果图,可以看出 STN 用于输入图像的确可以起到前景放大和特征对齐的目的。在网络训练时,STN 模块的学习率设置为初始学习率的 1/100,STN 学习率设置过大会导致输入图像变化剧烈,网络很难学习到共性特征。该模块用于 ReID\_PCB 网络,在 Market-1501 数据集上可以使  $mAP$  提升 0.62%,Rank-1 提升 0.21%。此外,本文进一步验证了多尺度特征提取网络和特征融合模块的有效性。采用 STN 增强局部特征同时在输入端进行水平切分的方式可以得到较显著的性能提升,在输入和输出层面的两种划分方式符合不同粒度的人体垂直视点分布,可以增

强局部图像的特征表示,Local\_3 代表在输入端切分得到的 3 个局部特征。同时,特征融合模块通过加权融合局部特征之间的关系来增强对相似样本的识别能力,LF 代表局部特征融合模块,该模块可以使  $mAP$  和 Rank-1 分别提升 1.65%和 0.6%。由表 2 消融实验对比可以看出所提各个模块的有效性。

表 2 消融实验  
Tab. 2 Ablation experiments

| Method                  | Rank-1 | $mAP$  |
|-------------------------|--------|--------|
| Baseline                | 92.22% | 78.30% |
| Baseline+STN            | 92.43% | 78.92% |
| Baseline+STN+Local_3    | 93.85% | 83.22% |
| Baseline+STN+Local_3+LF | 94.45% | 84.87% |

3.4 对比实验与分析

本文与近几年主流方法在 Market-1501 和 DukeMTMC-reID 数据集上进行了对比。这些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,与本文方法类似。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 3 和表 4 所示。

其中,PCB 是本文采用的 Baseline,该方法首次提出了将特征图水平切分提取局部特征,然后单独利用每个局部特征进行行人身份预测,取得了较好的识别效果。PCB+RPP 是对 PCB 网络的改进,通过注意力机制来对局部特征对应的空间分布进行权重分配,进而得到分布更合理的局部特征。但是,两种方法都没有使用全局特征,也没有考虑到特征对齐的影响。SNR 提出了一个即插即用的风格归一化模块,用于改善图像风格变化对识别的影响。AlignedReID 提出了基于最短路径的特征对齐方法,但是该方法并没有提取图像的局部特征,制约了复杂环境下模型的识别能力。HOREID 是一种通过人体骨架关键点检测模型来定位人体局部区域进而进行局部特征提取的方法,该方法获得了更加精细的人体局部特征,但是需要引入人体关键点检测模型,造成网络参数量庞大,而且已有研究已经证明,这种过于精细的划分方式在实际使用时泛化性较差。本文方法从参数量、泛化性等多方面入手来提取更丰富、更具有代表性的特征,展现出了更优的识别效果。



图 6 STN 可视化效果

Fig. 6 STN visualization

表 3 Market-1501 数据集实验结果对比

Tab. 3 Comparison of experimental results of Market-1501 dataset

| Method                      | $r=1$  | $mAP$  |
|-----------------------------|--------|--------|
| PABR <sup>[4]</sup>         | 88.8%  | 73.9%  |
| PCB <sup>[2]</sup>          | 92.3%  | 77.4%  |
| PCB+RPP <sup>[2]</sup>      | 93.8%  | 81.6%  |
| SNR <sup>[18]</sup>         | 94.4%  | 84.7%  |
| AlignedReID <sup>[19]</sup> | 92.6%  | 82.3%  |
| HOReID <sup>[6]</sup>       | 94.2%  | 84.9%  |
| Ours                        | 94.45% | 84.87% |

表 4 DukeMTMC-reID 数据集实验结果对比

Tab. 4 Comparison of experimental results of DukeMTMC-reID dataset

| Method                      | $r=1$ | $mAP$ |
|-----------------------------|-------|-------|
| PABR <sup>[4]</sup>         | 82.1% | 64.2% |
| PCB <sup>[2]</sup>          | 81.9% | 65.3% |
| PCB+RPP <sup>[2]</sup>      | 83.3% | 69.2% |
| SNR <sup>[18]</sup>         | 84.4% | 72.9% |
| AlignedReID <sup>[19]</sup> | 82.1% | 69.7% |
| HOReID <sup>[6]</sup>       | 86.9% | 75.6% |
| Ours                        | 87.5% | 76.7% |

### 3.5 识别效果可视化

本文选取了两对外观较相似的行人样本进行了可视化验证,如图 7 所示,1、2 号和 3、4 号分别是两对身穿白色和蓝色衣服的样本。两对样本在行人色彩外观以及穿着上十分相似,但在行人之间的细节信息存在明显差异。从图中可以看出,两对行人最



图 7 识别效果可视化

Fig. 7 Recognition effect visualization

明显的不同在于携带物的差别,在人体外观过于相似时,行人的背包、手提物以及身体上的配饰是模型识别最重要的线索。可视化结果展示了所提网络对显著细节信息强大的提取能力以及对相似样本的识别能力。

## 4 结 论

本文提出了一种基于多粒度信息与特征融合的 ReID 网络,重点分析了当前局部特征提取策略存在的弊端,并根据相关问题提出了有效的解决方法。具体地,在网络输出端和输入端对特征图和输入图像分别进行多粒度切分,同时利用 STN 对全局行人图像进行对齐,对局部图像进行了特征增强。此外,采用局部特征融合方式挖掘各个孤立局部特征之间的关联关系,避免水平切分造成的信息损失,增强网络对相似样本的识别能力。通过实验对比和可视化验证,证明了所提方法的有效性以及对复杂行人样本强大的特征提取能力。所提方法也为该领域今后的研究工作提供了新的思路和参考,同时,也可以运用到其他相关的计算机视觉领域当中用于目标识别与难样本特征挖掘。

## 参考文献:

- [1] ZHENG Z, ZHENG L, YANG Y. Pedestrian alignment network for large-scale person re-identification [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 29(10): 3037-3045.
- [2] SUN Y, ZHENG L, YANG Y, et al. Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline) [C]// European Conference on Computer Vision (ECCV), September 8-14, 2018, Munich, Germany. Berlin: Springer, 2018: 480-496.
- [3] WANG G, YUAN Y, CHEN X, et al. Learning discriminative features with multiple granularities for person re-identification [C]// 26th ACM International Conference on Multimedia, October 22-26, 2018, Seoul, Korea. New York: ACM, 2018: 274-282.
- [4] SUH Y, WANG J, TANG S, et al. Part-aligned bilinear representations for person re-identification [C]// European Conference on Computer Vision (ECCV), September 8-14, 2018, Munich, Germany. Berlin: Springer, 2018: 402-419.
- [5] HUANG H, YANG W, CHEN X, et al. EANet: enhancing alignment for cross-domain person re-identification [EB/OL]. (2018-12-29) [2021-12-29]. <https://arxiv.org/abs/1812.11369>.

- [6] WANG G N, YANG S, LIU H Y, et al. High-order information matters: learning relation and topology for occluded person re-identification [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 13-19, 2020, Seattle, WA, USA. New York: IEEE, 2020: 6448-6457.
- [7] ZHAO H Y, TIAN M Q, SUN S Y, 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, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 907-915.
- [8] WANG G, YUAN Y, LI J, et al. Receptive multi-granularity representation for person re-identification [J]. IEEE Transactions on Image Processing, 2020, 29: 6096-6109.
- [9] PARK H, HAM B. Relation network for person re-identification [C]//AAAI Conference on Artificial Intelligence, February 7-12, 2020, New York, NY, USA. Menlo Park: AAAI, 2020: 11839-11847.
- [10] JADERBERG M, SIMONYAN K, ZISSERMAN A. Spatial transformer networks [C]//Advances in Neural Information Processing Systems (NIPS 2015), December 7-12, Montreal, Canada. New York: Curran Associates, Inc., 2015: 2017-2025.
- [11] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition [C]//IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE, 2016: 770-778.
- [12] LUO H, GU Y Z, LIAO X Y, et al. Bag of tricks and a strong baseline for deep person re-identification [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, June 16-17, 2019, Long Beach, CA, USA. New York: IEEE, 2019: 1487-1495.
- [13] PENG P, TIAN Y, HUANG Y, et al. Discriminative spatial feature learning for person re-identification [C]//28th ACM International Conference on Multimedia, October 12-16, 2020, Seattle, WA, USA. New York: ACM, 2020: 274-283.
- [14] ZHANG Z, SABUNCU M. Generalized cross entropy loss for training deep neural networks with noisy labels [C]//Advances in Neural Information Processing Systems (NIPS 2018), December 3-8, 2018, Montreal, Canada. New York: Curran Associates, 2018: 8792-8802.
- [15] HERMANS A, BEYER L, LEIBE B. In defense of the triplet loss for person re-identification [EB/OL]. (2017-11-21) [2021-12-29]. <https://arxiv.org/abs/1703.07737>.
- [16] ZHENG L, SHEN L, TIAN L, et al. Scalable person re-identification: a benchmark [C]//IEEE International Conference on Computer Vision, December 7-13, 2016, Santiago, Chile. New York: IEEE, 2015: 1116-1124.
- [17] ZHENG Z, ZHENG L, YANG Y. Unlabeled samples generated by gan improve the person re-identification baseline in vitro [C]//IEEE International Conference on Computer Vision, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE, 2017: 3754-3762.
- [18] JIN X, LAN C L, ZENG W J, et al. Style normalization and restitution for generalizable person re-identification [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 13-19, 2020, Seattle, WA, USA. New York: IEEE, 2020: 3140-3149.
- [19] ZHANG X, LUO H, FAN X, et al. AlignedReID: surpassing human-level performance in person re-identification [EB/OL]. (2017-11-22) [2021-12-29]. <https://arxiv.org/abs/1711.08184>.

#### 作者简介:

张寿明 (1966—),男,博士,教授,硕士生导师,主要从事复杂工业控制方面的研究。