

中图法分类号: TP391.41 文献标识码: A 文章编号: 1006-8961(2023)05-1360-12

论文引用格式: Zhang H Y, Wang X Y and Peng X W. 2023. Foreground segmentation-relevant multi-feature fusion person re-identification. Journal of Image and Graphics, 28(05): 1360-1371 (张红颖, 王徐泳, 彭晓雯. 2023. 结合前景分割的多特征融合行人重识别. 中国图象图形学报, 28(05): 1360-1371) [DOI: 10.11834/jig.220683]

结合前景分割的多特征融合行人重识别

张红颖*, 王徐泳, 彭晓雯

中国民航大学电子信息与自动化学院, 天津 300300

摘要: 目的 行人重识别任务中, 同一行人在不同图像中的背景差异会导致识别准确率下降, 出现误识别的现象。针对此问题, 提出了一种结合前景分割的多特征融合行人重识别方法。**方法** 首先构建前景分割模块, 提取图像的前景, 并通过前景分割损失, 保持前景图像的平滑性和完整性; 然后提出了注意力共享策略和多尺度非局部运算方法, 将图像中的全局特征与局部特征、高维特征与低维特征结合, 实现不同特征之间的优势互补; 最后通过多损失函数对网络模型进行训练优化。**结果** 在3个公开数据集 Market-1501、DukeMTMC-reID (Duke multi-tracking multi-camera re-identification) 和 MSMT17 (multi-scene multi-time person ReID dataset) 上进行了消融实验和对比实验, 并以首位命中率 (rank-1 accuracy, Rank-1) 和平均精度均值 (mean average precision, mAP) 作为评价指标。实验结果显示, 在引入前景分割和多特征融合方法时, 网络的识别准确率均有一定提升。本文方法在 Market-1501、DukeMTMC-reID 和 MSMT17 数据集上 Rank-1 和 mAP 分别为 96.8% 和 91.5%、91.5% 和 82.3% 以及 83.9% 和 63.8%, 相比于对比算法, 本文方法具有较大优势。**结论** 本文提出的结合前景分割的多特征融合方法, 在提取前景的同时, 综合了不同尺度和不同粒度图像特征, 相较于已有模型, 提高了识别准确率。同时, 前景分割模块消除了无用背景, 缓解了背景差异导致的误识别现象, 使行人重识别模型的实用性得到加强, 在面对实际背景情况时, 也能有较好的识别效果。

关键词: 前景分割; 语义分割; 行人重识别 (ReID); 特征融合; 注意力机制

Foreground segmentation-relevant multi-feature fusion person re-identification

Zhang Hongying*, Wang Xuyong, Peng Xiaowen

College of Electronic Information and Automation, Civil Aviation University of China, Tianjin 300300, China

Abstract: Objective Person re-identification (ReID) is a computer vision-based cross camera recognition technology to target a pedestrian-specific in an image or video sequence. To obtain more discriminative features and achieve high accuracy, the deep learning based method of ReID is focused on processing personal features in recent years. The whole pedestrian image is often as the sample for the ReID model and each pixel feature is as the basis for recognition in the image. However, ReID, as a cross-camera recognition task, is required to deploy a wide range and number of camera locations, which will inevitably lead to background variations in pedestrian images. In other words, the heterogeneity of images-

收稿日期: 2022-07-07; 修回日期: 2022-09-28; 预印本日期: 2022-10-05

* 通信作者: 张红颖 carole_zhang0716@163.com

基金项目: 国家重点研发计划资助 (2018YFB1601200); 中国民航大学研究生科研创新资助项目 (2021YJS026)

Supported by: National Key R&D Program of China (2018YFB1601200); Graduate Research Innovation Grant Program of Civil Aviation University of China (2021YJS026)

captured is challenged for the interrelations between background similarity and its identity for both of single and multiple cameras. Therefore, it is necessary to rich background information to the pedestrian similarity metric in the ReID model. To resolve this problem, we develop a foreground segmentation-based multi-branch joint person re-identification method in terms of the residual network 50 (ResNet50). **Method** An integration of foreground segmentation and ReID method are employed. First, as the input for feature extraction, the foreground area of the pedestrian images is extracted by the foreground segmentation module. Then, to achieve mutual benefits between different features, the global features in the image are combined with local features and high-dimensional features with low-dimensional features using a multi-grain feature-guided branch and multi-scale feature fusion branch. For the foreground segmentation module, an attention mechanism is used to improve mask region-based convolutional neural network (Mask R-CNN). And, the foreground segmentation loss function is adopted to optimize the feature information loss derived of rough segmentation of the foreground. For multi-grain feature branch, the convolutional block attention module (CBAM) is improved initially in terms of a three-branch attention structure. The information-interacted between two dimensions is based on adding new branches between the spatial and channel dimensions. Furthermore, an attention-sharing strategy is implemented. To improve the effectiveness of feature extraction and avoid feature chunking-derived missing information, the attention information is shared in coarse-grained branches with fine-grained branches, which can yield global features to guide the extraction of local features simply. For multi-scale feature fusion branch, the features at different stages backbone network feature extraction are used as the input of multi-scale fusion straightforward. Additionally, the pyramid attention structure is used as well to get the feature information before fusion. Next, for the fusion module, to synthesize the global information and alleviate the loss of feature information, a non-local algorithm is illustrated in multiscale for feature fusion. Finally, as a joint loss function, the loss-incorporated is in relevance with foreground segmentation, TriHard, and Softmax. And, it is used to train the network for optimization further. **Result** The comparative analysis is based on 3 publicly available datasets (Market-1501, Duke multi-tracking multi-camera re-identification (DukeMTMC-reID) and multi-scene multi-time person ReID dataset (MSMT17)). The metrics-related evaluations consist of rank- n accuracy (Rank- n) and mean average precision (mAP). For the Market-1501 dataset, the Rank-1 and mAP can be reached 96.8% and 91.5% of each, which is a 0.6% improvement in Rank-1 and a 1% improvement in mAP compared to attention pyramid network (APNet-C); For the DukeMTMC-reID dataset, the Rank-1 and mAP can be reached to 91.5% and 82.3% of each, as well as an improvement of 1.1% in Rank-1 and an improvement of 0.8% in mAP compared to the model APNet-C; For the MSMT17 dataset, the Rank-1 and mAP can be reached to 83.9% and 63.8% of each, which is increased by 0.2% in Rank-1 and 0.3% in mAP compared to APNet-C; **Conclusion** We facilitate a foreground segmentation based multi-branch joint model. It can be focused on foreground extraction-based in accordance with an integration of multiple scales and grains image features. At the same time, the foreground segmentation module can wipe out ineffective background and alleviate the background-differentiated false recognition more.

Key words: foreground segmentation; semantic segmentation; person re-identification (ReID); feature fusion; attention mechanism

0 引言

行人重识别(person re-identification, ReID)是利用计算机视觉技术判断图像或视频序列中是否存在特定行人的跨镜头识别技术,对嫌疑人追踪和失踪人员检索等有重大意义(Deng等, 2021; Liu等, 2022)。基于深度学习的行人重识别方法能够获取更具判别性的特征,成为行人重识别领域的主要研究方法(沈庆等, 2020; 李擎等, 2022)。但由于背景变化、局部遮挡和光照变化等问题,如何从行人图像

中学习更有效的特征从而实现高精度的行人重识别仍然是一个挑战(郑鑫等, 2020)。

在早期,大多数方法将整个行人图像作为ReID模型的输入,将图像中的每一个像素特征都作为识别的依据。然而,在最近的研究中,Tian等人(2018)发现,如果行人图像中的背景发生变化,ReID模型将会出现过拟合、可迁移能力下降等问题。Huang等人(2019)通过对训练和测试数据的处理发现,如果训练和测试数据的背景差异很大,ReID方法的鲁棒性会受到极大影响,在实际应用时性能会出现明显下降。刘志刚等人(2022)指出行人图像中前景和

背景之前存在隐含关系,当同一行人的背景变化较大时,ReID模型可能会出现误识别的现象。由此可见,行人图像中背景变化产生的差异会对ReID的识别结果产生影响。

为了解决背景变化导致的误识别问题,一些方法将前景分割技术与ReID结合,利用前景分割提取行人图像中的前景部分。龚毓秀(2020)将提取的前景图像与原始图像拼接的结果作为ReID网络的输入,提高了识别的准确性。Gong等人(2021)通过前景分割方法,直接删除行人图像中的背景部分,建立了一个新的数据集。但是,直接使用前景分割方法删除背景会存在以下缺点:1)利用身体掩膜进行处理可能会破坏原始图像的结构化信息和平滑度;2)前景图像可能会丢失身体的某些部位,从而丢失部分细节信息;3)虽然身体掩膜可用于滤除背景干扰,但可能会丢失一些能够作为判别信息的背景线索。所以,为了缓解直接删除背景导致的信息丢失,本文在现有前景分割的方法中,加入了注意力机制和多损失融合的思想,通过注意力机制和损失函数的双重作用约束前景分割模块,在提取前景部分的同时,保持原始图像的结构信息和平滑度。

此外,除了通过前景分割方法提取前景,还有一些方法通过提取不同特性的特征缓解背景变化导致的误识别问题。李艳凤等人(2020)在使用监督网络提取有用的前景特征的同时,构建了一个多池融合网络来描述全局特征与局部特征,并实现多个细粒度的特征表达。宋晓茹等人(2022)通过对不同尺度的特征进行融合,在保留高维特征中的语义信息的同时,结合了低维特征中更多的位置与细节信息。Gong等人(2021)在直接删除背景的基础上,利用联合对齐信息驱动网络从原始图像中学习对齐特征,加强局部和全局特征的结合。种衍文等人(2022)在实现图像对齐的基础上,进行多粒度特征的融合,避免小粒度信息的丢失。以上方法通过特征融合,提取了更具鉴别性的行人特征,取得了较好的识别效果,但存在以下缺陷:1)很多方法直接将高维特征进行反卷积的结果作为低维特征并进行融合,会忽视反卷积过程中出现的特征信息变化问题;2)通过不同分支分别提取全局特征与局部特征,导致局部特征在提取过程中会因为图像分块出现特征丢失,而且不同分支的特征提取能力存在差别,直接将两种特征融合将会影响特征融合的有效性;3)大多数方

法没有同时关注多尺度特征和多粒度特征,丢失了部分具有鉴别性的信息。针对这些问题,本文首先在多粒度特征提取中,提出了一种注意力共享策略,通过全局特征引导局部特征的提取,提高局部特征的有效性,使其能更好地与全局特征融合。其次,在多尺度特征提取中,将特征提取过程各个阶段的结果作为多尺度特征融合的输入,并提出了一种多尺度非局部运算方法进行尺度融合,避免特征丢失。

综上所述,本文将前景分割与行人重识别相结合,并通过特征融合获取更具有判别力的特征,提高网络的特征表示能力,从而缓解背景变化的影响。本文主要有以下贡献:1)通过对图像中语义信息的分析,提出了一种基于注意力机制的前景分割模块,并设计了一种前景分割损失函数,优化了前景分割模块对前景的提取能力,同时保留部分有用的背景信息;2)在多尺度特征融合时,提出了一种多尺度非局部运算方法进行尺度融合,缓解了由高维特征在反卷积过程中以及卷积下采样操作中的特征丢失问题;3)在多粒度特征提取时,提出了一种基于注意力共享策略的多粒度特征引导方法,通过注意力机制引导局部特征的提取,不仅利用注意力机制聚焦网络关注点的特性,提高特征的有效性,而且将全局与局部联系起来,得到了对图像中行人的判别性表达。

1 结合前景分割的多特征融合网络

结合前景分割方法和特征融合对ReID任务的有效性,本文提出了一种结合前景分割的多特征融合行人重识别方法(multi-feature fusion person re-identification method combined with foreground segmentation, MFSNet),网络结构如图1所示,包括骨干网络、前景分割模块(foreground segmentation module, FSM)、多粒度特征引导分支(multi-granularity feature-guided branch, MFGB)和多尺度特征融合分支(multi-scale feature fusion branch, MFFB)4个部分。

1.1 网络结构

MFSNet网络首先在主干网络进行特征提取之前,通过前景分割模块提取前景,然后将提取结果作为特征提取的输入。在特征提取中,主干网络采用ResNet50(residual network 50),并且将经过ResNet50后的网络划分成3个分支。第1个分支沿

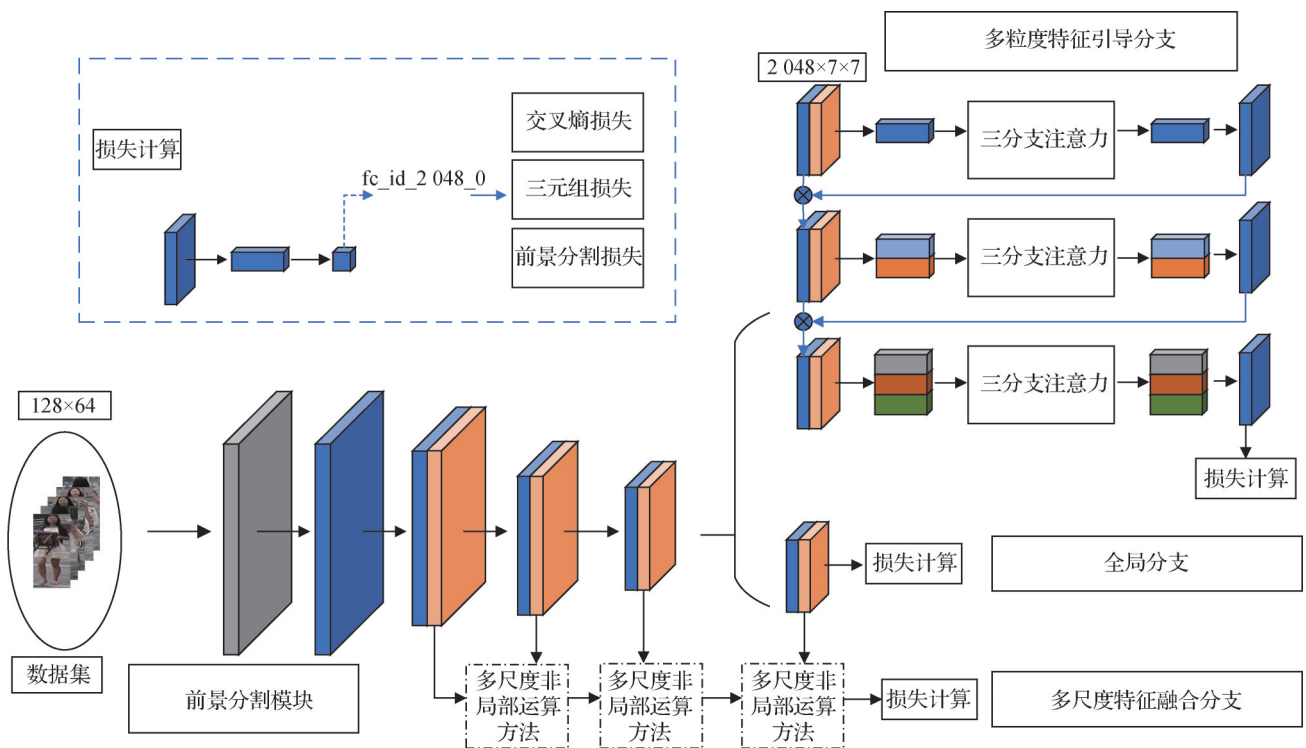


图1 结合前景语义分割的多特征融合网络

Fig. 1 Multi-branch joint network based on foreground semantic segmentation

用原始 ResNet50 相同的网络形式,提取行人图像的全局特征。对于第2个分支,为了实现全局特征引导局部特征,MFCB 将 ResNet50 第5个模块 block res_conv5_1 及之后的部分分离出来,并且去除了全连接层。随后将分离出来的部分作为 MFCB 模块的输入,实现全局特征对局部特征的引导提取,并获取融合特征。对于第3个分支,为了缓解多尺度融合过程中的特征丢失,MFCB 采用了逐级分离融合结构,在特征提取过程中分离出尺度信息并进行融合,即在 ResNet50 的第2,3,4,5层进行分离特征并进行融合。同时,为了进一步提升特征的有效性,在特征融合之前先使用金字塔注意力聚焦网络的特征提取能力,然后通过特征融合得到最后的多尺度融合特征。

1.2 前景分割模块

由于跨摄像头的特殊性,ReID 任务中不可避免地存在背景变化的情况,当相同身份 ID 行人的背景差异较大时,会出现误识别的现象。因此,本文提出一种前景分割模块提取图像中的有效前景,该模块分为掩膜提取和掩膜融合两部分,首先利用掩膜提取网络生成二值掩膜图像,然后利用哈达玛积将原始图像与其掩膜的融合,并将融合结果作为行人重

识别特征提取网络的输入。前景分割模块的结构如图2所示。

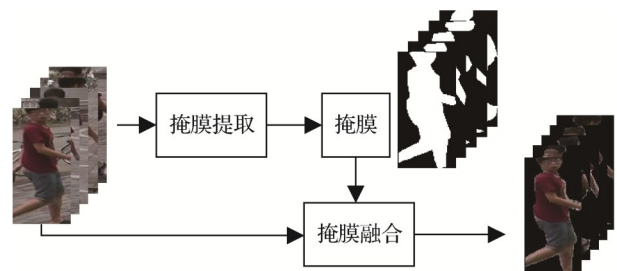


图2 前景分割模块

Fig. 2 Foreground segmentation module

在现有的掩膜提取网络中,Mask R-CNN(region based convolutional neural network)方法(He 等, 2017)采用特征金字塔(feature pyramid network, FPN)结构,并且通过 ROI Align(region of interest align)解决了 Faster R-CNN(faster region proposal convolutional neural network)中 ROI Pooling(region of interest pooling)两次取整产生误差的问题,能够较好地分割出有用信息,但其无法保持原始图像的结构信息和平滑度,会造成行人部分的细节信息丢失。因此,本文在 Mask R-CNN 网络的基础上,将 ReID 损失与原 Mask 损失结合,提出了前景分割损失函数,

并通过注意力机制使前景分割网络在提取前景时更关注与 ReID 相关的有效信息,避免硬性地从图像中的前后景分割,造成信息丢失,改进后的网络结构如图 3 所示。

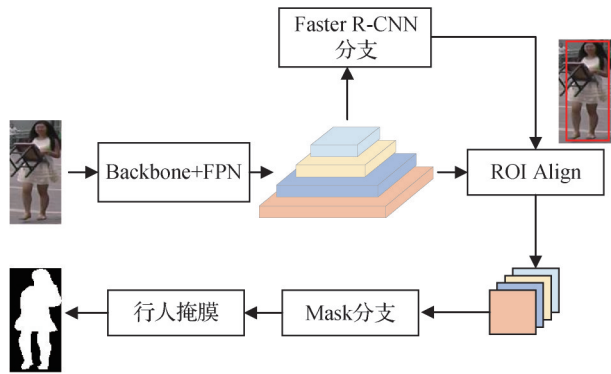


图 3 行人掩膜提取网络结构图
Fig. 3 Pedestrian mask extraction network structure diagram

1.3 多粒度特征引导分支

在行人重识别任务中,全局特征与局部特征的融合能够使网络联合图像全局信息和各个粒度的局部信息,增强图像中信息的利用率和有效性,从而提高网络的鲁棒性。为了更有效地利用局部信息,减少特征丢失,本文提出了多粒度特征引导分支 (MFGB),结构如图 4 所示。

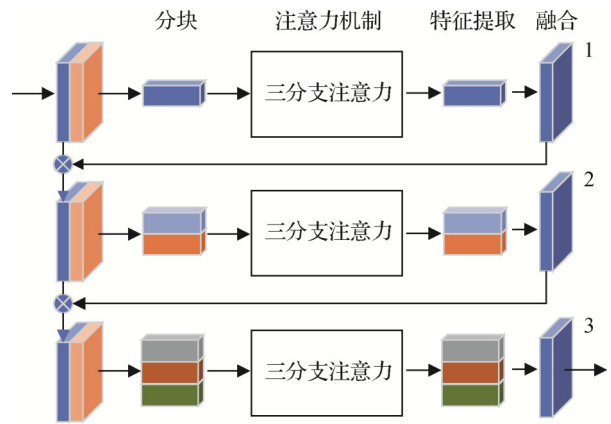


图 4 多粒度特征引导分支
Fig. 4 Multi-granularity feature-guided branch

多粒度特征引导分支主要由 3 个分支堆叠构成。每一个分支都采用分块—注意力机制—特征提取—融合的模式。其中,分块是对特征图进行切分处理,分块的数目越多,越能使网络关注到局部的特征信息。注意力机制和特征提取则是在聚焦网络的特征提取能力的同时,利用如图 5 所示的特征提取

残差块(每个残差块都有 3 层结构)提取更深层次的特征。最后,通过融合将局部的特征转化成与原输出特征大小一致,在得到每一分支的最终特征的同时,为实现特征引导打下基础。

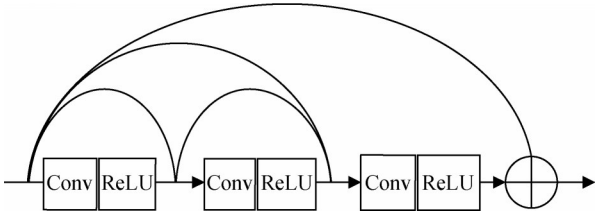


图 5 特征提取残差块
Fig. 5 Feature extraction residual blocks

对于其中的注意力机制,本文在 CBAM (convolutional block attention module) 的基础上新增了两个维度交互分支,提出了三分支注意力机制,其结构如图 6 所示。在新分支 1 和新分支 2 中,首先对特征进行重排列操作,将原本输入的特征(结构是 $C \times H \times W$)通过 Permute 函数进行通道维度 C 和空间维度 H 、 W 的交换,得到 $H \times C \times W$ 和 $W \times H \times C$ 的特征结构形式。然后对重排列之后的特征进行池化、卷积和正则化操作,进一步加深维度之间的融合,并重排列至原始 $C \times H \times W$ 形式(其中, Z 变换代表同时使用平均池化和最大池化)。最后将 3 个分支的输出特征进行融合处理,得到注意力信息。这样做的好处是,空间信息和通道信息两个维度之间的信息进行了交互,缓解了 CBAM 中将空间注意力和通道注意

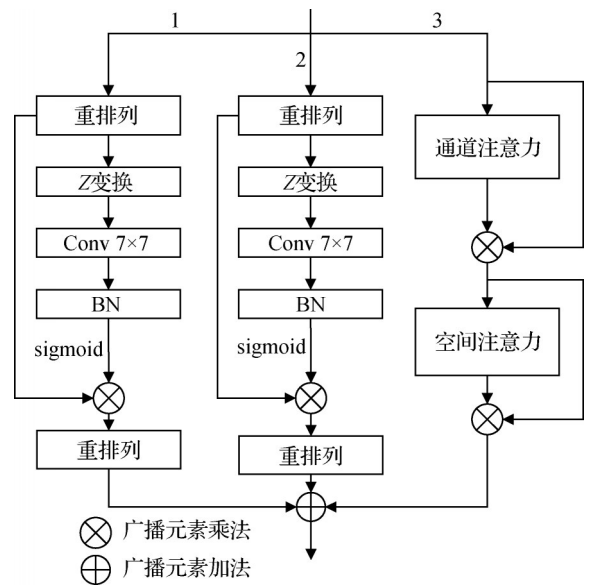


图 6 三分支注意力
Fig. 6 Three-branch attention

力进行串并联造成的信息丢失(张红颖和包雯静, 2022; Misra等, 2021)。

在3个分支对不同粒度特征进行提取的基础上, 本文提出的MFGB中的特征引导过程主要由残差堆叠策略和注意力共享策略实现。具体地, 在MFGB的第2分支和第3分支中, 分别使用两种策略将前一分支中的关键性信息(包括粗粒度的特征和注意力关注点信息)结合到细粒度特征的提取过程中, 达到通过粗粒度的信息引导细粒度特征提取的目的。

残差堆叠策略是使用残差结构将3个分支堆叠结合, 其结构为

$$Y_i = G(X_{i-1} + X_i) \quad (1)$$

式中, X_{i-1} 表示前一分支的特征, X_i 表示原始输入, Y_i 表示现分支的特征, $G(x)$ 表示特征引导过程, 包含了注意力机制。

注意力共享策略是通过权重共享, 将粗粒度分支中的注意力信息共享到细粒度分支中, 其结构为

$$Y'_i = H(X'_i) \quad (2)$$

式中, X'_i 表示输入特征, Y'_i 表示经过注意力机制处理后的特征, $H(x)$ 表示前一分支的注意力机制。

式(1)表示将前一分支的输出与本分支的输入结合作为本分支的新输入; 式(2)表示将本分支的注意力替换为上一分支的注意力, 即将粗粒度分支的注意力权重作为本分支注意力机制的初始权重。将式(1)(2)结合可得, MFGB的特征引导机制主要实现过程为

$$Y_i = G'(H(X_{i-1} + X_i)) \quad (3)$$

式中, $G'(x)$ 表示细粒度分支中的特征引导过程, 不包含注意力机制。

由式(3)可以看到, 首先, 残差堆叠策略将更粗粒度的信息与原输入结合, 作为细粒度特征提取的输入。然后, 注意力共享策略将粗粒度特征中注意力机制的关注点信息引入细粒度特征中, 通过粗粒度特征约束细粒度特征的提取。由此, 通过残差堆叠策略和注意力共享策略能够以粗粒度特征为蓝本进行细粒度特征的提取, 一方面能够用粗粒度特征弥补细粒度特征提取过程中的信息丢失, 另一方面能够用注意力机制的关注点信息约束细粒度特征的提取, 提高细粒度特征的有效性。

1.4 多尺度特征融合分支

在行人重识别任务中, 多尺度特征中的低维特征大多由高维特征经过上采样得到, 但是上采样得

到的低维特征不一定能真实地代表原始图像中的特征, 会出现特征信息丢失。针对此问题, 本文提出了多尺度特征融合分支(MFFB), 将骨干网络特征提取中的特征直接作为多尺度特征融合的输入, 减少特征信息丢失, 并在融合模块前使用金字塔注意力结构对特征进行处理, 提高特征的表达能力, 其主要结构如图7所示。

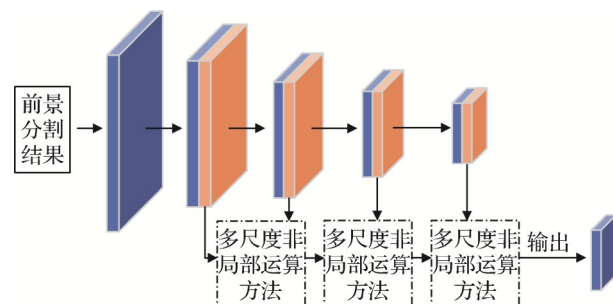


图7 多尺度特征融合分支

Fig. 7 Multi-scale feature fusion branch

在图7所示的结构中, 本文利用单尺度非局部运算考虑全局信息的思想(Wang等, 2018), 提出了一种多尺度非局部运算方法, 直接将不同尺度的信息进行融合, 以获得更有效的特征, 其结构如图8所示。

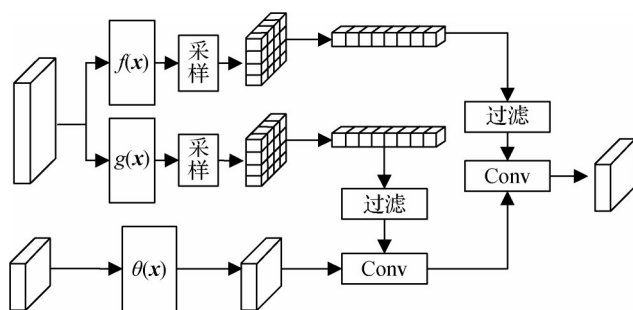


图8 融合模块

Fig. 8 Fusion module

多尺度非局部运算方法在原方法的基础上, 首先对输入 x (大小为 $H \times W$)进行缩放, 获得一个描述映射 k (大小为 $\frac{H}{s} \times \frac{W}{s}$), 然后将不同尺度的相关性通过 k 转换到同一尺度, 并由 x^i 和区域描述子 k^j 表示, 计算为

$$y^i = \frac{1}{\sigma(x, k)} \sum_{s \in [1, 2]} \sum_{j \in k} \phi(x^i, k^j) \theta(k^j) \quad (4)$$

式中, i 和 j 分别是输入 x 和输出 y 的索引, s 是尺度因子; ϕ 是系数函数, 用来计算两个特征之间的相似性; θ 表示特征转换函数; $\sigma(x)$ 代表归一化参数。

经过尺度因子的缩放操作后,不同尺度的信息达到了统一尺度。因此,在系数函数 ϕ 的选取中,仍使用嵌入高斯函数,计算为

$$\phi(\mathbf{x}^i, \mathbf{k}^j) = \exp(f(\mathbf{x}^i)^T g(\mathbf{k}^j)) \quad (5)$$

最后,将式(5)代入式(4),得到最终多尺度非局部运算方法的表达式,具体为

$$\mathbf{y}^i = \frac{1}{\sigma(\mathbf{x})} \sum_{s \in [1,2]} \sum_{j \in k} \exp(f(\mathbf{x}^i)^T g(\mathbf{k}^j)) \theta(\mathbf{k}^j) \quad (6)$$

式中, $f(\mathbf{x}^i) = \mathbf{W}_f \mathbf{x}^i$, $g(\mathbf{k}^j) = \mathbf{W}_g \mathbf{k}^j$ 。 θ 使用原非局部方法中的线性嵌入, $\theta = \mathbf{W}_\theta \mathbf{k}^j$ 。 $\sigma(\mathbf{x}) = \sum_{j,j'} f(\mathbf{x}^i, \mathbf{k}^j)$ 。

1.5 联合损失函数

为了更好地优化网络的训练过程,本文设计了前景分割损失,并联合难样本采样三元组损失和softmax损失对网络进行约束训练。

前景分割网络能够得到图像中的行人掩膜,但是由于前景分割任务和行人重识别任务的差异性,直接将前景分割的结果作为输入并不一定能取得良好效果。为了有机地将前景分割与行人重识别结合,本文在Mask R-CNN网络mask损失 L_M 的基础上,通过行人重识别任务的识别结果约束掩膜预测的结果,提出了前景分割损失函数,计算为

$$L_M(P) = A \times \frac{\exp(\mathbf{W}_{y_i}^T \mathbf{x}_i + b_{y_i})}{\sum_{j=1}^n \exp(\mathbf{W}_j^T \mathbf{x}_i + b_j)} + B \quad (7)$$

式中, A 和 B 分别表示原始损失 L_M 中的有效特征对应损失和非掩膜部分的特征对应损失。对于 A 的权重系数, n 为类别数, $\mathbf{x}_i$ 为第 y_i 类中第 i 个样本的特征向量, $\mathbf{W}$ 为权值, b 为偏置量。 A 和 B 计算为

$$A = -\frac{1}{N} \sum_{i=1}^N P_i^* \times \log(\sigma(P_i)) \quad (8)$$

$$B = -\frac{1}{N} \sum_{i=1}^N (1 - P_i^*) \times \log(1 - \sigma(P_i)) \quad (9)$$

式中, N 表示一个预测掩膜中的像素总数, σ 表示sigmoid函数, P_i 表示掩膜中的一个像素, P_i^* 则为其相应的真值像素(ground truth)。

难样本采样三元组损失是在三元组损失的基础上,将简单随机抽样改为一种基于训练批量的在线难样本采样方法,计算为

$$L_{Th} = \frac{\sum_{a \in batch} [\max_{p \in C} d_{ap}(\eta) - \min_{n \in D} d_{an}(\eta) + \alpha]}{P \times K} \quad (10)$$

式中, α 为预设的阈值参数,表示距离间隔; P 为随机

挑选的行人个数, K 为每个行人对应的照片数, $P \times K$ 的值表示每批次训练的样本数量; a, p, n 分别为目标样本、正样本和负样本; $batch$ 表示每个训练步骤中选择的样本数; C 和 D 分别对应相同身份和不同身份的图像集; $d_{ap}(\eta)$ 和 $d_{an}(\eta)$ 分别表示目标样本与正负样本之间的距离。

softmax损失是使用了softmax函数的交叉熵损失,计算为

$$L_s = -\frac{1}{N} \sum_{i=1}^N \lg \frac{\exp(\mathbf{W}_{y_i}^T \mathbf{x}_i + b_{y_i})}{\sum_{j=1}^n \exp(\mathbf{W}_j^T \mathbf{x}_i + b_j)} \quad (11)$$

式中, $\mathbf{W}$ 为权值参数, b 为偏置量, n 为训练的类别数, $\mathbf{x}_i$ 为第 y_i 类中对应于第 i 个样本的特征向量, N 为每一批次训练样本的总数。

为了提高前景分割模块与行人重识别的结合效果,本文采用多损失函数联合策略,将以上3种损失函数结合,作为总体的行人重识别网络的训练损失,其表达式为

$$L = L_M + L_{Th} + L_s \quad (12)$$

2 实验结果及分析

2.1 数据集设置及评价标准

为了验证本文方法的效果,使用ReID常用的数据集Market1501(Zheng等,2015)、DukeMTMC-reID(Duke multi-tracking multi-camera re-identification)(Ristani等,2016)和MSMT17(multi-scene multi-time person reid dataset)(Wei等,2018)进行实验,并与相关的行人重识别算法进行对比。每个数据集划分为训练集和测试集两部分。Market1501分为训练图像12 936幅、测试图像19 732幅,共包含6个摄像头拍摄的1 501个行人的图像。DukeMTMC-reID分为训练集16 522幅、测试集17 661幅和query集2 228幅,共包括1 404个行人在8个摄像头下的36 411幅行人图像。MSMT17数据集分为训练集32 621幅,测试集93 820幅,共包含15个摄像头拍摄的4 101个行人的126 441幅图像。

同时,本文使用Rank-1和平均精度均值(mean average precision, mAP)两个行人重识别任务中常用的评价指标作为对比的依据。其中,Rank-1是指被查询图像中行人与识别结果中第1位图像中行人是同一行人的概率。mAP的值则是一个求和平均数,

代表了整体的识别水平,计算为

$$mAP = \frac{\sum_{k=0}^C AP_k}{C} \tag{13}$$

式中, $\sum_{k=0}^C AP_k$ 表示将每一个类别的平均精度求和, C 表示总类别数。

2.2 实验环境及参数设置

实验基于一台 Ubuntu18.04 的戴尔工作站,其 CPU 参数为 Intel®Xeon(R)Silver4112CPU@2.6 GHz, 内存为 64 GB,并使用双 GeForce RTX 2080Ti 显卡加速运算。算法使用 Pytorch1.7.1 深度学习框架。网络实现分为前景分割和行人重识别两部分。前景分割部分采用预训练过的 Mask R-CNN 进行行人掩膜提取;行人重识别部分的骨干网络采用 ImageNet 预训练过的 ResNet50 作为基础模型,其他部分的权重的初始化主要通过 Kaiming 初始化完成。模型总共训练 160 个 epoch,采用 Adam 优化器对模型的训练进行优化,初始学习率设置为 0.0004,并在每经过 40 个 epoch 后缩小为原来的 1/10,直到网络收敛。

2.3 消融实验

本文提出的 MFSNet 首先利用前景分割模块获取行人图像中的前景(行人)部分,然后通过两个分支分别融合图像中的局部与全局信息、高维与低维信息,获得分辨力更强的行人特征表达。因此,消融实验主要是验证前景分割模块、多粒度特征引导分支和多尺度特征融合分支的有效性,实验结果如表 1 所示。其中,Baseline 表示只包含图 1 网络结构中的

ResNet50 主干网络和全局分支,直接通过残差网络的特征提取结果进行行人重识别;Baseline+MFGB、Baseline+MFFB 和 Baseline+FSM 分别是在 Baseline 的基础上单独加入 3 个结构,用以证明所提出网络结构的有效性。Baseline+MFGB+MFFB 和 Baseline+MFGB+MFFB+FSM 是在多粒度特征引导的基础上逐步添加网络结构,用以证明 3 个模块之间的耦合效果。

从表 1 可以看出,在 3 个数据集中,本文所提出的 MFGB 和 MFFB 结构均能对 Baseline 起到一定的提升效果。原因在于,在 MFGB 和 MFFB 中,本文考虑到了不同特性的特征内蕴含的独特信息,并通过构建注意力共享策略和多尺度非局部运算方法,对不同特性的特征采取了针对性的方式,减少特征丢失,实现特征之间相互补充。

同时,在加入 FSM 模块后,网络的精度也获得一定提升。原因在于,前景分割方法提取前景后,能够降低同一行人在不同摄像头拍摄的图像之间的背景差异,缓解 Baseline 中同一行人背景差异导致的特征相似度下降问题。并且,随着背景被消除,同一摄像头下,不同行人之间背景相似性被消除,其特征相似度下降,也对识别效果的提升起到了促进作用。

2.4 对比实验

为了验证本文方法的有效性,在 Market-1501 和 DukeMTMC-reID 数据集上进行对比实验,对比方法包括基于语义信息前景分割的 ReID 方法 SSP-ReID (saliency-semantic parsing ReID) (Quispe 和 Pedrini,

表 1 在 Market-1501、DukeMTMC-reID 和 MSMT17 上的消融实验结果
Table 1 Results of ablation experiments on Market-1501, DukeMTMC-reID and MSMT17 dataset

方法	Market-1501		DukeMTMC-reID		MSMT17	
	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
	/%					
Baseline	78.1	92	65.9	81.5	52.1	76.9
Baseline + MFGB	83.4	93.6	74.4	85.1	56.5	77.6
Baseline + MFFB	86.2	95.1	78.2	88.5	57.4	80.1
Baseline + FSM	85.8	94.2	76.3	86.1	55.6	78.2
Baseline + MFGB + MFFB	89.6	96.1	80.7	90.2	61.8	82.6
Baseline + MFGB + MFFB + FSM(MFSNet)	91.2	96.8	82.3	91.5	63.8	83.9

注:加粗字体表示各列最优结果。

2018) 和 FS-MENet (foreground fegmentation and multi-loss ensemble network)(龚毓秀, 2020); 基于多特征融合的 ReID 方法 CAL(counterfactual attention learning method)(Rao 等, 2021)、FPB(feature pyramid branch)(Zhang 等, 2021)、MGN(multiple granularity network)(Wang 等, 2018)和 PyrNet(aggregating deep pyramidal representations)(Martinel 等, 2019); 基于注意力机制的 ReID 方法 APNet-C(attention pyramid network)(Chen 等, 2021)。实验结果如表 2 和表 3 所示。

表 2 在 Market-1501 上与相关方法的比较
Table 2 Comparison with mainstream related method on Market-1501 dataset

方法	mAP	/%
		Rank-1
FS-MENet(龚毓秀, 2020)	88.6	93.2
SSP-ReID(RR)(Quispe 和 Pedrini, 2018)	90.8	93.7
APNet-C(Chen 等, 2021)	90.5	96.2
FPB(Zhang 等, 2021)	90.6	96.1
PyrNet(multi-shot)(Martinel 等, 2019)	94.0	96.1
MGN(Wang 等, 2018)	86.9	95.7
CAL(Rao 等, 2021)	89.5	95.5
本文(MFSNet)	91.2	96.8

注:加粗字体表示各列最优结果。

表 3 在 DukeMTMC-reID 上与相关方法的比较
Table 3 Comparison with related method on DukeMTMC-reID dataset

方法	mAP	/%
		Rank-1
FS-MENet(龚毓秀, 2020)	80.1	86.7
SSP-ReID(RR)(Quispe 和 Pedrini, 2018)	83.7	86.4
APNet-C(Chen 等, 2021)	81.5	90.4
FPB(Zhang 等, 2021)	82.9	91.2
PyrNet(multi-shot)(Martinel 等, 2019)	74.0	87.1
MGN(Wang 等, 2018)	78.4	88.7
CAL(Rao 等, 2021)	80.5	90.0
本文(MFSNet)	82.3	91.5

注:加粗字体表示各列最优结果。

同时,为了进一步验证所提模型对背景变化问题的效果,在 MSMT17 数据集上与部分主流方法进行比较。在对比方法中新增了 MPN(multi-task part-aware network)(Ding 等, 2022)、ABD-Net(attentive but diverse network)(Chen 等, 2019)、Deep Miner(a deep and multi-branch network which mines rich and diverse features)(Benzine 等, 2021)等 3 种与本文结构相似的方法。实验结果如表 4 所示。

表 4 在 MSMT17 上与相关方法的比较
Table 4 Comparison with related methods on MSMT17 dataset

方法	mAP	/%
		Rank-1
MPN(w/o RK)(Ding 等, 2022)	62.7	83.5
APNet-C(Chen 等, 2021)	63.5	83.7
ABD-Net(Chen 等, 2019)	60.8	82.3
Deep Miner(w/o RK)(Benzine 等, 2021)	67.3	85.6
MGN(Wang 等, 2018)	52.1	76.9
CAL(Rao 等, 2021)	64.0	84.2
本文(MFSNet)	63.8	83.9

注:w/o 表示 without,加粗字体表示各列最优结果。

从表 2—表 4 可以看出,在 Market1501、DukeMTMC-reID 和 MSMT17 数据集上,本文方法均优于目前已有的一部分算法。原因在于,本文方法在现有方法的基础上加入了不同尺度不同粒度特征的信息,增加了特征信息量。同时,通过模型结构的设计,提高了行人特征信息的利用率,减少了特征丢失,并在此基础上加入了前景分割部分,通过提取前景缓解了背景信息对 ReID 模型识别精度的影响。

2.5 可视化结果

2.5.1 前景分割可视化结果

为了更好地展示本文方法在提取前景上的效果,在 Market1501 数据集上随机选择行人图像进行前景分割,并进行可视化。图 9 展示了身份 ID 为 0041 和 0001 的两位行人前景提取效果图。其中,第 1 行是数据集中的原始图像,第 2 行是 Mask R-CNN 网络生成的行人掩膜图像,第 3 行是提取前景后的图像。

可以看到,本文使用的前景分割模块能够较好地实现前景部分的提取。同时,能够保留部分行人



图9 Market-1501数据集上部分行人图像前景提取效果图

Fig. 9 Result of partial pedestrian image foreground extraction on the Market-1501 dataset ((a) ID 0041; (b) ID 0001)

轮廓边缘的图像信息,保持了前景图像中行人部分图像的完整性和平滑性。

2.5.2 结合前景分割的多特征融合行人重识别方法识别结果可视化

为了更好地展示本文方法的识别效果,在 Market1501 数据集上随机选择行人进行识别结果可视化,如图 10 所示。因为前景分割模块是加入到特征提取网络之中的,所以最终的结果呈现都有背景。

3 结 论

行人重识别作为一种跨镜头任务,摄像头的数量多,覆盖的范围大,同一行人在不同摄像头拍摄到的图像中会存在较大的背景差异,使得同一行人特



图 10 在 Market-1501 数据集上部分查询结果可视化

Fig. 10 The visualization of partial query results on the Market-1501 dataset

征相似度下降。对此,本文提出了一种结合前景分割的多特征融合行人重识别方法。首先通过损失函数的设计和注意力机制的应用增强前景分割与行人重识别任务的耦合效果,在提取前景同时保持前景图像的平滑性和完整性;然后通过多粒度特征引导分支和多尺度特征融合分支提取更有鉴别性的特征;最后通过联合损失函数促进网络的训练优化。在3个主流数据集上的实验结果表明,本文方法在一定程度上缓解了背景变化导致的误识别问题,提高了识别效果。与其他ReID方法相比,本文方法通过前景分割的方式删除掉了行人图像中差异较大的背景部分,提高了同一行人特征的相似性;并联合行人图像中的多尺度、多粒度特征得到了更具判别力的行人特征表达。但是,本文方法直接将前景分割网络作为模块添加到ReID任务中,极大增加了模型的参数量和训练时间,会影响ReID任务的时效性。并且,前景分割与行人重识别作为两种任务,如何更好地结合是一个巨大挑战。下一步将着眼于二者的结合,一方面通过网络结构的优化,在保持网络模型轻量化的同时保持较好的识别效果;另一方面通过对结合两种任务的研究以取得更好的耦合效果。

参考文献(References)

- Benzine A, El Amine Seddik M and Desmarais J. 2021. Deep miner: a deep and multi-branch network which mines rich and diverse features for person re-identification [EB/OL]. [2022-07-07]. <https://arxiv.org/pdf/2102.09321.pdf>
- Chen G Y, Gu T P, Lu J W, Bao J A and Zhou J. 2021. Person re-identification via attention pyramid. *IEEE Transactions on Image Processing*, 30: 7663-7676 [DOI: 10.1109/TIP.2021.3107211]
- Chen T L, Ding S J, Xie J Y, Yuan Y, Chen W Y, Yang Y, Ren Z and Wang Z Y. 2019. ABD-Net: attentive but diverse person re-identification//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. Seoul, Korea (South): IEEE: 8350-8360 [DOI: 10.1109/ICCV.2019.00844]
- Chong Y W, Zhang C, Feng W Q and Pan S M. 2022. Person re-identification based on multi-level and generated alignment network. *Journal of Huazhong University of Science and Technology (Natural Science Edition)*, 50(4): 64-70 (种衍文, 章榔, 冯文强, 潘少明. 2022. 基于多粒度生成对齐网络的行人重识别. *华中科技大学学报(自然科学版)*, 50(4): 64-70) [DOI: 10.13245/j.hust.220411]
- Deng J H, Hao Y, Khokhar M S, Kumar R, Cai J Y, Kumar J and Aftab M U. 2021. Trends in vehicle re-identification past, present, and future: a comprehensive review. *Mathematics*, 9(24): #3162 [DOI: 10.3390/math9243162]
- Ding C X, Wang K, Wang P F and Tao D C. 2022. Multi-task learning with coarse priors for robust part-aware person re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(3): 1474-1488 [DOI: 10.1109/TPAMI.2020.3024900]
- Gong Y X. 2020. Research on Person Re-identification Method Based on Foreground Segmentation and Multi-loss Ensemble. Hefei: Hefei University of Technology (龚毓秀. 2020. 基于前景分割与多损失融合的行人重识别方法研究. 合肥: 合肥工业大学)
- Gong Y X, Wang R G, Yang J, Xue L X and Hu M. 2021. Person re-identification with global-local background_bias net. *Journal of Visual Communication and Image Representation*, 74: #102961 [DOI: 10.1016/j.jvcir.2020.102961]
- He K M, Gkioxari G, Dollár P and Girshick R. 2017. Mask R-CNN//*Proceedings of 2017 IEEE International Conference on Computer Vision*. Venice, Italy: IEEE: 2980-2988 [DOI: 10.1109/ICCV.2017.322]
- Huang Y, Wu Q, Xu J S and Zhong Y. 2019. SBSGAN: suppression of inter-domain background shift for person re-identification//*Proceedings of 2019 IEEE/CVF International Conference on Computer Vision*. Seoul, Korea (South): IEEE: 9526-9535 [DOI: 10.1109/ICCV.2019.00962]
- Li Q, Hu W Y, Li J Y, Liu Y and Li M X. 2022. A survey of person re-identification based on deep learning. *Chinese Journal of Engineering*, 44(5): 920-932 (李擎, 胡伟阳, 李江昀, 刘艳, 李梦璇. 2022. 基于深度学习的行人重识别方法综述. *工程科学学报*, 44(5): 920-932) [DOI: 10.13374/j.issn2095-9389.2020.12.22.004]
- Li Y F, Zhang B, Sun J, Chen H J and Zhu J L. 2020. Cross-dataset person re-identification method based on multi-pool fusion and background elimination network. *Journal on Communications*, 41(10): 70-79 (李艳凤, 张斌, 孙嘉, 陈后金, 朱锦雷. 2020. 基于多池化融合与背景消除网络的跨数据集行人再识别方法. *通信学报*, 41(10): 70-79) [DOI: 10.11959/j.issn.1000-436x.2020181]
- Liu M J, Zhao J Q, Zhou Y, Zhu H C, Yao R and Chen Y. 2022. Survey for person re-identification based on coarse-to-fine feature learning. *Multimedia Tools and Applications*, 81(15): 21939-21973 [DOI: 10.1007/s11042-022-12510-1]
- Liu Z G, Huang Z, Xie D J, Tian F and Li T Y. 2022. Person re-identification for suppressing background interference. *Journal of Computer-Aided Design and Computer Graphics*, 34(4): 563-569 (刘志刚, 黄朝, 谢东军, 田枫, 李婷玉. 2022. 抑制背景干扰的行人重识别方法. *计算机辅助设计与图形学学报*, 34(4): 563-569)
- Martinel N, Foresti G L and Micheloni C. 2019. Aggregating deep pyramidal representations for person re-identification//*Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Rec-*

- ognition Workshops (CVPRW). Long Beach, USA: IEEE: 1544-1554 [DOI: 10.1109/CVPRW.2019.00196]
- Misra D, Nalamada T, Arasanipalai A U and Hou Q B. 2021. Rotate to attend: convolutional triplet attention module//Proceedings of 2021 IEEE Winter Conference on Applications of Computer Vision. Waikoloa, USA: IEEE: 3138-3147 [DOI: 10.1109/WACV48630.2021.00318]
- Quispe R and Pedrini H. 2019. Improved person re-identification based on saliency and semantic parsing with deep neural network models. *Image and Vision Computing*, 92: #103809 [DOI: 10.1016/j.imavis.2019.07.009]
- Rao Y M, Chen G Y, Lu J W and Zhou J. 2021. Counterfactual attention learning for fine-grained visual categorization and re-identification//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 1005-1014 [DOI: 10.1109/ICCV48922.2021.00106]
- Ristani E, Solera F, Zou R, Cucchiara R and Tomasi C. 2016. Performance measures and a data set for multi-target, multi-camera tracking//Proceedings of 2016 European Conference on Computer Vision. Amsterdam, the Netherlands: Springer: 17-35 [DOI: 10.1007/978-3-319-48881-3_2]
- Shen Q, Tian C, Wang J B, Jiao S S and Du L. 2020. Multi-resolution feature attention fusion method for person re-identification. *Journal of Image and Graphics*, 25(5): 946-955 (沈庆, 田畅, 王家宝, 焦珊珊, 杜麟. 2020. 多分辨率特征注意力融合行人再识别. *中国图象图形学报*, 25(5): 946-955) [DOI: 10.11834/jig.19023]
- Song X R, Yang J, Gao S, Chen C B and Song S. 2022. Person re-identification method based on attention mechanism and multi-scale feature fusion. *Science Technology and Engineering*, 22(4): 1526-1533 (宋晓茹, 杨佳, 高嵩, 陈超波, 宋爽. 2022. 基于注意力机制与多尺度特征融合的行人重识别方法. *科学技术与工程*, 22(4): 1526-1533) [DOI: 10.3969/j.issn.1671-1815.2022.04.029]
- Tian M Q, Yi S, Li H S, Li S H, Zhang X S, Shi J P, Yan J J and Wang X G. 2018. Eliminating background-bias for robust person re-identification//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 5794-5803 [DOI: 10.1109/CVPR.2018.00607]
- Wang G S, Yuan Y F, Chen X, Li J W and Zhou X. 2018. Learning discriminative features with multiple granularities for person re-identification//Proceedings of the 26th ACM International Conference on Multimedia. Seoul, Korea (South): ACM: 274-282 [DOI: 10.1145/3240508.3240552]
- Wei L H, Zhang S L, Gao W and Tian Q. 2018. Person transfer GAN to bridge domain gap for person re-identification//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 79-88 [DOI: 10.1109/CVPR.2018.00016]
- Zhang H Y and Bao W J. 2022. The cross-view gait recognition analysis based on generative adversarial networks derived of self-attention mechanism. *Journal of Image and Graphics*, 27(4): 1097-1109 (张红颖, 包雯静. 2022. 融合自注意力机制的生成对抗网络跨视角步态识别. *中国图象图形学报*, 27(4): 1097-1109) [DOI: 10.11834/jig.200482]
- Zhang S F, Yin Z R, Wu X, Wang K, Zhou Q and Kang B. 2021. FPB: feature pyramid branch for person re-identification [EB/OL]. [2022-07-07]. <https://arxiv.org/pdf/2108.01901.pdf>
- Zheng L, Shen L Y, Tian L, Wang S J, Wang J D and Tian Q. 2015. Scalable person re-identification: a benchmark//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE: 1116-1124 [DOI: 10.1109/ICCV.2015.133]
- Zheng X, Lin L, Ye M, Wang L and He C L. 2020. Improving person re-identification by attention and multi-attributes. *Journal of Image and Graphics*, 25(5): 936-945 (郑鑫, 林兰, 叶茂, 王丽, 贺春林. 2020. 结合注意力机制和多属性分类的行人再识别. *中国图象图形学报*, 25(5): 936-945) [DOI: 10.11834/jig.19018]

作者简介

张红颖,女,教授,主要研究方向为机场智能信息处理、图像处理与计算机视觉。E-mail: carole_zhang0716@163.com

王徐泳,男,硕士研究生,主要研究方向为计算机视觉、行人重识别和图像超分辨率。E-mail: thresyong@163.com

彭晓雯,女,硕士研究生,主要研究方向为计算机视觉和目标跟踪。E-mail: 1280074374@qq.com