

# 基于注意力机制的多级特征级联行人重识别

张正一, 丁建伟\*, 魏慧雯, 萧晓彤

中国人民公安大学信息安全学院, 北京 100038

**摘要** 针对现有行人重识别算法因细节信息关注不足导致的判别性不强问题, 提出了一种基于注意力机制的多级特征级联行人重识别算法。首先, 通过级联不同深度的特征实现对不同层级特征的充分利用, 以补充高层级特征中的细节信息。然后, 引入一对互补的注意力机制模块, 以融合特征图中相似的像素及通道, 弥补特征中的空间位置信息, 提高特征的判别性。最后, 在 Market-1501、DukeMTMC-ReID、CUHK03 数据集上进行了大量实验。结果表明, 本算法的识别精度和平均准确率优于大多数当前的主流算法。

**关键词** 机器视觉; 行人重识别; 多级特征级联; 通道注意力机制; 空间注意力机制

中图分类号 TP391.4

文献标志码 A

doi: 10.3788/LOP202158.2215003

## Multi-Level Features Cascade for Person Re-Identification Based on Attention Mechanism

Zhang Zhengyi, Ding Jianwei\*, Wei Huiwen, Xiao Xiaotong

College of Information and Cyber Security, People's Public Security University of China, Beijing 100038, China

**Abstract** To address the problem of limited discriminative power in existing person re-identification algorithms owing to the loss of details, a multi-level features cascade for person re-identification algorithm based on attention mechanism is proposed in this paper. First, the algorithm is used to cascade features at different depths to fully utilize the features of various levels and replenish detailed information in high-level feature maps. Then, a pair of complementary attention mechanism modules is introduced to integrate similar pixels and channels in the high-level feature maps, compensate for the space location information in the features, and improve the discriminativeness of the features. Finally, extensive experiments are performed on Market-1501, DukeMTMC-ReID, and CUHK03 data sets. Results show that the algorithm shows better recognition and average accuracies than most current mainstream algorithms.

**Key words** machine vision; person re-identification; multi-level features cascade; channel attention mechanism; position attention mechanism

**OCIS codes** 150.0155; 100.4996; 100.4999

## 1 引言

行人重识别是计算机视觉领域的研究热点, 指在不同的摄像头视角和地点匹配同一个感兴趣的人。行人重识别在智能视频监控系统、智慧安防和人机交互中有着很多的应用, 逐渐成为维护公共安全和社会稳定的重要手段, 如搜索犯罪嫌疑人、跨境追踪和轨

迹分析。虽然行人重识别技术已经得到了快速发展, 但在开放的视频监控场景下, 受摄像头视角变化、光照不均、穿着相近以及遮挡等因素的影响<sup>[1]</sup>, 同一行人在不同摄像头下存在较大的外观差异或不同行人之间的外观差异十分微小, 导致提取的特征判别性不高, 从而影响了行人重识别的准确性和时效性。

在深度学习技术出现之前, 传统的行人重识别

收稿日期: 2020-11-30; 修回日期: 2020-12-23; 录用日期: 2021-01-21

基金项目: 国家重点研发计划(A19808)、中央高校基本科研业务费专项资金(2020JKF301)

通信作者: \*jwding@foxmail.com

算法主要利用手工设计的特征进行识别,如颜色特征、纹理特征、梯度直方图(HOG)特征<sup>[2]</sup>。这些手工设计的特征描述能力十分有限,存在很大的局限性,难以适应复杂场景下大数据量的行人重识别任务。随着深度学习技术的应用,利用卷积神经网络(CNN)从原始图像中自动提取复杂特征的算法极大促进了行人重识别的发展。Wu 等<sup>[3]</sup>提出了 PersonNet,利用小尺度卷积块提取图像中的细粒度特征。Zheng 等<sup>[4]</sup>提出了一个类别区分性嵌入(IDE)表示模型,将行人重识别的训练过程看作一个多种类的分类问题。Qian 等<sup>[5]</sup>提出了一个多尺度深度表征学习模型,可自适应调整尺度,获取图像中不同尺度且具有判别性的特征。张涛等<sup>[6]</sup>提出了一种基于全局特征的行人重识别改进算法,通过改进原始的三元组损失函数,减小类内距离并增加类间距离进行训练。上述算法大多通过提取行人的全局特征进行重识别,在遇到遮挡、视角变化等导致局部信息缺失的场景时,识别精度和平均准确率会下降。为了解决上述问题,补充提取特征中的局部信息,提高算法的鲁棒性,人们采用全局特征与局部特征相结合的方式对行人重识别。Zhao 等<sup>[7]</sup>将行人分解成多个局部区域,根据每个局部区域之间的相似度进行行人重识别。Suh 等<sup>[8]</sup>提出一个基于身体部位对齐的双流网络模型,将人体姿势表示为局部特征图,并将其与外观整体特征图结合,增强了网络的特征表述能力。徐龙壮等<sup>[9]</sup>提出了一种多尺度卷积特征融合算法,解决了全局特征判别信息不足的问题。但这些算法大多利用神经网络中的高层特征图进行训练、学习,无法充分挖掘浅层特征图中包含的细节信息,在遇到相近的行人图像时,由于特征

的判别性不强,导致算法的识别精度不高。

为解决特征提取过程中细节信息关注不足导致的判别性不强问题,本文提出了一种基于注意力机制的多级特征级联(AMFC)算法。该算法结合多级特征融合和注意力机制模块,将不同层级的特征进行级联(Cascade),充分利用卷积层中不同深度的特征,有效补充了用于分类或距离度量的高层级特征中的细节信息。将通道注意模块(CAM)<sup>[10]</sup>和空间注意模块(PAM)<sup>[11]</sup>进行级联,先提高神经网络在训练过程中对各个身体部位位置信息的识别能力。然后根据得到的位置信息选择更具判别性的身体特征,丰富特征空间信息的同时提高网络获取具有判别性特征的能力。最后在仅使用 Softmax 损失函数的情况下优化网络模型的学习过程。在数据集 Market-1501<sup>[12]</sup>、DukeMTMC-ReID<sup>[13]</sup>、CUHK03<sup>[14]</sup>上的实验结果表明,本算法具有鲁棒性强、判别性高等优点,且能取得较高的精确度和平均准确率。

## 2 基本原理

行人重识别算法一般分为特征提取和距离度量<sup>[5-11]</sup>两部分,本算法的网络结构如图 1 所示。其中,特征提取网络(FEN)用于特征提取,由骨干网络(Backbone)、多级特征级联网络(MFC-Net)、位置和通道注意力机制(PCAM)模块组成。首先用 Backbone 提取特征,将提取到的不同层级特征送入 MFC-Net 中进行级联操作,实现细节信息的补充。然后将级联得到的特征通过 PCAM,增强网络提取判别性特征的能力。最后将不同分支的特征经过平均池化(Avg Pooling)操作后计算损失,进行网络优化学习。

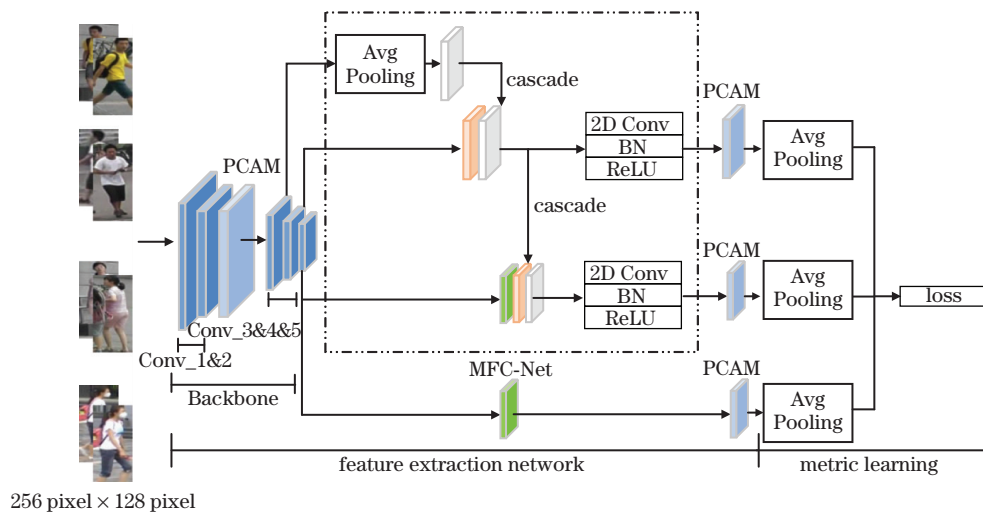


图 1 AMFC 算法的结构

Fig. 1 Structure of the AMFC algorithm

特征提取网络中的 Backbone 采用残差网络 (ResNet50), 并在第 2 个卷积层 (Conv\_2) 后插入了 PCAM, 之后分成了四个分支, 其中三个分支分别用于提取 ResNet50 最后 3 个卷积层 (Conv\_3&4&5) 的特征图, 另一个分支用于保存最后一层得到的特征图。Backbone 得到的特征图经过 MFC-Net 实现不同层级特征图的级联。MFC-Net 和 PCAM 主要对获取的特征进行处理, 补充特征中的细节信息, 使网络关注行人图像中更具判别性的空间区域和身体部位。最后将级联的特征图与最后一层的特征图通过 PCAM, 得到注意力机制加权的特征图。距离度量时先对最终的特征图进行平均池化, 降低参数量、整合全局空间信息, 得到最初的重识别结果, 计算损失并优化网络学习。测试时对实验结果进行重排序 (RK)<sup>[15]</sup>, 即基于  $k$  个最近邻编码的方式对实验结果中的图像进行二次排序, 提升网络的识别精度。

## 2.1 多级特征级联网络

CNN 层数越深, 提取到的特征抽样性就越高。高层级卷积块获取的特征图感受野较大, 其语义表达能力也较强, 常用于最终的分类或相似度度量。但这类特征图的分辨率较低, 缺乏空间几何等细节信息。因此, 当遇到相近的行人图像时, 仅使用高层

级特征图进行行人重识别的精确度较低。此时, 行人图像中的细节信息就成了重识别的关键, 如鞋子的颜色、衣服上的花纹等微小的细节信息。低层级获取到的特征图分辨率较高, 包含很多细节信息, 但感受野较小、噪声较多, 特征的判别性不高, 不能用于最终的分类或相似度度量。MFC-Net 将低层级特征图与高层级特征图进行级联, 以补充高层级特征图中的细节信息, 增加最终用于分类或相似度度量特征向量的判别性, 提高行人重识别算法的鲁棒性。

特征提取网络中第 3 个卷积层特征图的尺寸为  $28 \text{ pixel} \times 28 \text{ pixel}$ , 首先, 对该层特征图进行平均池化操作, 将该图像的尺寸变为  $16 \text{ pixel} \times 8 \text{ pixel}$ , 与后面卷积层得到的特征图尺寸相同。然后, 将其与第 4 个卷积层的特征图在通道维度上进行叠加。叠加后将得到的特征图与第 5 个卷积层的特征图进行叠加。由于仅叠加特征图不能实现信息的补充与通道的融合, 因此, 用  $1 \times 1$  二维卷积 (2D Conv)、正则化层 (BN) 以及线性整流单元 (ReLU) 激活函数融合各层的信息, 并将其恢复到原始各层的通道数, 实现不同层特征图信息的相互补充及融合。MFC-Net 的结构如图 2 所示。

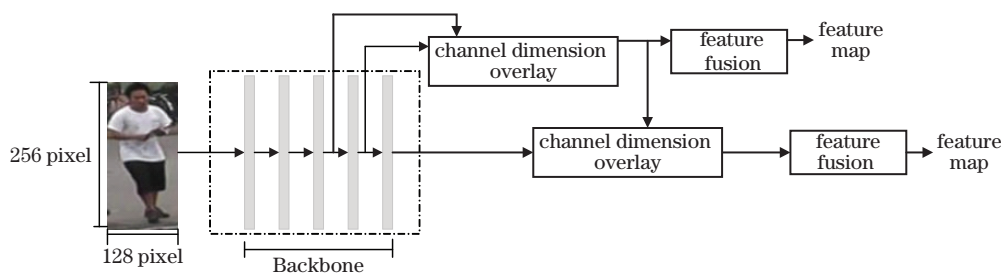


图 2 MFC-Net 的结构

Fig. 2 Structure of the MFC-Net

## 2.2 位置与通道注意力机制模块

PCAM 模块包含一对互补的注意力机制模块 CAM 和 PAM。CAM 采用逐通道特征水平的信息聚合方式聚合语义相关的通道, 提高某一语义特征的显著性, 使网络在学习过程中关注更具判别性的语义特征。PAM 采用聚合相似像素点的方式丰富

特征图中身体部位等空间位置信息。PCAM 将 PAM 和 CAM 进行级联, 使神经网络在训练过程中提高对不同身体部位位置信息的识别能力, 并根据得到的位置信息选择更具判别性的身体特征。PCAM 的结构如图 3 所示。



图 3 PCAM 的结构

Fig. 3 Structure of the PCAM

### 2.2.1 通道注意模块

在神经网络获取特征的过程中, 通常认为高层次卷积块产生特征图的通道语义相关性较高, 即一

些通道共享行人图像中的属性, 如行人、遮挡物、背景。引入 CAM 的主要目的是聚合这些语义相关的通道, 增加提取特征的判别性。CAM 的结构如图 4

所示,设卷积块得到的特征图为  $\mathbf{F} \in \mathbf{R}^{C \times H \times W}$ , 其中,  $C$  为通道数,  $H \times W$  为特征图的尺寸。首先, 重塑 (Reshape)  $\mathbf{F}$  的尺寸, 得到  $\mathbf{F}' \in \mathbf{R}^{C \times N}$ , 其中,  $N = H \times W$ , 将  $\mathbf{F}'$  与其转置相乘, 计算通道的相似性矩阵  $\mathbf{Q} = \mathbf{R}^{C \times C}$ , 得到

$$q_{i,j} = \frac{\exp(\mathbf{F}'_i, \mathbf{F}'_j)}{\sum_{j=1}^C \exp(\mathbf{F}'_i, \mathbf{F}'_j)}, \quad (1)$$

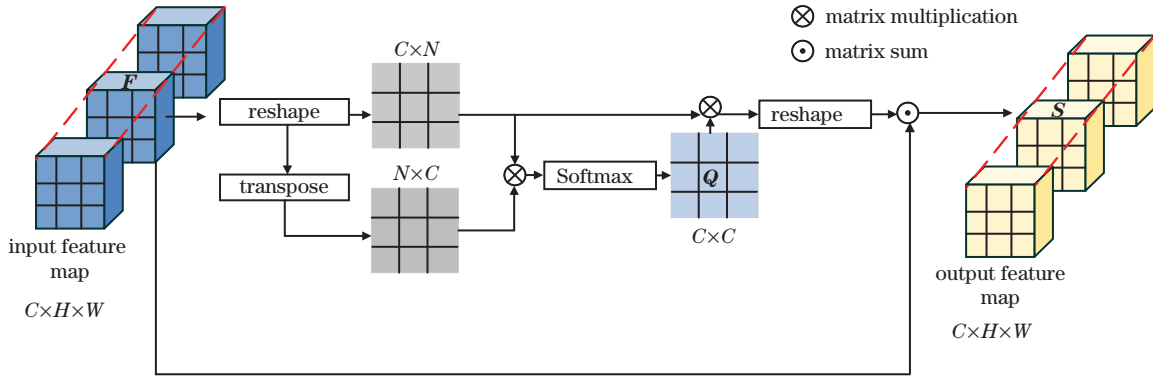


图 4 CAM 的结构

Fig. 4 Structure of the CAM

### 2.2.2 空间注意模块

行人图像中很多像素点的语义是相关的, 为了补充提取特征中的空间位置信息, 引入了 PAM。PAM 与 CAM 类似, 不同的是需要计算像素的相似性矩阵  $\mathbf{P} = \mathbf{R}^{N \times N}$ , 然后再经过矩阵相乘、重塑等运算与原始特征图在对应像素点进行叠加, 就能得到最终的空间注意力加权特征图。

### 2.3 重排序方法

重排序方法用  $k$  个相互最近邻算法对最初的检索结果进行重新排序, 以提高最终的识别精度。主要思路: 如果  $\mathbf{p}$  为查询图像,  $\mathbf{g}_z$  为图库中的一张图像, 且二者同属一个人, 则检索结果中  $\mathbf{g}_z$  应该在  $\mathbf{p}$  的前  $k$  个近邻里。反之,  $\mathbf{p}$  也应该在  $\mathbf{g}_z$  的前  $k$  个近邻里, 符合该条件的图像就是最终的检索结果。首先, 基于马氏距离 (Mahalanobis distance) 计算  $\mathbf{p}$  与图库图像的距离, 可表示为

$$d(\mathbf{p}, \mathbf{g}_z) = (\mathbf{x}_p, \mathbf{x}_{g_z})^T \mathbf{M} (\mathbf{x}_p - \mathbf{x}_{g_z}), \quad (4)$$

式中,  $\mathbf{x}_p$  为  $\mathbf{p}$  的特征向量,  $\mathbf{x}_{g_z}$  为  $\mathbf{g}_z$  的特征向量,  $\mathbf{M}$  为协方差矩阵。根据  $d(\mathbf{p}, \mathbf{g}_z)$  取前  $k$  张图像作为最初的检索结果, 可表示为

$$\mathbf{H}(\mathbf{p}, k) = \mathbf{g}_1^0, \dots, \mathbf{g}_k^0, \quad |\mathbf{H}(\mathbf{p}, k)| = k, \quad (5)$$

式中,  $|\cdot|$  为最初检索结果中图像的数目。然后, 基于雅卡尔距离 (Jaccard distance) 计算  $\mathbf{p}$  与  $\mathbf{H}(\mathbf{p}, k)$  中每张图库图像的距离, 可表示为

式中,  $q_{i,j}$  为通道  $i$  对通道  $j$  的影响。将  $\mathbf{Q}$  与  $\mathbf{F}'$  相乘, 赋予相似通道较大的权重, 得到的矩阵  $\mathbf{V}$  可表示为

$$\mathbf{V} = \mathbf{Q} \times \mathbf{F}', \quad (2)$$

将重塑后的  $\mathbf{V}$  与  $\mathbf{F}$  在对应像素点进行叠加, 得到通道注意力机制加权的特征图为

$$\mathbf{S} = \alpha \mathbf{V} + \mathbf{F}, \quad (3)$$

式中,  $\alpha$  为控制通道注意力模块的权重。

$$d_j(\mathbf{p}, \mathbf{g}_z) = 1 - \frac{|\mathcal{R}^*(\mathbf{p}, k) \cap \mathcal{R}^*(\mathbf{g}_z, k)|}{|\mathcal{R}^*(\mathbf{p}, k) \cup \mathcal{R}^*(\mathbf{g}_z, k)|}, \quad (6)$$

式中,  $\mathcal{R}^*(\mathbf{p}, k) = \{ [\mathbf{g}_z \in \mathbf{H}(\mathbf{p}, k)] \cap [\mathbf{p}_z \in \mathbf{H}(\mathbf{g}_z, k)] \}$  为检索到的所有符合  $k$  个相互最近邻的图像集合。最后, 将得到的马氏距离  $d(\mathbf{p}, \mathbf{g}_z)$  与雅卡尔距离  $d_j(\mathbf{p}, \mathbf{g}_z)$  进行加权, 得到的最终距离可表示为

$$d_f = (1 - \beta)d_j(\mathbf{p}, \mathbf{g}_z) + \beta d(\mathbf{p}, \mathbf{g}_z), \quad (7)$$

式中, 参数  $\beta$  可控制两种距离的权重, 根据最终的距离可得到重排序的结果。

## 3 实验结果与分析

为了验证本算法的有效性, 在三个公开的大规模行人重识别数据集 Market-1501、DukeMTMC-ReID 和 CUHK03 上进行实验。首先, 通过对比实验验证 MFC-Net 以及 PCAM 对算法性能的影响; 然后, 将本算法与当前效果最好的几种算法进行对比, 并将 Baseline、加入 MFC-Net 的 Baseline 和 AMFC 在特征图上关注的区域进行可视化, 检验各个模块对算法的影响; 最后, 将 Baseline、加入 MFC-Net 的 Baseline 和 AMFC 在摄像头视角变化、不同行人穿着相近、光照变化以及存在遮挡等复杂场景下的检索结果进行可视化, 进一步展现各模块的作用及本算法的有效性。



### 3.1 数据集

Market-1501 数据集中的图像由 6 个摄像头采集(5 个高清摄像头和 1 个低清摄像头),共包含 1501 个行人的 32668 张有标签图像。其中,每个行人至少出现在 2 个摄像头下;且每个摄像头中不同行人可能有多张图像。训练集包含 751 个行人的 12936 张图像。测试阶段使用 750 个行人的 3368 张图像作为查询图像,其余 19732 张图像作为图库图像。

DukeMTMC-ReID 数据集是 DukeMTMC 数据集的一个子集,专门用于行人重识别问题,其包含的图像由 8 个高清摄像头采集,共 1812 个行人的 36411 张图像;且 1404 个行人至少由 2 个摄像头捕获,408 个行人仅出现在 1 个摄像头下。与 Market-1501 数据集相似,用 702 个行人的 16522 张图像作为训练集,其余 702 个行人的图像作为测试集,包含 2228 张查询图像和 17661 张图库图像。

CUHK03 数据集由 5 对摄像头采集,共包含 1467 个不同的行人。CUHK03 数据集的测试协议有两种,第一种测试协议舍弃第 4、第 5 对摄像头采集的行人图像,在其余 1360 个行人中随机选取 100 个行人的图像作为测试集,选取 1160 个行人的图像作为验证集。第二种测试协议使用所有行人图像,其中,训练集包含 767 个行人的图像,测试集包含 700 个行人的图像;且根据行人检测框的获取方式将数据集分成人工标注(labeled)和基于可变形零件模型(DPM)标注(detected)的两个子集,labeled 子集包含 7368 张训练图像、1400 张查询图像和 5328 张图库图像,detected 子集包含 7365 张训练图像、1400 张查询图像和 5332 张图库图像。实验采用是第二种测试协议中的 labeled 子集。

### 3.2 实验细节

实验用 ResNet50 作为 Backbone,先进行预训练,然后对网络进行微调,将不同层的特征进行级联,并添加注意力模块,将最后一个卷积块的步长由 2 变成 1,同时去除最后的池化层和全连接层。训练阶段,将输入图像的尺寸调整为 256 pixel $\times$ 128 pixel,使用随机裁剪、随机水平翻转以及随机擦除等图像增强方法以及像素归一化方法对图像数据进行预处理,训练总轮数 epoch 为 65,训练批尺寸 batchsize 为 8。使用 Warmup 学习率优化方法激活网络,设置初始学习率为 0.002,预热 epoch 为 5 个,稳定后的基本学习率为 0.02;且学习率每经过 30 个 epoch 衰减一次,衰减系数为 0.1。采用随机梯度下降

(SGD)法优化网络模型的参数,设置的权重衰减参数为 0.0005,动量为 0.9。网络的损失为 3 个分支特征损失的平均值。测试阶段不使用图像增强算法,仅使用像素归一化预处理方法,输入图像的尺寸为 256 pixel $\times$ 128 pixel。

本网络用 1 块 NVIDIA GeForce RTX 2060 GPU 进行训练,操作系统为 Ubuntu16.04,CUDA 版本为 10.0,采用基于 Python3.6.10 的深度学习框架 Pytorch1.2.0 实现算法的编程。网络模型的训练大约需要 3 h,采用行人重识别领域中两个最常用的评价指标首位命中率(Rank-1)和平均精度均值(mAP)对网络的性能进行评估。Rank- $n$  表示检索结果中前  $n$  张图像与查询图像都属于同一个人的概率,但行人重识别领域通常以 Rank-1 作为衡量标准,即第 1 张图像就是正确结果的概率。mAP 用于评价模型在所有类别上检测结果的好坏,其结合查全率与查准率,可全面衡量行人重识别算法的性能。

### 3.3 仿真实验与结果分析

为了验证 MFC-Net、PCAM 模块和 AMFC 算法的有效性,在 Market-1501、DukeMTMC-ReID、CUHK03 数据集上进行对比实验。基础网络 Baseline 为使用交叉熵损失的 ResNet50,且仅用最后一个卷积层得到的特征计算损失。对实验结果进行重排序,以提高识别精度。此外,将模型在特征图上的关注区域、检索结果进行可视化,以直观体现每个模块对算法的作用效果。

#### 3.3.1 MFC-Net 的性能评估

MFC-Net 将特征提取网络中第 3、第 4、第 5 个卷积层得到特征图进行级联,并计算损失。为了验证 MFC-Net 对算法性能的提升效果,将单一层级与多层级级联对网络性能的影响进行了对比实验,结果如表 1 所示。可以发现,加入 MFC-Net 后,网络在 Market-1501 数据集上的 Rank-1 达到了 93.2%,mAP 达到了 83.5%,相比 Baseline 分别提升了 1.8 和 5.7 个百分点。在 DukeMTMC-ReID 数据集上的 Rank-1 达到了 85.6%,mAP 达到了 73.0%,相比 Baseline 分别提升了 3.5 和 6.2 个百分点。在 CUHK03 数据集上的 Rank-1 达到了 65.0%,mAP 达到了 60.6%,相比 Baseline 分别提升了 6.6 和 8.0 个百分点。经过 RK 后,网络在 Market-1501 数据集上 Rank-1 达到了 94.5%,mAP 达到了 92.2%,相比 Baseline 分别提升了 1.3 和 1.9 个百分点。在 DukeMTMC-ReID 数据集上的 Rank-1 达

到了 89.1%, mAP 达到了 86.4%, 相比 Baseline 分别提升了 1.6 和 1.9 个百分点。在 CUHK03 数据集上的 Rank-1 达到了 75.5%, mAP 达到了

76.2%, 相比 Baseline 分别提升了 5.8 和 5.5 个百分点。这表明加入 MFC-Net 可大幅度提升模型的性能。

表 1 MFC-Net 对算法性能的影响

Table 1 Influence of MFC-Net on the performance of the algorithm

unit: %

Algorithm	Market-1501		DukeMTMC-ReID		CUHK03	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
Baseline	91.4	77.8	82.1	66.8	58.4	52.6
MFC-Net	93.2	83.5	85.6	73.0	65.0	60.6
Baseline(+RK)	93.2	90.3	87.5	84.5	69.7	70.7
<b>MFC-Net(+RK)</b>	<b>94.5</b>	<b>92.2</b>	<b>89.1</b>	<b>86.4</b>	<b>75.5</b>	<b>76.2</b>

### 3.3.2 PCAM 的性能评估

为检验 PCAM 对网络性能的提升效果, 保留 MFC-Net 并在第 2 个卷积层的特征图、第 5 个卷积层的特征图以及级联得到的特征图后添加 PCAM 模块构成 AMFC, 对比实验结果如表 2 所示。可以发现, 加入 PCAM 后, 在 Market-1501 数据集上网络的 Rank-1 与 mAP 分别达到了 93.8% 和 84.1%, 相比结合 Baseline 和 MFC-Net 模块的网络, 均提升了 0.6 个百分点。在 DukeMTMC-ReID 数据集上网络的 Rank-1 与 mAP 分别为 86.2% 和 73.9%, 相比结合 Baseline 和 MFC-Net 模块的网络, 分别提升了 0.6 和

0.9 个百分点。在 CUHK03 数据集上网络的 Rank-1 与 mAP 分别为 65.7% 和 60.5%, 相比结合 Baseline 和 MFC-Net 模块的网络, Rank-1 提升了 0.7 个百分点, mAP 基本持平。经过 RK 后, 在 Market-1501 数据集上, 结合 Baseline 和 MFC-Net 模块的网络与 AMFC 的 Rank-1 基本持平, 但 AMFC 的 mAP 提升了 0.3 个百分点。在 DukeMTMC-ReID 数据集上 Baseline 加上 MFC-Net 与 AMFC 的两项指标基本一致。在 CUHK03 数据集上 Rank-1 基本持平, mAP 提升了 0.5 个百分点, 这表明添加 PCAM 可以提升网络的识别能力和识别精度。

表 2 PCAM 对算法性能的影响

Table 2 Influence of PCAM on algorithm performance

unit: %

Algorithm	Market-1501		DukeMTMC-ReID		CUHK03	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
MFC-Net	93.2	83.5	85.6	73.0	65.0	60.6
AMFC	93.8	84.1	86.2	73.9	65.7	60.5
MFC-Net(+RK)	94.5	92.2	89.1	86.4	75.5	76.2
<b>AMFC(+RK)</b>	<b>94.6</b>	<b>92.5</b>	<b>89.1</b>	<b>86.5</b>	<b>75.6</b>	<b>76.7</b>

### 3.3.3 不同算法的对比

为验证本算法的有效性, 分别在 Market-1501、DukeMTMC-ReID、CUHK03 数据集上对比了本算法与目前主流算法的性能, 结果如表 3 所示。可以发现, 本算法在 Rank-1 和 mAP 两项评价指标上的性能远优于大多数主流算法。相比基于部分卷积基

线(PCB)的行人重识别、AlignedReID++、DaRe 等算法, 本算法在 Market-1501、DukeMTMC-ReID、CUHK03 数据集上的各项性能指标均最好。这表明 MFC-Net 与 PCAM 的结合能很好地提升网络性能, 也验证了 AMFC 的有效性。

表 3 不同算法的检测性能

Table 3 Detection performance of different algorithms

unit: %

Algorithm	Market-1501		DukeMTMC-ReID		CUHK03	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
SVDNet <sup>[16]</sup>	82.3	62.1	76.7	56.8	40.9	37.8
PCB <sup>[17]</sup>	92.3	77.4	81.9	65.3	59.7	53.2
PCB+RPP <sup>[17]</sup>	93.8	81.6	83.3	69.2	62.8	56.7
DaRe <sup>[18]</sup>	89.0	76.0	80.2	64.5	60.2	64.5
AlignedReID++ <sup>[19]</sup>	91.0	77.6	80.7	68.0	60.9	59.7
CASN(PCB) <sup>[20]</sup>	94.4	82.8	87.7	73.7	73.7	68.0
<b>AMFC(+RK)</b>	<b>94.6</b>	<b>92.5</b>	<b>89.1</b>	<b>86.5</b>	<b>75.6</b>	<b>76.7</b>

### 3.3.4 关注区域的可视化

选取 Market-1501 数据集中 在摄像头视角变化、不同行人穿着相近、光照变化及存在遮挡等复杂场景下拍摄的行人图像,并用 Baseline、加入 MFC-Net 的 Baseline 以及 AMFC 进行特征提取,将最后 1 个卷积层得到的特征图关注区域进行可视化,结果如图 5 所示。其中,深色区域表示网络主要关注的区域,即网络认为判别性高的区域。可以发现,Baseline 仅关注了图像中部分具有判别性的区域,加入 MFC-Net 后扩大了网络关注的范围,使网络在学习过程中捕捉到了更多细节信息,获取了较多具有判别性的区域。而 AMFC 关注的区域更加集中,获取的特征更准确、判别性更高。

### 3.3.5 可视化效果的对比

为了更直观地展现 MFC-Net、PCAM 的作用以及 AMFC 算法的有效性,选取 Market-1501 数据集中 在复杂场景下捕获的行人图像,用 Baseline、加入 MFC-Net 的 Baseline 以及 AMFC 算法对同一张查询图像进行测试,并将 Rank-5 的结果进行可视化。Baseline、加入 MFC-Net 的 Baseline 以及 AMFC 算法对遮挡、摄像头视角变化图像的处理结果如图 6 所示。可以发现,Baseline 算法的检索结果图 6(a5)为错误结果,效果不好,无法很好地应对遮挡、摄像头视角变换的影响;加入 MFC-Net 的 Baseline 算法检索结果中图 6(b6)是错误的,可在一定程度上

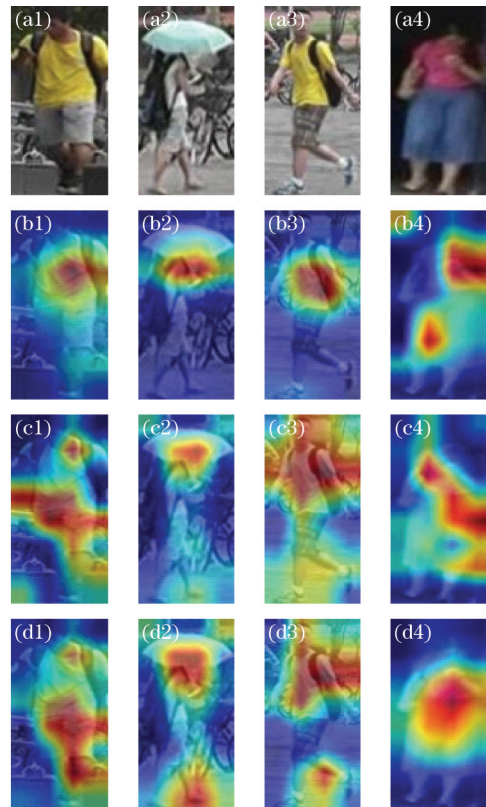


图 5 不同算法关注区域的可视化结果。(a)原始图像;  
(b) Baseline;(c)加入 MFC-Net 的 Baseline;(d) AMFC  
Fig. 5 Visualization results of the area of interest of different algorithms. (a) Original image;  
(b) Baseline; (c) Baseline with MFC-Net;  
(d) AMFC

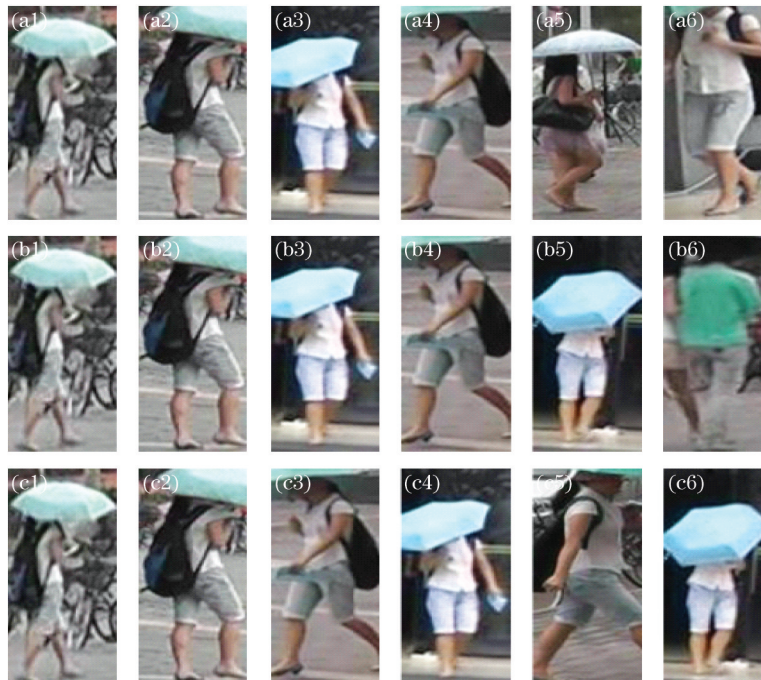


图 6 不同算法对遮挡、摄像头视角变化图像的处理结果。(a) Baseline;(b)加入 MFC-Net 的 Baseline;(c) AMFC  
Fig. 6 Processing results of different algorithms on the occlusion and camera viewing angle changes.

(a) Baseline; (b) Baseline with MFC-Net; (c) AMFC



减小因遮挡、摄像头视角变化带来的影响; AMFC 算法的检索结果全部正确, 可有效减小遮挡、摄像头视角变化的影响。

Baseline、加入 MFC-Net 的 Baseline 以及 AMFC 算法对光照变化图像的处理结果如图 7 所



图 7 不同算法对光照变化图像的处理结果。(a) Baseline; (b) 加入 MFC-Net 的 Baseline; (c) AMFC

Fig. 7 Processing results of different algorithms on the images of various illumination.

(a) Baseline; (b) Baseline with MFC-Net; (c) AMFC

Baseline、加入 MFC-Net 的 Baseline 以及 AMFC 算法对穿着相似图像的处理效果如图 8 所示。可以发现, Baseline 算法的检索结果中图 8(a2) 和图 8(a4) 为错误结果, 不能有效应对穿着相似的

示。可以发现, Baseline 算法的检索结果中最后 3 张图像均为错误结果, 效果不好; 加入 MFC-Net 的 Baseline 算法以及 AMFC 算法的检索结果中最后 2 张图像为错误结果, 可在一定程度上减小光照变化的影响。

影响; 加入 MFC-Net 的 Baseline 算法检索结果中仅最后 1 张图像检索错误, 可在一定程度上减小穿着相似的影响; AMFC 算法的检索结果全部正确, 可有效减小穿着相似的影响。

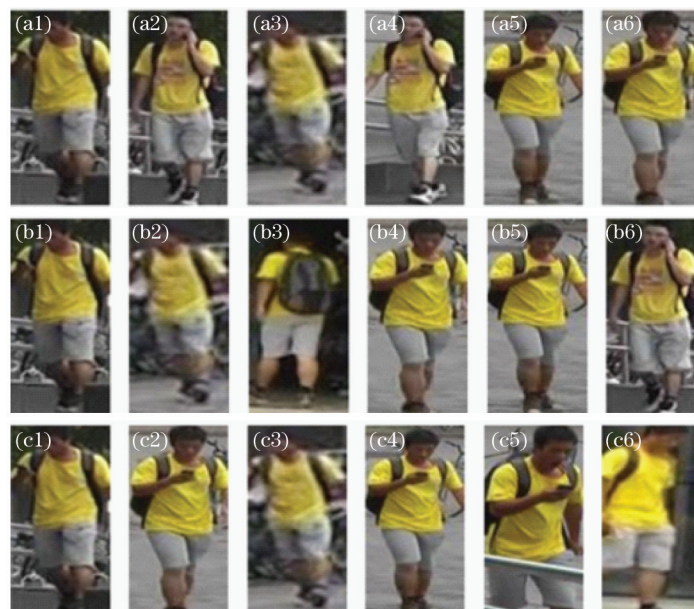


图 8 不同算法对穿着相似图像的处理结果。(a) Baseline; (b) 加入 MFC-Net 的 Baseline; (c) AMFC

Fig. 8 Processing results of different algorithms on images with similarly dressed person.

(a) Baseline; (b) Baseline with MFC-Net; (c) AMFC



## 4 结 论

提出了一种基于注意力机制的多级特征级联行人重识别算法。该算法将不同层级的特征图进行级联,充分利用了浅层特征图的细节信息,实现高层级特征中信息的补充。将级联得到的特征图通过注意力模块,聚合语义相关的像素及通道,丰富特征中的空间位置信息,增强特征的判别性。在 Market-1501、DukeMTMC-ReID、CUHK03 数据集上的实验结果表明,相比其他主流算法,本算法在复杂场景下的精确度更高、鲁棒性更强。

### 参 考 文 献

- [1] Zhu X B, Che J. Person re-identification algorithm based on feature fusion and subspace learning [J]. *Laser & Optoelectronics Progress*, 2019, 56 (2): 021503.  
朱小波, 车进. 基于特征融合与子空间学习的行人重识别算法[J]. *激光与光电子学进展*, 2019, 56(2): 021503.
- [2] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C] // 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), June 20-25, 2005, San Diego, CA, USA. New York: IEEE Press, 2005: 886-893.
- [3] Wu L, Shen C H, Hengel A. PersonNet: person re-identification with deep convolutional neural networks [EB/OL]. (2016-01-27) [2020-11-20]. <https://arxiv.org/abs/1601.07255>.
- [4] Zheng L, Zhang H H, Sun S Y, et al. Person re-identification in the wild [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 3346-3355.
- [5] Qian X L, Fu Y W, Jiang Y G, et al. Multi-scale deep learning architectures for person re-identification [C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 5409-5418.
- [6] Zhang T, Yi Z M, Li X, et al. Improved algorithm for person re-identification based on global features [J]. *Laser & Optoelectronics Progress*, 2020, 57 (24): 241503.  
张涛, 易争明, 李璇, 等. 一种基于全局特征的行人重识别改进算法[J]. *激光与光电子学进展*, 2020, 57(24): 241503.
- [7] Zhao L M, Li X, Zhuang Y T, et al. Deeply-learned part-aligned representations for person re-identification [C] // 2017 IEEE International Conference on Computer Vision (ICCV), October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 3239-3248.
- [8] Suh Y, Wang J D, Tang S Y, et al. Part-aligned bilinear representations for person re-identification [M] // Ferrari V, Hebert M, Sminchisescu C, et al. *Computer vision-ECCV 2018. Lecture notes in computer science*. Cham: Springer, 2018, 11218: 418-437.
- [9] Xu L Z, Peng L. Person reidentification based on multiscale convolutional feature fusion [J]. *Laser & Optoelectronics Progress*, 2019, 56(14): 141504.  
徐龙壮, 彭力. 基于多尺度卷积特征融合的行人重识别[J]. *激光与光电子学进展*, 2019, 56(14): 141504.
- [10] Chen T L, Ding S J, Xie J Y, et al. ABD-Net: attentive but diverse person re-identification [C] // 2019 IEEE/CVF International Conference on Computer Vision (ICCV), October 27-November 2, 2019, Seoul, Korea (South). New York: IEEE Press, 2019: 8350-8360.
- [11] Zhang H, Goodfellow I, Metaxas D, et al. Self-attention generative adversarial networks [EB/OL]. (2018-05-21) [2020-11-20]. <http://export.arxiv.org/abs/1805.08318>.
- [12] Zheng L, Shen L Y, Tian L, et al. Scalable person re-identification: a benchmark [C] // 2015 IEEE International Conference on Computer Vision (ICCV), December 7-13, 2015, Santiago, Chile. New York: IEEE Press, 2015: 1116-1124.
- [13] Ristani E, Solera F, Zou R, et al. Performance measures and a data set for multi-target, multi-camera tracking [M] // Hua G, Jégou H. *Computer vision-ECCV 2016. Lecture notes in computer science*. Cham: Springer, 2016, 9914: 17-35.
- [14] Li W, Zhao R, Xiao T, et al. DeepReID: deep filter pairing neural network for person re-identification [C] // 2014 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2014, Columbus, OH, USA. New York: IEEE Press, 2014: 152-159.
- [15] Zhong Z, Zheng L, Cao D L, et al. Re-ranking person re-identification with  $k$ -reciprocal encoding [C] // 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 3652-3661.
- [16] Sun Y F, Zheng L, Deng W J, et al. SVDNet for pedestrian retrieval [C] // 2017 IEEE International Conference on Computer Vision (ICCV), October

- 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 3820-3828.
- [17] Sun Y F, Zheng L, Yang Y, et al. Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline)[M]//Ferrari V, Hebert M, Sminchisescu C, et al. Computer vision-ECCV 2018. Lecture notes in computer science. Cham: Springer, 2018, 11208: 501-518.
- [18] Wang Y, Wang L Q, You Y R, et al. Resource aware person re-identification across multiple resolutions[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 8042-8051.
- [19] Luo H, Jiang W, Zhang X, et al. AlignedReID++: dynamically matching local information for person re-identification[J]. Pattern Recognition, 2019, 94: 53-61.
- [20] Zheng M, Karanam S, Wu Z Y, et al. Re-identification with consistent attentive siamese networks[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 5728-5737.