双路注意力机制行人重识别方法

张媛媛[1)], 宋存利[2)], 张雪松[2)*]

[1)] (大连交通大学计算机与通信工程学院 大连 116028)
[2)] (大连交通大学软件学院 大连 116028)
(zhangxuesong@djtu.edu.cn)

摘 要: 为解决目前Re-ID方法中对非显著可辨别特征关注不足, 以及提取的行人关键特征表达不充分的问题, 提出一种基于双路注意力机制特征提取网络, 由双路注意力主干网络和增强注意特征融合模块组成. 其中, 双路注意力网络使模型关注到不同显著程度的有效特征区域, 可分别用于挖掘显著和潜在非显著可辨别特征, 强调潜在关键特征的重要性; 增强注意特征融合模块用于完成特征信息互补, 同时采用反事实干预强化习得注意力特征图的质量和有效性, 从而得到更具有判别性的最终特征表示. 在 Market1501, DukeMTMC-reID 和 MSMT17 数据集上进行了广泛实验, 结果表明, mAP 值分别达到了 89.3%, 80.0%, 58.4%; Rank-1 值分别达到了 95.7%, 89.8%, 80.7%, 充分证明了该方法的优越性.

关键词: Re-ID; 深度学习; 注意力; 非显著特征; 反事实干预
中图法分类号: TP391.41 **DOI:** 10.3724/SP.J.1089.2023-00416

Person Re-Identification via Dual-Path Attention

Zhang Yuanyuan[1)], Song Cunli[2)], and Zhang Xuesong[2)*]

[1)] (*School of Computer and Communication engineering, Dalian Jiaotong University, Dalian* 116028)
[2)] (*School of Software, Dalian Jiaotong University, Dalian* 116028)

Abstract: In order to solve the problem of insufficient attention to non-salient distinguishable features and insufficient expression of extracted key features of pedestrians in current Re-ID methods, a feature extraction network based on a dual-path attention mechanism is proposed (Dual-Path Attention Network), which consists of a dual-path attention backbone network. It is composed of an enhanced attention feature fusion module. Among them, the dual-path attention network allows the model to focus on effective feature areas with different salience levels, which can be used to mine significant and potentially non-salient distinguishable features respectively, emphasizing the importance of potential key features.The enhanced attention feature fusion module is used to complete feature information complementation, and counterfactual intervention is used to enhance the quality and effectiveness of the learned attention feature map, so as to make the final feature representation more discriminative. Extensive experiments were conducted on Market1501, DukeMTMC-reID and MSMT17 data sets, and the results showed that the mAP values reached 89.3%, 80.0%, 58.4% respectively; the Rank-1 values reached 95.7%, 89.8%, 80.7% respectively, which fully proved the superiority of this method.

Key words: person re-identification; deep learning; attention mechanism; non-distinctive; counterfactual intervention

收稿日期: 2023-07-06; 修回日期: 2024-05-02. **基金项目**: 国家自然科学基金(62276042); 辽宁省教育厅项目(LJKZ0486). **张媛媛**(1993—), 女, 硕士研究生, 主要研究方向为计算机视觉; **宋存利**(1975—), 女, 博士, 教授, 硕士生导师, 主要研究方向为计算机视觉、智能优化算法; **张雪松**(1980—), 男, 博士, 副教授, 硕士生导师, CCF 会员, 论文通信作者, 主要研究方向为计算机视觉、人工智能等.

行人重识别[1](person re-identification, Re-ID)是图像检索领域中的一个子问题，旨在根据给定的特定行人图像，从非重叠摄像机中捕获该行人的图像或视频. 近些年来, Re-ID 问题得到许多国内外学者的关注和研究，常与行人检测或行人跟踪技术相结合，并被广泛应用于公共安全和智慧交通等实际领域，为实现人员监控、安全防范等任务提供了关键的技术支持. 尽管目前相关研究已取得了巨大的进展，但在现实中受光照变化、摄像机视角、行人姿态和遮挡等因素影响，行人图像的类内差异较大，使得在庞大数据库中学习具有判别性的特征并成功地识别出特定行人仍具有巨大的挑战性.

传统 Re-ID 方法主要依靠人工设计图像特征(如纹理、颜色和外观等)，并通过相似度度量判别是否为同一行人. 因此，在早期的 Re-ID 研究中，主要关注如何获取更具有表征能力的行人特征，如 RGB, HOG[2], SIFT[3], 局部二值模式 LBP[4]和 LOMO 等，以及学习鲁棒性更高的相似度度量，如 XQDA[5], LMNN[6], KISSME[7]和 LFDN[8]. 但随着数据量的不断增加和实际任务需求的提高，传统的 Re-ID 方法已经无法适应日益复杂场景下识别任务的需求. 随着深度学习技术在计算机视觉领域的快速发展，基于深度学习的 Re-ID 方法在识别精度和检索时间上都取得了巨大突破. 2014 年，深度学习被 Li 等[9]成功地应用到 Re-ID 任务中，可在滤波配对神经网络(filter pairing neural network, FPNN)模型下获取更具鲁棒性行人特征，实现了端对端的 Re-ID.

早期基于深度学习的 Re-ID 方法主要侧重于捕捉行人的全局特征，但由于全局特征表达信息的能力有限，无法挖掘到更具辨别能力的行人特征，因此，仅仅利用全局特征无法进一步提升模型性能. 为了解决这一问题，Sun 等[10]提出了基于图像水平分割的局部卷积(part-based convolutional baseline, PCB)算法，将行人特征图水平分为 6 块，然后通过对每个水平块计算分类损失来学习行人的局部特征. 针对行人图像未对齐问题，Luo 等[11]提出了 Aligned-ReID 算法，通过对特征图进行水平分块，计算 2 幅图像局部特征之间的最短距离来实现水平块间的动态对齐，有效地缓解了图像块对齐问题. 基于姿态估计也是提取局部特征方法之一, Zhao 等[12]提出了一种基于人体区域引导的、多阶段特征分解和树形结构的竞争特征融合网络结构 SpindleNet，通过预估 14 个人体关键点信息，使用模型提取到 7 个人体区域的局部特征，有效地避免了图像错位导致的人体区域特征未对齐问题. 虽然增强了局部细节信息的表示能力，却无法避免由身体姿势和遮挡带来的性能提升瓶颈. 大多基于部分的研究方法只关注到固定区域的语义信息，缺乏挖掘不同层级和粒度判别特征的能力. 针对此类问题，Wang 等[13]提出一种综合全局和局部信息的不同粒度特征学习策略，设计的多粒度网络(multiple granularity network, MGN)结构能够引导模型获取到不同粒度的特征表示. Park 等[14]提出一种关系网络，考虑了各个身体部位和其他部分之间的关系，其模型让单一的部分级特征同时包含了身体其他部位的部分信息，从而能够获取具有更强辨别力的行人特征.

尽管上述方法可以获得表达能力较强的局部特征信息，但无法避免与身份无关信息的干扰. 因此，许多研究工作通过引入注意力机制引导模型关注到更为显著的关键信息，忽略甚至过滤掉无关信息. Woo 等[15]提出了一种轻量级卷积块注意力模块(convolutional block attention module, CBAM)，它沿着通道和空间 2 个维度顺序推断注意映射，然后将注意映射和输入特征图相乘，进而实现自适应特征细化. Chen 等[16]提出一个残差双注意力模块(residual dual attention module, RDAM)，使模型能够挖掘出更多必要的语义信息. Zhang 等[17]提出了一个关系感知全局注意力(relation-aware global attention, RGA)模块，来引导模型捕获全局范围和局部外观信息的结构信息，并显著地提高特征的表示能力. 但这些注意力方法大多只关注显著区域特征，而忽略了被显著区域掩盖的非显著可辨别特征，从而限制了网络进一步提升性能的能力. Rao 等[18]提出了一个基于因果推理的反事实注意学习方法来学习更有效的注意，它提供了一种测量注意质量的工具，以及一个强大的监督信号来指导学习过程. 该方法通过使用反事实干预方法分析学习到的视觉注意对网络预测的影响，来最大限度地提高影响，以鼓励网络学习更多有用的注意.

基于以上问题，并受上述文献的启发，本文提出一种基于双路注意力机制特征提取的 Re-ID 方法(person re-identification via dual-path attention, DPA-Net)，不仅能够使模型关注到不同显著程度的有效识别特征，还可以强化获得的关键特征信息. 首先，本文提出一个双路注意力模块(dual-path attention module, DPAM)，使用基于通道注意力挤压和激励(squeeze-and-excitation, SE)模块[19]的通道分区块对行人特征信息通道进行自动学习，将行人特征图划

分为强信号通道区域和弱信号通道区域；然后引入基于坐标注意力[20]的双路注意力分支，分别挖掘显著特征和潜在的非显著可辨别特征；其次，设计一个增强注意特征融合模块(enhanced attention feature fusion module, EAFFM)，通过自主学习，对不同强度特征设置不同的权重，丰富语义信息并完成信息互补；同时，采用反事实干预强化习得注意力特征图的质量和有效性，从而得到更具有判别性的最终特征表示．为了满足更复杂的现实环境要求，在随机遮挡的基础上，通过增加随机灰度和亮度的变化来实现数据增强，提升了模型的泛化能力．

1　DPANet 网络模型

DPANet 方法针对现有方法中对非显著可辨别特征关注不足以及提取的行人关键特征表达不充分而提出，其模型结构如图 1 所示. DPANet 的结构由 2 个主要部分组成，分别是嵌入双路注意力结构的残差网络模块 ResNet_DPAM 和 EAFFM. 首先，对低层特征使用通道注意学习，并将其划分为强信号通道区域和弱信号通道区域；然后，使用 DPAM 分别对 2 个区域进行关键特征挖掘，获得显著性关键特征和潜在非显著可辨别特征；它们保留了更多语义线索信息，弥补了不同区域的语义损失；最后使用 EAFFM 根据关键信息表达程度的不同，对 DPAM 的获得的特征设置不同的权重，实现显著可辨别特征和潜在非显著可辨别特征的互补；同时，通过反事实注意(未纠正的注意)生成的特征，即反事实特征对最终预测的影响来强化习得注意力特征图的质量.

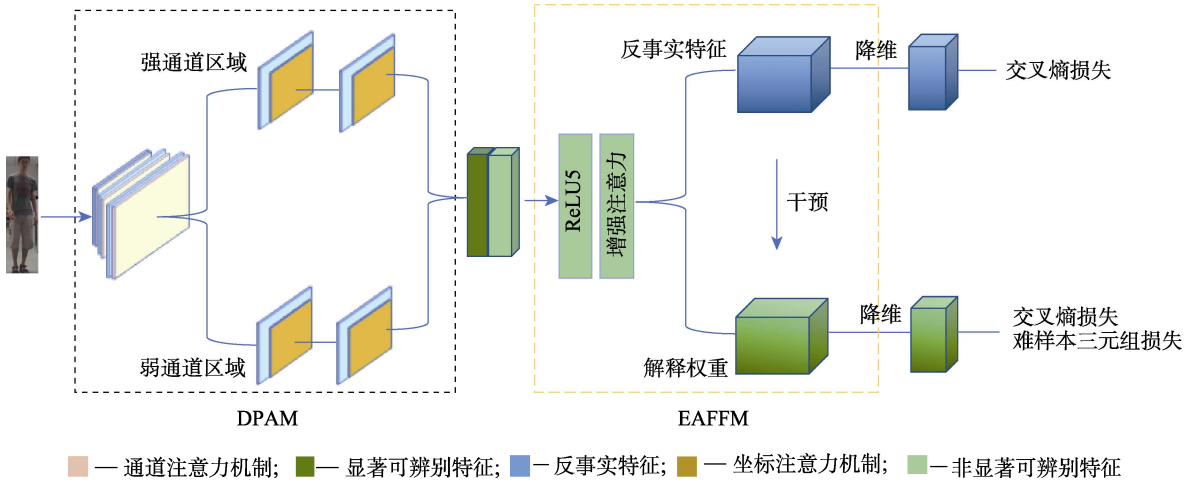


图 1　本文模型结构

1.1　网络模块

1.1.1　ResNet_DPAM

本文选择 ResNet50 网络作为骨干网络，为了增强特征的表达能力，引入了注意力机制．但在实验中发现，引入注意力虽然可以在某种程度上加强关键特征，削弱无关特征的干扰，增强行人特征的判别能力，但是同时也会忽略掉部分有用信息，造成信息缺失．因此，为解决此问题，本文在该网络框架中提出双路注意力网络 ResNet_DPAM.

如图 1 所示的黑色虚线区域，ResNet_DPAM 采用 ResNet50 的前 3 层卷积模块作为浅层特征提取的主干网络，并在前 3 层加入通道注意学习，在第 3 层卷积块之后将其划分为强信号通道区域 R_s 弱信号通道区域 R_w，并在每个卷积块之后引入坐标注意力分别提取各自的关键特征．其中，通过引入通道注意力 SE 模块，自动学习通道重要程度来完成通道分区.

在结构上，SE 模块首先通过 Squeeze 操作将全局空间信息压缩至通道维度中，并使用全局平均池化得到特征压缩后第 c 个通道的特征向量

$$\boldsymbol{Z}_c = F_{\mathrm{sq}}(U_c) = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} U_c(i,j).$$

其中，U_c 表示第 c 个通道大小为 $H \times W \times C$ 的特征图；F 为卷积变换函数.

然后，经过 Excitation 操作，充分利用 Squeeze 操作中的聚合信息，通过 2 个全连接层和激活函数完成通道之间的非线性交互，得到与输入特征数目相同的权重值，用于表示通道的重要程度，即

$$S = F_{\mathrm{ex}}(Z,W) = \sigma(g(Z,W)) = \sigma(W_2\delta(W_1 Z)).$$

其中，W_1 为第 1 个全连接层权重；δ 为 ReLU5 函数；W_2 为第 2 个全连接层函数；σ 为 Sigmoid 函数.

本文将得到的 S 作为强信号通道区域的通道权重值，而弱通道区域是为了解决强信号通道区域忽略掉的非显著重要特征，即强信号通道区域的通道重要程度互换的结果. 因此将 $E-S$ 作为弱信号通道区域的通道权重值，得到激励后的特征图分别为

$$\tilde{X}_{c,\text{strong}} = F_{\text{scale}}(U_c, S_c) = S_c U_c,$$

$$\tilde{X}_{c,\text{weak}} = F_{\text{scale}}(U_c, E_c - S_c) = (E_c - S_c) U_c.$$

其中，$\tilde{X}_c = [\tilde{x}_1, \tilde{x}_2, \cdots, \tilde{x}_c]$ 表示 S 和 U 在通道级相乘得到的特征图；$F_{\text{scale}}(\cdot,\cdot)$ 表示标量与特征映射之间的通道乘法；E 表示重要程度均为 1 的权重值；strong 表示强信号通道区域；weak 表示弱信号通道区域.

尽管 SE 模块在关注内部通道信息表现出色，但它忽略了位置信息的重要性，因此本文在第 4 层和第 5 层卷积块层中通过嵌入坐标注意力(coordinate attention, CA)，以高效并准确地提取行人关键特征. 不同于 SE，CA 在结构上通过 2 个空间维度 $H\times1$ 和 $1\times W$ 分别沿水平坐标和垂直坐标对通道进行编码，生成一对方向感知特征图，分别为

$$Z_c^{\text{h}}(h) = \frac{1}{W}\sum_{0\leqslant i\leqslant W} U_c(h,i),$$

$$Z_c^{\text{w}}(w) = \frac{1}{H}\sum_{0\leqslant j\leqslant H} U_c(j,w).$$

之后，通过系列卷积演化变换得到与输入特征数目相同的权重值 S，并将它们用作注意力权重，分别为

$$S^{\text{h}} = \sigma\left(F_{\text{h}}(f^{\text{h}})\right),$$

$$S^{\text{w}} = \sigma\left(F_{\text{w}}(f^{\text{w}})\right).$$

其中，$f = \delta\left(F_1\left(\left[Z^{\text{h}}, Z^{\text{w}}\right]\right)\right)$，$[\cdot,\cdot]$ 表示沿空间维度的拼接操作.

最后，得到坐标注意特征图 $\tilde{X}_c = U_c \times S_c^{\text{h}} \times S_c^{\text{w}}$. 将 SE 模块和 CA 模块嵌入到 ResNet50 主干网络中，重构为 DPANet，既可以使模型关注到不同显著程度的可辨别特征，弥补单路注意力网络造成的信息缺失，也可使改进后模型所习得的特征表示更具有判别性，特征表达的语义信息更为丰富.

1.1.2 EAFFM

EAFFM 主要是针对 ResNet_DPAM 输出的双注意力特征图. 而设计的特征融合模块，具体结构如图 2 所示. 对于 ResNet_DPAM 输出的 2 个注意特征图，EAFFM 首先对 $\tilde{X}_{\text{strong}}$ 和 $\tilde{X}_{\text{weak}}$ 实现通道维度的拼接，针对两者特征数据对判别结果的贡献度不同，本文设置了 2 个权重参数 α 和 β. 之后，为了捕捉两者通道之间的非线性交互，实现不同显著程度特征块的通道级别语义信息感知，增强数据表征能力，使增强特征具有更为丰富的语义信息，本文将拼接后的特征输入到一个卷积块中，得到融合模块的输入特征 $\tilde{X}_c = F\left(\left[\alpha\cdot\tilde{X}_{\text{strong}}, \beta\cdot\tilde{X}_{\text{weak}}\right]\right)$. 其中，$[\cdot,\cdot]$ 为 cat 操作.

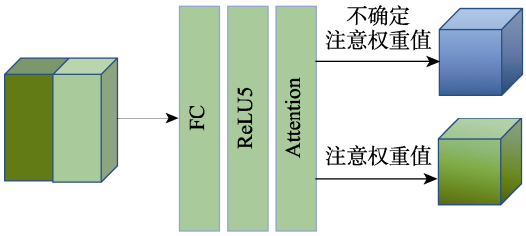


图 2 EAFFM 整体结构

其次，通过 2 个空间维度池化完成 R_{s} 和 R_{w} 特征语义信息间的空间交互，将显著关键特征和非显著可辨别特征建立联系，完成信息互补. 之后通过注意力对完整信息特征图进行自主学习，生成注意权重值，准确地获得突出显示感兴趣的区域的注意力特征图 $X = \tilde{X}\times S_{\text{h}} \times S_{\text{w}}$. 其中，$S_{\text{h}}$ 和 S_{w} 分别为水平和垂直方向习得的注意权重值.

最后，将得到的特征 X 经过由一个 1×1 卷积层、归一化层和激活层构成的卷积块，降低其特征维度并输入到分类器中，用于预测行人的类别概率.

然而，由于模型的不稳定性，习得的注意权重值和特征之间的“伪关系”可能会降低行人特征的泛化性. 因此，为了鼓励网络生成更具有影响力的注意力图，本文在特征融合阶段使用反事实因果关系来强化习得注意力图的质量，提高模型的稳定性和可解释性. 具体来说，假设存在因果模型 M，考虑任意变量 X，Z 和 Y，通过操纵变量的值获得修正后的模型 M_x，进而使用 M_x 和 $Z=z$ 得到反事实结果 $Y_{M_x}(z)$. 在因果推理文献中，利用 do 算子改变模型，记为 do(·)，该形式称为干预.

在本文实验中，预测结果 y 由最终习得的注意力特征决定，即 $y = Y\left(S=S, X=\tilde{X}\right)$. 其中，$S$ 表示习得的注意权重值；$\tilde{X}$ 表示 R_{s} 和 R_{w} 在通道维度拼接后的特征.

受文献[18]启发，本文生成一个不确定注意权

重值 S'，代替习得的注意权重值 S，进以获得新的注意力特征. 对学习到的最终注意力特征进行反事实干预，可得到干预后的最终预测结果

$$y' = Y\left(\mathrm{do}(S = S'), X = \tilde{X}\right).$$

对新目标进行优化，进而优化注意学习过程，通过反事实特征对结果的预测指导注意模型发现更具判别力的特征区域. 优化策略采用交叉熵损失 $L_{\mathrm{id}} = L_{\mathrm{id}}^{y} + L_{\mathrm{id}}^{y'}$. 其中，$L_{\mathrm{id}}^{y}$ 表示正常习得注意力特征得到的交叉熵损失；$L_{\mathrm{id}}^{y'}$ 表示干预后反事实特征得到的交叉熵损失.

1.2　损失函数

为了进一步提高模型对行人的分辨能力，本文采用多损失函数联合对网络模型进行训练和优化. 即交叉熵损失和三元组损失.

1.2.1　交叉熵损失

交叉熵损失被常用于分类问题中，其定义为

$$L_{\mathrm{id}} = -\sum_{i=1}^{n} q_i \lg p_i .$$

其中，p_i 表示输入图像属于第 i 位行人 id 的概率 $p_i = \mathrm{e}^{z_i} \Big/ \sum_{n=1}^{N} \mathrm{e}^{z_n}$，$q_i$ 通过输入图像的 id 标签得到，即

$$q_i = \begin{cases} 1, & y = i \\ 0, & y \neq i \end{cases}.$$

为了防止训练中出现过拟合问题，本文对标签加入平滑操作. 设置一个常数 γ，使图像的 id 标签为 i 时，q_i 无限接近于 1；反之接近于 0. 即

$$q_i = \begin{cases} 1 - \dfrac{N}{N-1}\gamma, & y = i \\ \dfrac{\gamma}{N}, & y \neq i \end{cases}.$$

1.2.2　三元组损失

三元组损失函数 L_{tri} 常用于度量学习，旨在减小类内距离，扩大类间距离. 对于输入的三元组样本集合 $A = \left[I_{\mathrm{a}}, I_{\mathrm{p}}, I_{\mathrm{n}}\right]$，$I_{\mathrm{a}}$ 代表锚点样本，I_{p} 代表正样本，I_{n} 代表负样本. 三元组损失函数的目的是使 $\left\langle I_{\mathrm{a}}, I_{\mathrm{p}} \right\rangle$ 互相靠近，使 $\left\langle I_{\mathrm{a}}, I_{\mathrm{n}} \right\rangle$ 互相远离. 三元组损失函数 $L_{\mathrm{tri}} = \left(d_{\mathrm{a,p}} - d_{\mathrm{a,n}} + a\right)_{+}$. 其中，$a$ 为设定的控制距离差异的超参数.

由于选择样本对中的样本差异对网络性能优化有重要影响，因此为了提高模型性能，加快网络收敛速度，本文实验使用难样本采样三元组损失函数进行训练. 难样本三元组损失函数计算公式为

$$L_{\mathrm{tri}} = \left(\max d_{\mathrm{a,p}} - \min d_{\mathrm{a,n}} + a\right)_{+}.$$

1.2.3　联合损失函数

最终的优化目标函数由加入平滑的交叉熵损失和难样本采样三元组损失组成，即

$$L_{\mathrm{total}} = L_{\mathrm{tri}} + L_{\mathrm{id}} = L_{\mathrm{tri}} + L_{\mathrm{id}}^{y} + L_{\mathrm{id}}^{y'}.$$

2　实验与结果分析

2.1　数据集

本文实验在 Market1501，DukeMTMC-reID 和 MSMT17 这 3 个广泛应用的数据集上展开，以验证本文方法的有效性.

Market1501 数据集中的图像由 6 个不同的摄像头拍摄而成，其中一个为低像素摄像头，包含 1 501 位行人共 32 668 幅图像. 其中训练集包含 751 位行人，12 936 幅图像，测试集包含 750 位行人，19 732 幅图像. DukeMTMC-reID 数据集由 8 个不同摄像头拍摄，包含 1 404 位行人共 36 411 幅图像，图像中的行人边界框经过人工标注完成. 其中，训练集包含 702 位行人，16 522 幅图像，测试集包含 702 位行人，17 661 幅图像. MSMT17 是最大的 Re-ID 数据集之一，由 15 个不同的摄像头拍摄而成，包含 4 101 位行人共 126 411 幅图像. 其中训练集包含 1 041 位行人，30 248 幅图像，其余 3 060 位行人用于测试集.

2.2　实验设置和评价指标

本文实验在 PyTorch 深度学习框架基础上实现，实验硬件配置为单块 NVIDIA Tesla P100 GPU，显卡 16 GB，内存 48 GB. 在数据预处理阶段，通过随机擦除、随机水平翻转、随机灰度和随机亮度变化实现数据增强. 在训练阶段，使用在 ImageNet 上预训练的 ResNet50 作为主干网络，输入图像大小为 256×128，批大小设置为 64，初始学习率设置为 0.000 35，并在第 40，80 和 120 个 epoch 时将学习率衰减为原来的十分之一，共训练 160 个 epoch. 在数据集 Market1501, DukeMTMC-reID 和 MSMT17 上训练时间分别约为 5 h, 7 h 和 12 h.

本文实验采用平均精度均值(mean average precision，mAP)和累计匹配特性曲线(cumulative match characteristic，CMC)作为对改进模型性能评估的衡量指标. 其中，Rank-k 表示排序的前 k 个结果中存在检索目标的命中率.

2.3　对比实验

为了验证 DPANet 方法的有效性，分别在 Market1501，DukeMTMC-reID 和 MSMT17 数据集上进行了训练与测试，并与 Re-ID 方法 MGN[13]，

RGA[17], Pyramid[21]和 APNet[22]等进行了对比. 主要评价指标为 Rank-1, Rank-5 以及 mAP, 对比结果如表 1 所示. 此外, 本文实验均未采用重排序算法(re-ranking[23]).

表 1 不同方法在 3 个数据集上的性能比较 %

方法	Market1501			DukeMTMC-reID			MSMT17		
	mAP	Rank-1	Rank-5	mAP	Rank-1	Rank-5	mAP	Rank-1	Rank-5
PCB[10]	77.4	92.3	97.2	78.4	88.7		40.4	68.2	
PCB+RPP[10]	81.6	93.8	97.5	69.2	83.3				
Aligned-ReID[11]	79.3	91.8							
MGN[13]	86.9	95.7		78.4	88.7				
RGA-SC[17]	88.4	**96.1**					57.5	80.7	
RGA-S[17]	88.0	95.7	98.5	77.5	88.8	94.7	55.8	78.2	88.7
CAL[18]	87.0	94.5	97.9	76.4	87.2	94.1	56.2	79.5	89.0
Pyramid[21]	88.2	95.7	98.4	79.0	89.0				
APNet-S[22]	89.0	96.1	**98.7**	78.8	89.3	85.0	59.0	80.8	89.8
OSNet[24]	84.9	94.8		73.5	88.6		52.9	78.7	
DCAL[25]	87.2	94.5		**80.2**	89.6		**62.3**	**83.1**	
BNNeck[26]	85.7	94.1		75.9	86.2				
DPANet	**89.3**	95.7	98.6	80.0	**89.8**	**95.3**	58.4	80.7	**89.9**

注. 粗体表示最优值.

如表 1 所示, 在 Market1501 数据集上, DPANet 与其他方法相比表现出良好的性能效果, 其中, Rank-1 达到 95.7%, mAP 达到 89.3%. 与 MGN[13], RGA[17], Pyramid[21]和 APNet[22]方法相比, Rank-1 精度结果相近, mAP 分别提高了 2.4 个百分点、1.3 个百分点、1.1 个百分点和 0.3 个百分点. 在 DukeMTMC-reID 数据集上, DPANet 方法与其他方法比较表现相对较好. 其中, Rank-1 达到 89.8%, mAP 达到 80.0%. 与 RGA[17]和 APNet[22]方法相比, Rank-1 分别提高 1.0 个百分点和 0.5 个百分点, mAP 分别提高 2.5 个百分点和 1.2 个百分点. 与大多 Re-ID 方法相比, DPANet 在整体性能上表现较优, 充分证明了 DPANet 方法能够获得更具有判别性的特征表示, 更好地区分不同行人之间的特征, 使得 Re-ID 方法更具有鲁棒性和较好的泛化性.

由于本文是基于全局特征的方法, 选择 CAL[18]方法进行检索对比实验. 表 2 展示了本文方法和 CAL[18]方法在 Market1501 数据集上的检索结果, 序号为绿色字体表示匹配正确的结果, 序号为红色字体代表匹配错误. 本文随机选取了 4 幅具有代表性的行人图像, 可以看出, 当出现不同遮挡和分辨率, 以及姿势变化时, 本文方法能够较好地克服这些挑战, 成功匹配到目标图像, 得到正确结果, 进一步验证了本文方法的有效性和可靠性.

2.4 消融实验

为了进一步检验 DPANet 方法中各个模块的有效性, 在 Market1501 数据上进行了实验验证. 其中, 以使用 BNNeck 和一个 FC 层作为分类器的 ResNet50 作为基线(Baseline)模型. 为了充分说明每个模块的贡献率, 本文将逐个模块添加到模型中进行对比.

如表 3 所示, 与 Baseline 模型相比, 加入单路注意力(single path attention module, SPAM)结构的 mAP 和 Rank-1 达分别到 88.6%和 95.0%, 而引入 DPAM 后 mAP 和 Rank-1 分别达到 89.2%和 95.3%; 与 Baseline 模型相比, mAP 和 Rank-1 分别明显提高了 1.7 个百分点和 0.5 个百分点, 与加入 SPAM 结构相比, mAP 和 Rank-1 分别提高了 0.6 个百分点和 0.3 个百分点. 在此组合基础上加入 EAFFM, mAP 和 Rank-1 分别提高了 0.1 个百分点和 0.4 个百分点. 实验结果充分证明了 DPAM 和 EAFFM 的有效性.

2.4.1 DPAM 分析

DPAM 主要作用是在强化显著性全局关键特征的同时, 关注到非显著潜在可辨别特征, 以弥补由学习不变特征而造成的细节信息丢失. 为了更直观地观察 DPAM 的有效性, 在 Market1501 数据集上, 分别对不同视角行人图像训练得到 Baseline, SPAM 和 DPAM, 通过 Grad-CAM[27]方法得到的可视化结果如图 3 所示. 通过观察图 3 中的可视化结果, 并

表 2　本文和 CAL[18]方法在 Market1501 数据集上的检索结果对比

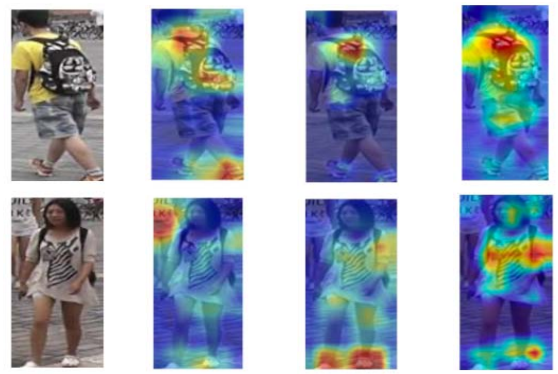
方法	查询图像	检索结果
本文		
		
		
		
CAL[18]		
		
		
		

对比不同视角、不同姿态、衣服配饰差异，可以看出，与 Baseline 和 SPAM 相比, DPAM 能够准确地定位与身份相关区域，捕捉到不同部分区域的显著特征，如人脸、书包、衣服、鞋子和配饰等. 由此可以证明，DPAM 能够提取到更具鲁棒性的可辨别特征，能够有效地应对复杂情况下的图像识别问题.

表 3 不同模块在 Market1501 实验结果比较 %

不同模块组合	mAP	Rank-1	Rank-5	Rank-10
Baseline	87.5	94.8	98.3	99.1
Baseline+SPAM	88.6	95.0	98.4	99.0
Baseline+DPAM	89.2	95.3	98.3	99.1
Baseline+DPAM+EAFFM	**89.3**	**95.7**	**98.6**	**99.4**

注. 粗体表示最优值.



a. 原始图片 b. Baseline c. Baseline+SPAM d. Baseline+DPAM

图 3 Grad-CAM[27]注意特征提取可视化

由于浅层网络多关注图像的边缘信息，深层网络关注图像的细节信息，因此本文在前 3 层分别加入 SE 模块，从通道域的角度赋予图像不同位置的不同权重，激励有用特征，抑制无用特征，从而将获得的更重要的边缘特征信息传递到深层网络以提取关键特征. 然而，注意力机制忽略了非显著潜在特征的重要性. 本文通过在 ResNet50_conv3 进行区域划分，将 SPAM 重构为 DPAM，分别在强弱通道区域提取关键细节特征. 为了更好地说明在 ResNet50_conv3 进行区域划分对于 DPAM 的优越性，本文对剩余层进行了分区实验(未加入反事实干扰)，实验结果如表 4 所示，表中列出了在不同层卷积块分区下的实验评价指标 mAP 和 Rank-1 的数值；同时，图 4 展示了 mAP 和 Rank-1 精度随分区层级变化的曲线图. 综合表 4 和图 4 的结果可以得出，在第 3 层卷积块进行分区时，可以获得较高的 mAP 和 Rank-1 精度，层级较低或较高都不利于识别语义线索的发现，验证了在 ResNet50_conv3 进行区域划分为最优方案.

表 4 不同层区域划分在 Market1501 实验结果比较 %

不同层区域划分	mAP	Rank-1	Rank-5	Rank-10
DPAM_conv2	88.6	94.8	98.3	99.1
DPAM_conv4	88.1	94.8	**98.4**	98.9
DPAM_conv3	**89.2**	**95.3**	98.3	**99.1**

注. 粗体表示最优值.

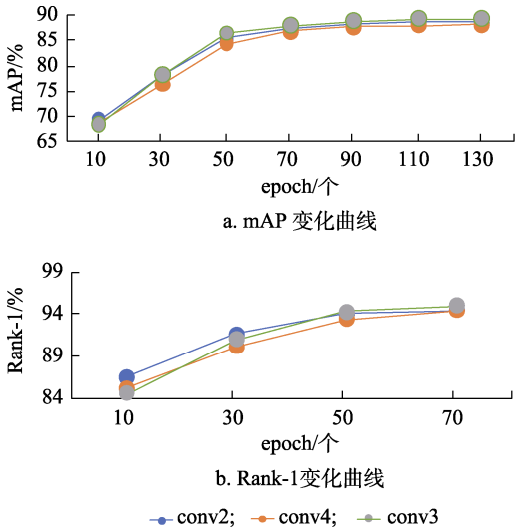


图 4 Market1501 上不同层区域划分的 mAP 和 Rank-1 的变化曲线

同时，本文将加入 SE 和 CA 的表现与 Baseline 模型进行了比较，结果如表 5 所示. 与加入 SE 相比，在第 3 层和第 5 层卷积块加入 CA，可以获得更高精度的 mAP 和 Rank-1，说明加入 CA 后的模型能够提取更为准确的行人关键特征，此方案更优.

表 5 分别加入 SE 和 CA 后的 DPAM 在 Market1501 实验结果比较 %

不同注意力方案	mAP	Rank-1	Rank-5	Rank-10
Baseline	87.5	94.8	98.3	99.1
Baseline+DPAM(SE)	88.8	95.0	98.5	99.0
Baseline+DPAM(CA)	**89.2**	**95.3**	**98.3**	**99.1**

注. 粗体表示最优值.

2.4.2 EAFFM 分析

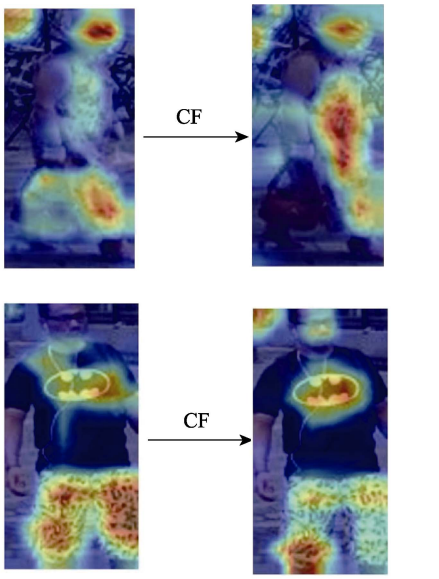
EAFFM 主要对 ResNet_DPAM 输出的双区域注意特征采用增强注意策略，进行空间位置交互和权重学习，完成信息互补，得到语义信息丰富且表征能力强的特征表示(semantic enhancement feature, SEF)，并通过反事实干预生成的反事实特征(counterfactual feature, CF)指导模型习得更高质量的注意力特征图. 为了展示 CF 和 SEF 对模型识别结果的影响效果，本文在 Market1501 数据集上进行了 3 组实验，实验结果如表 6 所示. 根据表 6 的数据可以发现，与简单通道拼接相比，采用 SEF 作为最后特征表示的 mAP 能够提升 0.2 个百分点；加入 CF 干预后，mAP 和 Rank-1 分别提升了 0.3 和 0.5 个百分点，由此证明加入 SEF 和 CF 是有效的. 此外，本文增加了 SEF 和 SEF 在 CF 指导模型后的热力图对比如图 5 所示，加入 CF 后，模型能够获

得更为准确的可辨别特征，有效地提高了已习得注意特征的质量.

表 6　不同特征表示在 Market1501 的实验结果比较 %

特征表示	mAP	Rank-1	Rank-5	Rank-10
DPANet	89.0	95.2	98.4	99.1
DPANet(SEF)	89.2	95.3	98.3	99.1
DPANet(SEF+CF)	**89.3**	**95.7**	**98.6**	**99.4**

注. 粗体表示最优值.



a. SEF 特征热力图　　b. SEF+CF 特征热力图

图 5　SEF 和在 CF 指导模型后的 SEF 特征热力图对比

3　结　语

本文提出了一种基于双路注意力特征提取的 Re-ID 方法，首先，对 ResNet50 网络通过进行区域划分、引入注意力机制将其重构为双路注意力结构，突出非显著潜在可辨别特征的重要性，获得更具判别力的特征表示. 其次，对强、弱通道区域特征采用增强注意策略进行可学习权重融合，实现显著关键特征和潜在可辨别特征的信息互补，并引入反事实干预，增强模型输出的特征质量. 最后在 Re-ID 领域中的 3 个经典数据集上进行了仿真实验，证明了提出方法的可行性和有效性. 但是由于本文方法重点在于优化全局关键特征上，忽略局部关键特征的重要性，因此对于模糊且相似性较大的图像，会出现匹配错误的问题. 针对此问题，后续工作将考虑全局关键特征和局部关键特征的关系并有效地融合，在不增加模型复杂度的情况下，进一步提高模型的识别精度.

参考文献(References):

[1] Luo Hao, Jiang Wei, Fan Xing, *et al*. A survey on deep learning based person re-identification[J]. Acta Automatica Sinica, 2019, 45(11): 2032-2049(in Chinese)
(罗浩，姜伟，范星，等. 基于深度学习的行人重识别研究进展[J]. 自动化学报, 2019, 45(11): 2032-2049)

[2] Dalal N, Triggs B. Histograms of oriented gradients for human detection[C] //Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2005: 886-893

[3] Lowe D G. Object recognition from local scale-invariant features[C] //Proceedings of the 7th IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 1999: 1150-1157

[4] Ojala T, Pietikainen M, Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002, 24(7): 971-987

[5] Köstinger M, Hirzer M, Wohlhart P, *et al*. Large scale metric learning from equivalence constraints[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2012: 2288-2295

[6] Weinberger K Q, Saul L K. Distance metric learning for large margin nearest neighbor classification[J]. The Journal of Machine Learning Research, 2009, 10: 207-244

[7] Liao S C, Hu Y, Zhu X Y, *et al*. Person re-identification by local maximal occurrence representation and metric learning[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2015: 2197-2206

[8] Pedagadi S, Orwell J, Velastin S, *et al*. Local fisher discriminant analysis for pedestrian re-identification[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2013: 3318-3325

[9] Li W, Zhao R, Xiao T, *et al*. DeepReID: deep filter pairing neural network for person re-identification[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2014: 152-159

[10] Sun Y F, Zheng L, Yang Y, *et al*. Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline)[C] //Proceedings of the 15th European Conference on Computer Vision. Heidelberg: Springer, 2018: 501-518

[11] Luo H, Jiang W, Zhang X, *et al*. Alignedreid++: dynamically matching local information for person re-identification[J]. Pattern Recognition, 2019, 94: 53-61

[12] 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] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 907-915

[13] Wang G S, Yuan Y F, Chen X, *et al*. Learning discriminative features with multiple granularities for person re-identificat-

ion[C] //Proceedings of the 26th ACM International Conference on Multimedia. New York: ACM Press, 2018: 274-282

[14] Park H, Ham B. Relation network for person re-identification[C] //Proceedings of the 34th AAAI Conference on Artificial Intelligence. Palo Alto: AAAI Press, 2020: 11839-11847

[15] Woo S, Park J, Lee J Y, *et al*. CBAM: convolutional block attention module[C] //Proceedings of the 15th European Conference on Computer Vision. Heidelberg: Springer, 2018: 3-19

[16] Chen X S, Fu C M, Zhao Y, *et al*. Salience-guided cascaded suppression network for person re-identification[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 3297-3307

[17] Zhang Z Z, Lan C L, Zeng W J, *et al*. Relation-aware global attention for person re-identification[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2020: 3183-3192

[18] Rao Y M, Chen G Y, Lu J W, *et al*. Counterfactual attention learning for fine-grained visual categorization and re-identification[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2021: 1005-1014

[19] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2018: 7132-7141

[20] Hou Q B, Zhou D Q, Feng J S. Coordinate attention for efficient mobile network design[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2021: 13708-13717

[21] Zheng F, Deng C, Sun X, *et al*. Pyramidal person re-identification via multi-loss dynamic training[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2019: 8506-8514

[22] Chen G Y, Gu T P, Lu J W, *et al*. Person re-identification via attention pyramid[J]. IEEE Transactions on Image Processing, 2021, 30: 7663-7676

[23] Zhong Z, Zheng L, Cao D L, *et al*. Re-ranking person re-identification with k-reciprocal encoding[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 3652-3661

[24] Zhou K Y, Yang Y X, Cavallaro A, *et al*. Omni-scale feature learning for person re-identification[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2019: 3701-3711

[25] Zhu H W, Ke W J, Li D, *et al*. Dual cross-attention learning for fine-grained visual categorization and object re-identification[C] //Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2022: 4682-4692

[26] Luo H, Gu Y Z, Liao X Y, *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. Los Alamitos: IEEE Computer Society Press, 2019: 1487-1495

[27] Selvaraju R R, Cogswell M, Das A, *et al*. Grad-CAM: visual explanations from deep networks via gradient-based localization[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2017: 618-626