DOI:10.19651/j.cnki.emt.2519815

基于动态特征增强和分层门控融合的遮挡行人重识别*

任鹏霏 杨真真
(南京邮电大学理学院 南京 210023)

摘 要:行人重识别作为智能监控与智慧城市建设的核心支撑,在实际场景中常因遮挡问题导致识别精度不佳。现有基于卷积神经网络的行人重识别方法受限于局部感受野,难以捕捉跨遮挡区域的长距离依赖。基于 Transformer 的行人重识别方法虽具备全局建模能力,但存在局部与全局特征融合不足等问题,在严重遮挡场景下鲁棒性欠佳。针对上述问题,提出了一种基于动态特征增强和分层门控融合的端到端遮挡行人重识别方法,通过动态特征增强模块优化中层特征的局部细节与抗噪能力,借助分层多尺度门控融合模块缓解高层特征的语义稀释,构建"中层增强-高层提纯"的端到端特征处理链路。仿真实验对比了所提方法与现有方法在 Occluded-Duke、Occluded-ReID、Market1501 和 MSMT17 数据集上的识别性能,Rank-1 准确率分别达到了 74.8%、88.8%、96.7%和 90.9%,mAP 精度分别为 67.0%、86.3%、93.8%和 77.6%,验证了其在遮挡场景下的有效性与优越性。

关键词:行人重识别;遮挡;特征增强;Swin-Transformer;门控注意力;特征融合
中图分类号:TP391.41;TN911.7 **文献标识码**:A **国家标准学科分类代码**:510.40

Dynamic feature enhancement and hierarchical gated fusion for occluded person re-identification

Ren Pengfei Yang Zhenzhen
(College of Science, Nanjing University of Posts and Telecommunications, Nanjing 210023, China)

Abstract: Person re-identification, a core pillar of intelligent surveillance and smart city development, often exhibits significant accuracy degradation in real-world scenarios due to occlusion. Existing convolutional neural network-based person re-identification methods are constrained by local receptive fields, making long-range dependency capture across occluded regions difficult. Transformer-based person re-identification methods, despite global modeling capabilities, suffer from insufficient local-global feature fusion, leading to poor robustness in severely occluded scenarios. An end-to-end occluded person re-identification method based on dynamic feature enhancement and hierarchical gated fusion is proposed to tackle these problems. It employs a dynamic feature enhancement module to optimize local details and noise resistance of mid-level features, and a hierarchical multi-scale gated fusion module to mitigate semantic dilution in high-level features, constructing an end-to-end feature processing pipeline of " mid-level enhancement-high-level purification". The proposed method is compared with existing methods on Occluded-Duke, Occluded-ReID, Market1501 and MSMT17 datasets. Experimental results show Rank-1 accuracies of 74.8%, 88.8%, 96.7% and 90.9%, with mAP accuracies of 67.0%, 86.3%, 93.8% and 77.6%, respectively, validating its effectiveness and superiority in occluded scenarios.

Keywords: person re-identification;occlusion;feature enhancement;Swin-Transformer;gated attention;feature fusion

0 引 言

行人重识别(person re-identification, ReID)[1-5]作为跨非重叠摄像头实现人员追踪的核心技术,是全域安全防控网络的重要技术支撑。然而,现实监控环境存在如下挑战:1)光照变化、视角差异等环境因素导致特征分布剧烈波动;2)遮挡干扰使得服饰纹理、肢体轮廓等关键特征缺失,使基于全局特征的传统方法精度大幅下降,因此,遮挡行人重识别已成为该领域亟需突破的瓶颈[6-7]。

针对遮挡行人重识别,现有研究主要分为基于卷积神

收稿日期:2025-09-09
*基金项目:国家自然科学基金(62571269)、江苏省研究生科研与实践创新计划项目(KYCX24_1125, SJCX24_0279)资助

经网络(convolutional neural network, CNN)和基于视觉Transformer(vision transformer, ViT)两类方法。基于CNN的方法通过局部特征分块或姿态引导应对遮挡。例如,基于局部卷积基线(part-based convolutional baseline, PCB)[8]将特征图水平分割为多个块,通过区域分类器捕捉局部信息;姿态引导特征对齐(pose guided feature alignment, PGFA)[9]利用姿态估计模型生成注意力图,引导网络关注未遮挡区域。然而,这类方法存在固有缺陷:手工分割(如PCB)易丢失边缘信息,提取的特征粒度粗糙,难以适配复杂遮挡尺度;Zhu等[10]提出的一种身份引导的人体语义解析方法(identity-guided human semantic parsing, ISP),虽能定位关键区域,但辅助网络的引入会引发跨域适配问题,导致性能受限。更重要的是,CNN的局部感受野特性使其难以捕捉跨遮挡区域的长距离依赖(如被遮挡行人头部与脚部的关联),导致特征表示不完整。而Transformer通过全局建模能力捕捉图像中长距离特征关联,在姿态变化、光照差异及遮挡场景中表现出更优的特征表示能力。He等[11]首次将Transformer应用于ReID,通过图像分块序列建模全局语义,显著提升了遮挡场景的匹配精度;Zhang等[12]提出的动态补丁感知增强Transformer(dynamic patch-aware enrichment transformer, DPEFormer),借助动态补丁选择模块和特征融合模块,在无需外部检测器的情况下实现了精准的局部特征提取,并结合真实遮挡增强策略,有效提升了模型的鲁棒性,在遮挡与整体ReID基准测试中显著超越了现有方法。Yuan等[13]提出的无训练特征中心化ReID(training-free feature centralization ReID, Pose2ID)框架,基于特征中心化理念,首次实现了无训练条件下的高性能行人重识别,在Market1501、SYSU-MM01等基准数据集以及Occluded-ReID遮挡场景下表现卓越,为轻量化、通用化ReID系统开辟了新路径。Swin-Transformer[14]作为Transformer的重要分支,通过窗口自注意力与分层处理机制,在降低计算复杂度的同时兼顾局部特征提取与全局信息整合,在图像分类、分割等视觉任务中表现卓越。然而,其在行人重识别领域的应用仍较有限,尤其针对遮挡场景的专项优化不足,既未充分利用其分层特征优势捕捉遮挡区域的细粒度细节,也缺乏对局部与全局特征的动态融合策略,导致在严重遮挡下的性能提升受限。

为解决上述问题,本文在Swin-Transformer网络模型的基础上,提出了一种基于动态特征增强和分层门控融合(dynamic feature enhancement and hierarchical gated fusion, DFE-HGF)的遮挡行人重识别方法以提高识别效果。主要创新点如下:

1)针对Swin-Transformer中层特征易受局部遮挡和背景噪声污染的问题,设计了动态特征增强模块(dynamic feature reinforcement module, DFRM),采用动态归一化定向抑制噪声区域、双路径加权融合平衡去噪与细节保留,并通过多层深度可分离卷积覆盖不同遮挡尺度特征需求。突破了传统静态特征处理的局限,实现对动态遮挡场景的自适应特征净化与增强,弥补了单一尺度卷积在局部细节捕捉上的不足。

2)针对高层特征全局语义易被冗余信息覆盖的问题,构建了分层多尺度门控融合模块(hierarchical multi-scale gated fusion, HMSGF),通过多阶门控聚合适配不同遮挡尺度的语义关联,通道拆分专项融合分别聚焦局部细节与全局语义,并结合通道-空间双维注意力精准调控有效特征通过层级化提纯与双维度协同机制,缓解高层语义稀释与细节丢失,提升全局特征判别性。

3)仿真实验,实验结果显示在Occluded-Duke和Occluded-ReID两个遮挡专用数据集上,Rank-1准确率分别达到74.8%和88.8%,mAP精度分别为67.0%和86.3%;在Market1501和MSMT17两个通用数据集上,Rank-1准确率分别为96.7%和90.0%,mAP精度分别为93.8%和77.6%,显著优于现有主流行人重识别方法,充分证明了提出的DFE-HGF框架的有效性与优越性。

1 相关工作

1.1 Swin-Transformer网络

Transformer架构最初应用于自然语言处理,随后逐渐引入到计算机视觉任务中。Dosovitskiy等[15]提出的ViT将图像分割为固定尺寸的图像块并转换为序列输入,验证了Transformer在图像分类中的有效性。但全连接自注意力机制计算复杂度较高,难以适配高分辨率任务。Liu等[14]提出的Swin-Transformer通过局部窗口注意力降低复杂度,并通过滑动窗口策略在相邻窗口间建立关联,实现局部特征提取与全局信息整合的平衡。其分层化结构在图像分类、目标检测和语义分割任务中表现优异,成为视觉Transformer的主流架构之一。尤其对于行人图像,Swin-Transformer的窗口注意力天然适配局部-全局特征分布,分层输出的多尺度特征可捕捉边缘纹理、部件结构与整体语义,为遮挡场景下的特征学习提供潜力。

尽管Swin-Transformer建模能力强,但在遮挡场景下仍过度关注被遮挡区域,且难以保持跨尺度特征一致性。因此,构建适用于遮挡场景的模块,以进一步提升其主干的局部建模能力,成为研究的重点。

1.2 门控注意力机制

注意力机制通过动态分配权重聚焦关键信息,是提升视觉模型判别能力的核心工具。其中,门控注意力通过可学习的门控结构控制信息流动,在特征筛选与噪声抑制方面优势显著,尤其适用于遮挡行人重识别等需精准定位有效区域的任务。早期注意力机制的研究以通道或空间单维度建模为主。通道压缩激励网络(squeeze-and-excitation network, SENet)[16]通过压缩-激励机制捕获通道依赖、生成通道权重;卷积块注意力模块(convolutional block

attention module，CBAM)[17]融合空间注意力分支，实现通道-空间联合调控。这类方法虽能提升特征判别性，但权重主要依赖全局统计信息，对遮挡边界等动态局部场景的适应性不足。门控机制通过引入非线性激活函数生成“开关”信号，动态控制特征通过比例，在复杂场景中表现更优。在 Transformer 架构中，用门控机制优化自注意力输出，通过抑制噪声标记的权重提升特征纯度。例如，门控注意力 Transformer（gated attention transformer，GAT)[18]在多人姿势跟踪任务中，通过门控机制动态调节注意力层中外观嵌入与时间姿势相似性的权重，从而有效抑制运动模糊、拥挤场景或遮挡区域带来的干扰。

现有 Re-ID 方法虽利用门控机制提升了遮挡鲁棒性，但仍存在如下局限：1）门控权重多基于单一特征尺度生成，难以适配不同遮挡尺度；2）通道与空间门控的协同不足，未能充分挖掘跨维度的互补信息。为此，本文提出通道-空间门控增强模块（channel spatial gated enhancement，CSGE)，整合空间门控与 SE 注意力机制，结合局部卷积与残差增强策略，实现对遮挡区域的动态调控与有效信息融合，进一步提升模型在复杂遮挡条件下的特征判别能力。

2　研究方法

2.1　网络构架

针对遮挡场景下行人特征易受噪声干扰、判别性下降的问题，本文提出了一种基于动态特征增强和分层门控融合(DFE-HGF)的端到端遮挡行人重识别网络，以多阶段特征提取为基础，通过动态特征增强与分层特征提纯的协同处理，实现对遮挡敏感特征的精准捕捉与区分。如图 1 所示，架构以 Swin-Transformer 为骨干网络，通过 4 阶特征提取（stage1～stage4）将输入的行人图像转换为多尺度特征张量，其中，stage2 输出的中层特征 $\boldsymbol{X}_l \in \mathbf{R}^{B\times C\times H\times W}$（$B$ 为批量大小，C 为通道数，H、W 为特征图尺寸）送入动态特征增强模块；stage4 输出的高层特征 $\boldsymbol{X}_h \in \mathbf{R}^{B\times C\times H\times W}$ 送入分层多尺度门控融合模块。

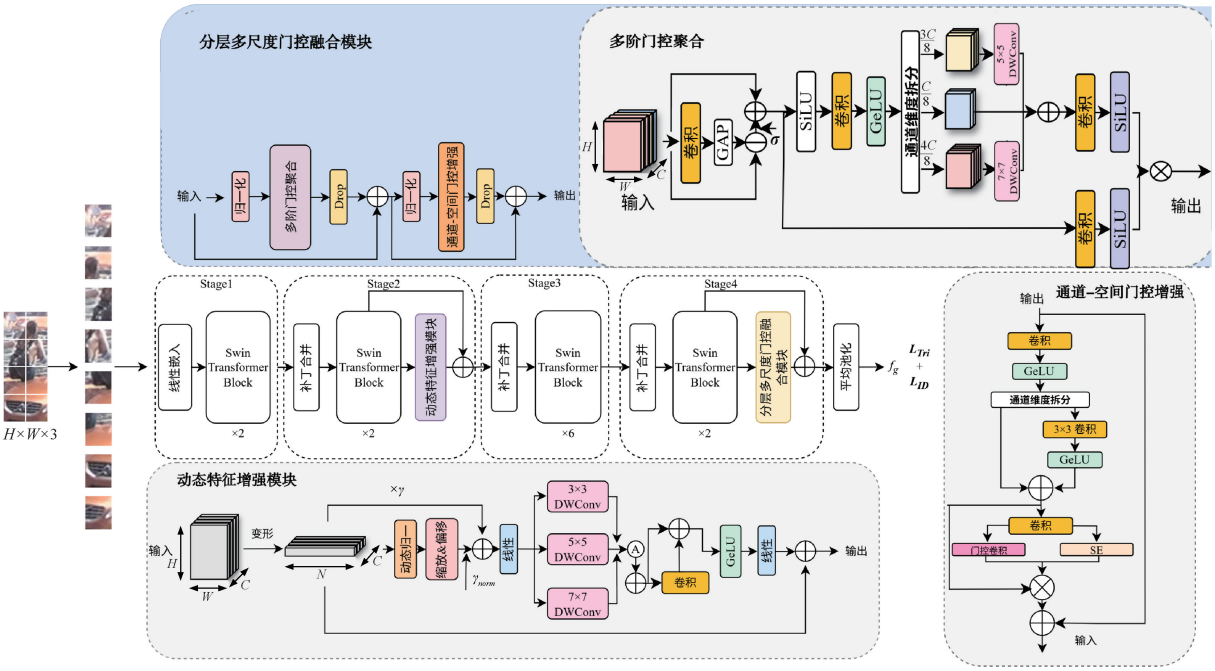

图 1　DFE-HGF 整体架构

Fig. 1　Overall architecture of DFE-HGF

动态特征增强模块针对中层特征 $\boldsymbol{X}_l$ 采用“去噪-增强-补强”三阶机制。通过动态归一化抑制噪声并保留关键特征，随后结合原始特征 $\boldsymbol{X}_l$ 进行加权融合以平衡去噪与细节，生成增强特征 $\boldsymbol{X}_{fusion}$；最后，3 个并行的深度可分离卷积捕捉不同遮挡尺度的局部特征，经平均融合与残差连接，输出最终中层增强特征 $\boldsymbol{X}_{enh}$，强化对局部遮挡残留细节的敏感性。

分层多尺度门控融合模块针对高层特征 $\boldsymbol{X}_h$ 实现分层提纯。将通过多阶门控聚合输出的初步聚合特征 $\boldsymbol{X}^1_{fusion}$ 映射至高维空间，并按固定比例拆分为细节子通道（局部结构）与主干子通道（全局语义），通过拼接投影形成“局部-全局”互补特征 $\boldsymbol{X}_{Fusion}$；最后通过通道-空间门控增强，输出判别性更强的高层特征 $\boldsymbol{X}_{Enh}$。

2.2　动态特征增强模块

Swin-Transformer 中 stage2 输出的中层特征包含丰富的局部纹理和空间细节，但易受局部遮挡和背景噪声影响。传统静态增强方法（如固定归一化、单一尺度卷积）难适配遮挡动态分布，容易压制有效特征或放大冗余。动态

特征增强模块针对遮挡场景下中层特征易受噪声污染及尺度适配性不足的核心问题,构建“去噪-增强-补强”3阶递进处理机制,通过动态特征调整与多尺度融合协同,提升局部特征的判别稳定性。

去噪(动态归一化机制):突破传统层归一化(LayerNorm)的静态统计范式,通过可学习参数$\alpha \in \mathbf{R}^{+}$对特征激活值进行差异化调控,实现对噪声的定向清理:Stage2输出中层特征张量$\boldsymbol{X}_l$,通过可学习参数α对特征进行非线性缩放,形成动态激活特征:

$$\boldsymbol{X}_{\alpha} = \tanh(\alpha \boldsymbol{X}_l) \tag{1}$$

其中,$\tanh(\cdot)$为tanh激活函数,α是可学习参数。引入可学习的通道权重$\boldsymbol{w} \in \mathbf{R}^{C}$和偏置$\boldsymbol{b} \in \mathbf{R}^{C}$,对$\boldsymbol{X}_{\alpha}$进行逐通道校准:

$$\boldsymbol{X}_{norm} = \boldsymbol{w} \odot \boldsymbol{X}_{\alpha} + \boldsymbol{b} \tag{2}$$

其中,$\odot$表示逐元素乘法,$\boldsymbol{w}$和$\boldsymbol{b}$分别用于增强判别性通道和抑制噪声通道。最终输出的归一化特征张量为$\boldsymbol{X}_{norm} \in \mathbf{R}^{B\times C\times H\times W}$,该张量为后续双路径融合提供去噪后的特征基础。

增强(双路径动态加权融合):对去噪特征与原始特征进行自适应加权,其加权公式为:

$$\boldsymbol{X}_{fusion} = \gamma_{norm} \boldsymbol{X}_{norm} + \gamma \boldsymbol{X}_l \tag{3}$$

其中,$\gamma_{norm} \in \mathbf{R}^{+}$和$\gamma \in \mathbf{R}^{+}$为可学习参数,通过梯度下降协同学习,实现“去噪不丢细节”的特征融合。最终输出的融合特征张量$\boldsymbol{X}_{fusion} \in \mathbf{R}^{B\times C\times H\times W}$,既继承了动态归一化对遮挡噪声的抑制能力,又保留了原始特征中关键的局部细节,为后续多尺度卷积聚合提供了更鲁棒的中层特征。

补强(多尺度卷积聚合):融合特征$\boldsymbol{X}_{fusion}$并行输入3种不同尺寸卷积核的深度可分离卷积层,其中,3×3卷积聚焦细微纹理与局部轮廓,捕捉小范围遮挡细节;5×5卷积扩大感受野,覆盖中等范围遮挡并提取上下文关联;7×7卷积覆盖更广泛的上下文信息,有效建模大尺度遮挡场景下的全局语义关联。其公式为:

$$\boldsymbol{X}_{dw}^{k} = \dot{\otimes}(\boldsymbol{X}_{fusion}, \boldsymbol{\omega}_{dw}^{k}) \tag{4}$$

其中,$\dot{\otimes}$为逐通道卷积操作,$k = 3,5,7$为卷积核尺寸。$\boldsymbol{\omega}_{dw}^{k} \in \mathbf{R}^{C_{in}^{D}\times k\times k\times C_{out}^{D}}$为$k\times k$尺寸的逐通道卷积核。并行输出的特征通过聚合平均,并与原始特征残差连接后经特征映射,形成最终增强特征:

$$\boldsymbol{X}_{pw}^{k} = \otimes(\boldsymbol{X}_{dw}^{k}, \boldsymbol{\omega}_{pw}^{k}) \tag{5}$$

$$\boldsymbol{X}_{enh} = \otimes\left(\boldsymbol{X}_{fusion} + avg\left(\sum_{k=3,5,7} \boldsymbol{X}_{pw}^{k}\right), \boldsymbol{\omega}_{proj}\right) \tag{6}$$

其中,$\otimes$为逐点卷积,$avg(\cdot)$为取平均操作,$\boldsymbol{\omega}_{pw}^{k}$为逐点卷积核,$\boldsymbol{\omega}_{proj}$为通道投影的1×1卷积核。

2.3 分层多尺度门控融合模块

Stage4输出的高层语义特征包含行人的全局信息,但在严重遮挡下存在两大瓶颈:1)语义稀释,关键身份特征易被冗余区域或遮挡部分干扰,导致判别性下降;2)细节丢失,高层特征的空间分辨率较低,使其难以保留遮挡残留的局部线索。传统的多尺度融合方法缺乏对空间与通道的精细调控,难以区分“有效语义”与“遮挡噪声”。

为此,分层多尺度门控融合模块提出“通道拆分-双维注意力-多阶聚合”架构,实现高层语义特征分层化提纯。模块首先对高层语义特征$\boldsymbol{X}_h$进行多阶门控聚合处理:通过特征分解操作分离全局与局部分量,1×1卷积映射后经全局平均池化提取全局上下文表示,与原始特征差分增强并引入可学习缩放因子δ,以加强局部区域的响应能力:

$$\boldsymbol{X}_{d} = SiLU(\boldsymbol{X}_{proj} + \delta(\boldsymbol{X}_{proj} - \boldsymbol{X}_{avg})) \tag{7}$$

其中,$SiLU(\cdot)$为SiLU激活函数,且$\boldsymbol{X}_{proj} = \otimes(\boldsymbol{X}_h, \boldsymbol{\omega}_{proj})$,$\boldsymbol{X}_{avg} = avg(\boldsymbol{X}_{proj})$。

此外,为适应遮挡区域在尺度与位置上的差异,将$\boldsymbol{X}_d$通道按照$C_1 = \lfloor C/8 \rfloor$,$C_2 = \lfloor 3C/8 \rfloor$,$C_3 = C - C_1 - C_2$比例拆分:$\boldsymbol{X}_d^1 \in \mathbf{R}^{C_1\times B\times H\times W}$感受野聚焦局部未遮挡细节;$\boldsymbol{X}_d^2 \in \mathbf{R}^{C_2\times B\times H\times W}$经5×5深度可分离卷积处理,感受野覆盖局部遮挡边缘;$\boldsymbol{X}_d^3 \in \mathbf{R}^{C_3\times B\times H\times W}$经7×7深度可分离卷积处理。3部分特征拼接后经1×1卷积融合,得到多阶特征,其公式为:

$$\boldsymbol{F}_i = \begin{cases} \boldsymbol{X}_d^1, & i = 1 \\ DSC_{k\times k}^{d}(\boldsymbol{X}_d^i), & i = 2,3 \end{cases} \tag{8}$$

$$\boldsymbol{Z} = Conv_{1\times 1}([\boldsymbol{F}_1 \,\|\, \boldsymbol{F}_2 \,\|\, \boldsymbol{F}_3]) \tag{9}$$

其中,$DSC_{k\times k}^{d}(\cdot)$表示带膨胀率$d$的$k\times k$深度可分离卷积算子,$\|$表示通道拼接操作,$Conv_{1\times 1}(\cdot)$为1×1卷积。将其与$\boldsymbol{X}_d$逐元素相乘,并通过残差连接输出初步聚合特征$\boldsymbol{X}_{fusion}^{1}$。其公式为:

$$\boldsymbol{X}_{fusion}^{1} = SiLU(\boldsymbol{Z}) \odot SiLU(Conv_{1\times 1}(\boldsymbol{X}_d)) + \boldsymbol{X}_h \tag{10}$$

之后,模块进入“通道拆分-细节融合”阶段,如图2所示,传统特征处理采用“全通道统一操作”,难以兼顾局部细节与全局语义提取。本模块基于“特征功能分工”思想,先经1×1卷积将通道映射至高位特征空间$\hat{\boldsymbol{X}}_{fusion}^{1} \in \mathbf{R}^{2C\times B\times H\times W}$以丰富表达。随后,将特征通道按固定比例($p = 0.25$)进行划分为细节子通道$\boldsymbol{X}_{det}$(占25%通道)与主干子通道$\boldsymbol{X}_{main}$(占75%通道)。细节子通道通过3×3深度可分离卷积提取局部结构,经GELU激活函数形成局部特征;主干子通道保留全局语义信息。两类子通道拼接后经1×1卷积投影回原通道维度,形成“局部-全局”互补特征,其公式为:

$$\boldsymbol{X}_{Fusion} = \otimes([GELU(\hat{\boldsymbol{X}}_{det})) \,\|\, \boldsymbol{X}_{main}], \boldsymbol{\omega}_{proj}) \tag{11}$$

其中,$\hat{\boldsymbol{X}}_{det} = DSC_{3\times 3}(\boldsymbol{X}_{det})$,$DSC_{3\times 3}(\cdot)$为3×3深度可分离卷积,$GELU(\cdot)$为GELU激活函数。该“通道拆分-细节提取-融合”设计可同时兼顾全局稳定性与局部细节敏感性,在遮挡条件下仍能保持较强的局部表征能力,为后续的注意力调控奠定良好基础。

为进一步增强特征的表达能力，引入通道-空间双维度注意力协同调控：空间注意力通过 3×3 深度可分离卷积与 Sigmoid 激活生成空间掩码，动态强化未遮挡关键区域（如头部、手部）并抑制遮挡噪声；通道注意力基于 SE 机制，先经全局平均池化获取通道全局描述子 $\boldsymbol{z}=GAP(\boldsymbol{X}_{Fusion})$ 后，通过两层感知机生成通道权重 $\boldsymbol{M}_c$：

$$\boldsymbol{M}_s=\sigma(DSC_{3\times3}(\boldsymbol{X}_{Fusion})) \tag{12}$$

$$\boldsymbol{M}_c=\sigma(\boldsymbol{W}_2\rho(\boldsymbol{W}_1\boldsymbol{z})) \tag{13}$$

其中，$\boldsymbol{M}_s$ 为生成的空间注意力掩码，$\sigma(\cdot)$ 为 Sigmoid 函数，$\boldsymbol{W}_1\in\mathbf{R}^{C/r\times C}$ 为降维映射矩阵，$\boldsymbol{W}_2\in\mathbf{R}^{C\times C/r}$ 为升维映射矩阵，r 为缩放系数，$\rho(\cdot)$ 为 ReLU 激活函数。最终，双维度注意力通过残差连接联合调控输出：

$$\boldsymbol{X}_{Enh}=\boldsymbol{X}_{Fusion}\odot\boldsymbol{M}_s\odot\boldsymbol{M}_c+\boldsymbol{X}_{Fusion} \tag{14}$$

该机制在遮挡复杂场景下有效突出关键特征、抑制冗余信息，实现高层语义特征的分层提纯与判别性增强，为后续分类和度量任务提供鲁棒高层表示。

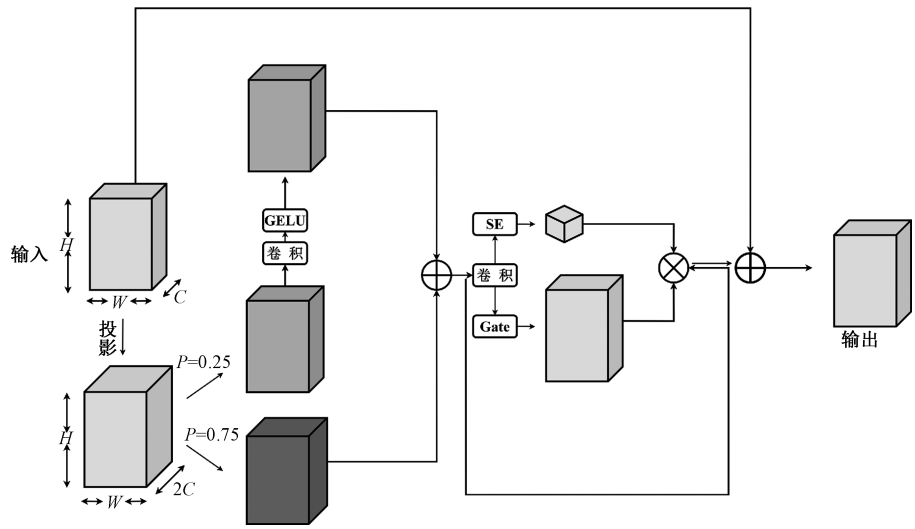

图 2　通道-空间门控增强模块

Fig. 2　Channel-spatial gated enhancement module

2.4　损失函数

本文采用多损失函数协同优化策略，结合带标签平滑的交叉熵损失和三元组损失。带权重的联合损失函数同时约束分类与特征学习，与协同优化，实现端到端训练，提高模型在复杂遮挡场景下的识别性能。其公式表示为：

$$L_{total}=\lambda_1 L_{ID}+L_{Tri} \tag{15}$$

$$L_{Tri}=\sum_{i=1}^{N}\left[\|f_i^a-f_i^p\|_2^2-\|f_i^a-f_i^n\|_2^2+m\right]_+ \tag{16}$$

$$L_{ID}=-\sum_{i=1}^{C}\tilde{y}_i\log(p_i) \tag{17}$$

其中，λ_1 为权重系数，$\tilde{y}_i$ 为标签平滑后的分布，p_i 为预测的第 i 类概率，f_i^a、f_i^p、f_i^n 分别为锚点、正样本与负样本的特征表示。

3　实验和结果分析

3.1　数据集和评价标准

为验证所提方法在不同场景下的有效性，实验选了取包含整体行人特征的通用数据集（Market1501、MSMT-17）以及遮挡场景下的两个专用数据集（Occluded-Duke、Occluded-ReID）。各数据集配置如下：Market1501 含 1 501 名行人，由 6 台不同分辨率相机拍摄；MSMT-17 为 15 台摄像机采集的大规模数据集，含 4 101 名行人共 126 441 张图像，其中训练集 32 621 张，测试集 93 820 张；Occluded-Duke 含训练集 15 618 张、查询集 2 210 张、图库集 17 661 张；Occluded-ReID 含 200 名行人，每人 5 张全身图与 5 张不同遮挡程度图。

评估指标采用行人重识别领域广泛认可的平均精度均值（mAP）与累积匹配曲线（CMC）。mAP 通过综合所有类别的平均精度并加权平均获得，反映算法的整体检索精度；CMC 以 Rank-n 准确率曲线形式呈现，直观评估不同排序位置下的识别性能。

3.2　实验设置

实验基 PyTorch 实现，在 GTX3090 上执行，以 Swin-Transformer 为骨干网络，在 ImageNet21K 上进行初始权值预训练，然后在 ImageNet1K 上进行微调。将输入图像的大小调整为 384×128。采用随机梯度下降（stochastic gradient descent，SGD）作为优化器，训练批大小为 64，初始学习率 0.008，以余弦方式衰减。

3.3　参数分析

在训练和测试阶段设置参数 α，其对 Occluded-Duke 数据集的影响如图 3 所示。实验表明，当 $\alpha=0.6$ 时模型性能最优。因此本文设置 $\alpha=0.6$。

3.4　对比试验

为验证所提 DFE-HGF 在遮挡场景下的性能优势，选

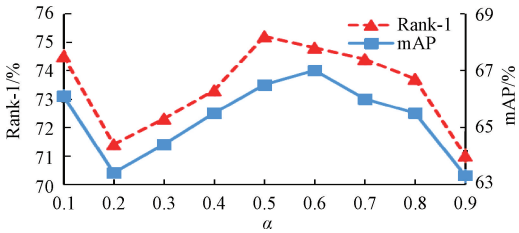

图 3 可学习参数 α 的参数分析

Fig. 3 Parameter analysis of learnable parameter α

取当前主流方法在 Occluded-Duke 和 Occluded-ReID 两个遮挡数据集上进行定量对比。对比方法涵盖基于姿态信息(HoReID、PGFA、PVPM)、手工分割(PCB)、Transformer 架构(TransReID、PFD、PAT)以及近年新方法(Solider、OPR-DAAO、RFMT),其中 HoReID、PCB、TransReID、PFD、Solider 和 SPT-main 的结果都在同一服务器上实现,其余结果引用自原论文。具体对比结果如表 1 所示。

表 1 在遮挡数据集上不同模型结果

Table 1 Results of different models on occluded datasets

%

算法	Occluded-Duke		Occluded-ReID	
	Rank-1	mAP	Rank-1	mAP
PCB[8]	41.2	33.4	40.9	36.5
PGFA[9]	51.4	37.3	—	—
TransReID[12]	67.1	59.6	81.5	76.2
HoReID[19]	54.9	43.6	79.3	70.1
BPBreID[20]	66.7	54.1	76.9	68.6
PVPM[21]	47.0	37.7	70.4	61.2
PFD[22]	66.3	59.1	80.5	76.7
Solider[23]	72.0	65.0	83.8	82.0
SPT-main[24]	62.1	54.3	83.5	78.6
PAT[25]	64.5	53.6	81.6	72.1
RTGAT[26]	61.0	50.1	71.8	51.0
DFE-HGF	**74.8**	**67.0**	**88.8**	**86.3**

实验结果表明,基于姿态信息的方法(HoReID、PGFA)受跨域适配限制,性能提升有限;手工分割方法(如 PCB)因特征粒度粗糙,适应性最差;Transformer 类方法(如 TransReID、PFD)虽优于 CNN,但局部-全局特征融合不足。近年来,Solider 在遮挡场景表现较优。相比之下,本文提出的 DFE-HGF 在 Occluded-Duke 和 Occluded-ReID 上的 mAP 分别达到 67.0%和 86.3%,均超过现有方法,验证了其在复杂遮挡下的优势。

为了验证 DFE-HGF 方法的普适性,同时在 Market-1501 和 MSMT-17 两个完整数据集上进行了仿真实验,结果如表 2 所示。

表 2 在完整数据集上不同模型结果

Table 2 Results of different models on complete datasets

%

算法	Market-1501		MSMT-17	
	Rank-1	mAP	Rank-1	mAP
PCB[8]	92.1	76.6	—	—
PGFA[9]	91.2	76.8	63.5	35.7
TransReID[12]	93.2	86.9	83.3	64.9
HoReID[19]	93.2	81.8	—	—
BPBreID[20]	95.1	87.0	—	—
PFD[22]	94.9	88.1	82.3	61.9
Solider[23]	96.6	93.6	90.0	75.6
SPT-main[24]	94.3	87.2	79.5	58.0
CLIP-Reid[27]	95.4	90.5	89.7	75.8
NFormer[28]	93.2	83.7	80.8	62.2
DRL-Net[29]	94.7	86.9	88.1	76.6
DFE-HGF	**96.7**	**93.8**	**90.9**	**77.6**

在完整数据集上,DFE-HGF 同样表现优异。在 Market-1501 上,Rank-1 和 mAP 分别为 96.7%和 93.8%,超过 CNN 基线(PCB)、Transformer 方法(TransReID),并较 Solider 略有提升。在更具挑战性的 MSMT17 上,Rank-1 和 mAP 分别为 90.9%和 77.6%,比 Solider 的 mAP 高 2.0%。这表明 DFE-HGF 不仅适用于遮挡场景,也能在常规大规模数据集上保持稳定优势。

3.5 消融实验

提出的 DFE-HGF 由动态特征增强和分层多尺度门控融合两个主要模块组成。为验证所提方法中各核心组件的有效性,通过构建不同变体,在 Occluded-Duke 数据集上进行消融实验,结果如表 3 所示。

表 3 Occluded-Duke 数据集上消融实验结果

Table 3 Ablation experimental results on Occluded-Duke dataset

%

方法	Rank-1	Rank-5	Rank-10	mAP
Baseline	72.0	83.3	87.2	65.0
Baseline-D	73.4	84.2	87.6	65.7
Baseline-H	72.8	84.4	88.1	65.9
Baseline-D-H	**74.8**	**84.9**	**89.1**	**67.0**

实验结果验证了 DFE-HGF 框架中核心模块的有效性,以未引入 DFRM 和 HMSGF 的原始 Swin-Transformer 骨干网络作为 Baseline,其在 Occluded-Duke 数据集上的 Rank-1 为 72.0%,mAP 为 65.0%;仅引入 DFRM 的 Baseline-D 变体,Rank-1 提升至 73.4%,mAP 提升至 65.7%;仅引入 HMSGF 模块的 Baseline-H 变体,

Rank-1 准确率提升至 72.8%，mAP 提升至 65.9%；同时引入两个模块 Baseline-D-H 变体性能实现显著跃升，Rank-1 为 74.8%（较 Baseline 提升 2.8%），mAP 为 67.0%（提升 2.0%），超过单一模块增益，证明二者在中层增强-高层提纯的协同链路形成互补，显著提升了整体性能。

3.6　可视化实验

为了可视化评估所提出模型的性能，本节展示了在 Occluded-Duke 数据集上的可视化结果，如图 4 所示。图 4 中每一行左侧为查询图像（Query），右侧为对应的 Top-10 检索结果，其中绿色编号表示正确匹配，红色编号（方框标出）表示错误匹配。

图 4　Occluded-Duke 数据集的可视化结果

Fig. 4　Visualization results on the Occluded-Duke dataset

可视化结果显示，Baseline 方法鲁棒性不足，当行人存在外观相似性或被事物遮挡时，误检率显著上升。而 DFE-HGF 的 Top-10 检索结果中正确匹配比例更高，且正确结果大多位于前列。即使在复杂遮挡场景中，仍能稳定聚焦未遮挡关键区域，有效抑制姿态变化、背景噪声及外观相似性的干扰，显著降低误检概率。

此外，在 Occluded-Duke 数据集还进行了类激活映射（class activation mapping，CAM）可视化实验，如图 5 所示。

CAM 可视化进一步说明，Baseline 易受背景或遮挡物的干扰，例如车辆、路牌等区域被错误地激活，导致模型未能充分聚焦在行人的关键部位。而 DFE-HGF 模块能准确地聚焦于背包、头部和躯干等判别性区域，同时抑制对背景的响应。这表明 DFE-HGF 能在复杂遮挡环境下挖掘更加鲁棒的判别特征。

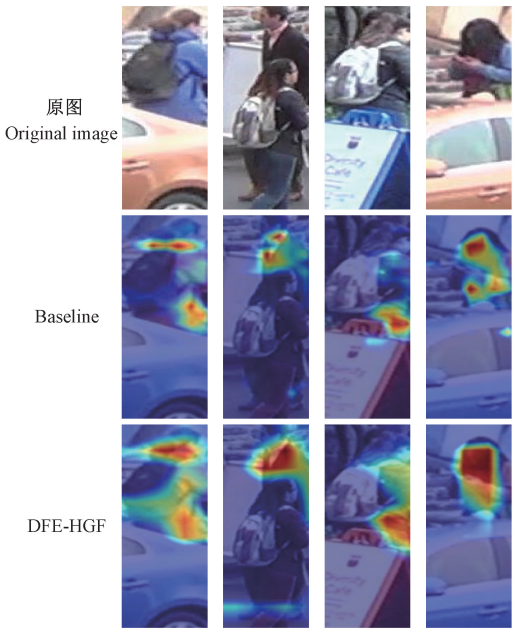

图 5　类激活映射图

Fig. 5　Class activation mapping map

4　结　　论

针对遮挡场景下行人重识别任务中局部特征易受噪声污染、全局语义易被稀释以及多尺度遮挡适配不足等挑战，提出了一种基于动态特征增强和分层门控融合的端到端行人重识别框架，通过动态特征增强模块和分层多尺度门控融合模块的协同设计，实现了对遮挡敏感特征的精准捕捉与判别。未来将进一步探索极端遮挡场景的特征补全技术，并结合轻量化策略推动行人重识别方法在实际监控场景的实际应用。

参考文献

[1]　姬晓飞，赵帅，宋京浩，等. 基于姿势估计和特征融合的行人重识别算法[J]. 电子测量与仪器学报，2024，38(4)：187-194.
JI X F，ZHAO SH，SONG J H. et al. Feature-enhanced based pedestrian re-identification algorithm[J]. Journal of Electronic Measurement and Instrumentation，2024，38(4)：187-194.

[2]　YE M，SHEN J B，LIN G J，et al. Deep learning for person reidentification：A survey and outlook[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2022，44(6)：2872-2893.

[3]　刘明洋，万九卿. 基于深度测量的行人体态特征提取与再识别方法[J]. 仪器仪表学报，2023，44(1)：201-211.
LIU M Y，WAN J Q. Person shape feature extraction and reidentification based on depth measurement[J]. Chinese Journal of Scientific Instrument，2023，44(1)：201-211.

[4]　胡玉玲，王鑫依，张一，等. 一种基于注意力的无监督行人重识别方法[J]. 电子测量技术，2025，48(10)：161-168.

HU Y L, WANG X Y, ZHANG Y, et al. An attention-based unsupervised approach to pedestrian re-identification [J]. Electronic Measurement Technology, 2025,48(10):161-168.

[5] 姬晓飞,孙英超,宋京浩.基于特征增强的行人重识别算法[J].电子测量技术,2025,48(22):198-205.
JI X F, SUN Y CH, SONG J H. Feature-enhanced based pedestrian re-identification algorithm [J]. Electronic Measurement Technology, 2025, 48(22): 198-205.

[6] PENG Y J, WU J L, XU B Q, et al. Deep learning based occluded person re-identification: A survey[J]. ACM Transactions on Multimedia Computing, Communications and Applications, 2023, 20(3): 1-27.

[7] NING EN H, WANG CH SH, ZHANG H, et al. Occluded person re-identification with deep learning: A survey and perspectives[J]. Expert Systems with Applications, 2023, 239:122419.

[8] SUN Y F, 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, 2018: 501-518.

[9] MIAO J X, WU Y, LIU P, et al. Pose-guided feature alignment for occluded person re-identification [C]. 2019 IEEE/CVF International Conference on Computer Vision, 2019: 542-551.

[10] ZHU K, GUO H Y, LIU ZH W, et al. Identity-guided human semantic parsing for person re-identification [C]. Computer Vision-ECCV, 2020, 12348: 346-363.

[11] HE SH T, LUO H, WANG P CH, et al. Transreid: Transformer based object re-identification[C]. 2021 IEEE/CVF International Conference on Computer Vision, 2021:14993-15002.

[12] ZHANG X, FU K R, ZHAO Q J. Dynamic patch-aware enr-ichment transformer for occluded person re-identification[J]. Knowledge-Based Systems, 2025, 327:114193.

[13] YUAN CH, ZHANG G W, MA CH X, et al. From poses to identity: Training-free person re-identification via feature centralization[C]. Computer Vision and Pattern Recognition Conference,2025:24409-24418.

[14] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]. IEEE/CVF International Conference on Computer Vision, 2021:9992-10002.

[15] DOSOVITSKIY A, BEYER L, KOLESNIKOV A, et al. An image is worth 16×16 words: Transformers for image recognition at scale [J]. ArXiv preprint arXiv:2010.11929, 2020.

[16] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. 2018 IEEE Conference on Computer Vision and Pattern Recognition, 2018:7132-7141.

[17] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]. European Conference on Computer Vision, 2018:3-19.

[18] DOERING A, GALL J. A gated attention transformer for multi-person pose tracking[C]. 2023 IEEE/CVF International Conference on Computer Vision, 2023: 3181-3190.

[19] WANG G AN, YANG SH, LIU H Y, et al. High-order information matters: Learning relation and topology for occluded person re-identification [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition,2020: 6448-6457.

[20] SOMERS V, VLEESCHOUWER C, ALAHI A. Body part-based representation learning for occluded person ReIdentification[C]. 2022 IEEE/CVF Winter Conference on Applications of Computer Vision, 2023:1613-1623.

[21] GAO SH, WANG J Y, LU H CH, et al. Pose-guided visible part matching for occluded person reid [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11744-11752.

[22] WANG T, LIU H, SONG P H, et al. Pose-guided feature disentangling for occluded person re-identification based on transformer [C]. AAAI Conference on Artificial Intelligence, 2022, 36(3): 2540-2549.

[23] CHEN W H, XU X ZH, JIA J, et al. Beyond appearance: A semantic controllable self-supervised learning framework for human-centric visual tasks[C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 15050-15061.

[24] TAN L, XIA J ER, LIU W F, et al. Occluded person re-identification via saliency-guided patch transfer[C]. AAAI Conference on Artificial Intelligence, 2024, 38(5): 5070-5078.

[25] LI Y L, HE J F, ZHANG T ZH, et al. Diverse part discovery:Occluded person re-identification with part-aware transformer[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 2897-2906.

[26] HUANG M Y, HOU CH P, YANG Q Y, et al. Reasoning and Tuning: Graph attention network for occluded person re-identification[J]. IEEE Transactions on Image Processing, 2023, 32: 1568-1582.

[27] LI S Y, SUN L, LI Q L. CLIP-ReID: Exploiting vision language model for image re-identification without concrete text labels[C]. AAAI Conference on Artificial Intelligence, 2023, 37(1): 1405-1413.

[28] WANG H CH, SHEN J Y, LIU Y T, et al. Nformer:Robust person re-identification with neighbor transformer [C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 7287-7297.

[29] JIA M X, CHENG X H, LU SH J, et al. Learning disentangled representation implicitly via transformer for occluded person re-identification [J]. IEEE Transactions on Multimedia,2022,25:1294-1305.

作者简介

任鹏霏,硕士研究生,主要研究方向为深度学习、计算机视觉。

E-mail:1223086824@njupt.edu.cn

杨真真(通信作者),博士,副教授,主要研究方向为深度学习、计算机视觉。

E-mail:yangzz@njupt.edu.cn