中图法分类号:TP391.4;TP18　　**文献标识码**:A　　**文章编号**:1006-8961(2026)11-3893-16

论文引用格式:Ye X Y, Xue Z Q, Jiang D Q, Dai Y C and Wang J X. 2026. Occluded person re-identification with significance-guided erasure and gating regularization. Journal of Image and Graphics, 31(11):3893-3908(叶学义, 薛智权, 蒋德琦, 戴渝琛, 王佳欣. 2026. 显著性引导擦除与门控正则化的遮挡行人重识别. 中国图象图形学报, 31(11):3893-3908)[DOI:10.11834/jig.250576]

显著性引导擦除与门控正则化的遮挡行人重识别

叶学义*,薛智权,蒋德琦,戴渝琛,王佳欣

杭州电子科技大学通信工程学院,杭州　310018

摘　要:**目的**　遮挡行人重识别因可见区域缺失和遮挡噪声干扰,导致特征判别性不足、鲁棒性差。现有方法多局限于单视图表征,存在通道判别性建模不充分、显著区域过拟合等问题。**方法**　提出一种融合通道注意力增强、显著性引导擦除与门控正则化的多视角学习框架。该方法首先采用挤压激励残差网络(squeeze-and-excitation residual network,SE-ResNet)作为骨干网络,通过通道注意力机制自适应增强判别性特征。进而,设计Top-k显著性引导软擦除模块,利用类激活图定位关键区域并进行平滑抑制,迫使网络发掘多样化辅助特征,以缓解过拟合。此外,引入门控正则化机制,对特征通道选择概率施加熵约束,提升选择过程的稀疏性与稳定性。在训练阶段,通过加权融合多视图特征并借助特征对齐损失实现信息协同;推理阶段仅需单视图特征即可完成高效检索。**结果**　在4个代表性数据集上的实验表明,所提方法均取得领先性能。其中,在专门针对遮挡行人重识别的数据集Occluded-REID(occluded person re-identification)上,平均精度均值(mean average precision,mAP)与首位命中率(rank-1 accuracy,Rank-1)分别达到83.4%与92.8%;在大型多场景复杂数据集MSMT17(multi-scene multi-time)上,分别达到71.4%与89.6%,较先进方法MVIIP(multi-view information integration and propagation)有显著提升。在基于Duke-MTMC(Duke multi-target multi-camera)构建的合成遮挡数据集P-DukeMTMC和经典通用行人重识别基准数据集Market-1501上,所提方法同样表现优异,验证了其针对合成与自然遮挡的普适性。**结论**　所提方法在遮挡与多场景任务中均表现出良好的适应性、鲁棒性和泛化能力,适用于复杂监控场景下的行人检索。

关键词:遮挡行人重识别;判别性特征挖掘;通道注意力增强;显著性引导擦除;门控正则化

Occluded person re-identification with saliency-guided erasure and gating regularization

Ye Xueyi*, Xue Zhiquan, Jiang Deqi, Dai Yuchen, Wang Jiaxin

School of Communication Engineering, Hangzhou Dianzi University, Hangzhou　310018, China

Abstract: **Objective** Occluded person re-identification (Re-ID) aims to match individuals across non-overlapping camera views despite frequent partial obstruction, making it a challenging problem for intelligent surveillance and public security. The key difficulties arise from two sources: the loss of discriminative identity information in occluded regions and the intrusion of noise from occluders. Although many methods have been proposed, most remain limited to single-view paradigms and therefore struggle to model channel-wise feature importance adequately, avoid overfitting to the most salient regions, and enforce stable feature selection under severe visual ambiguity. These limitations reduce both robustness and generaliz-

收稿日期:2025-11-14;**修回日期**:2026-02-03;**预印本日期**:2026-02-10

*通信作者:叶学义　xueyiye@hdu.edu.cn

基金项目:国家自然科学基金项目(U19B2016)

Supported by:National Natural Science Foundation of China(U19B2016)

ability. **Method** To address these issues holistically, this paper proposes a novel multi-view gated and soft-regularized network (MGSR-Net), a synergistic framework designed to improve feature discriminability, diversity, and stability. MGSR-Net consists of an efficient single-view inference branch and a robust multi-view regularization branch, both of which share a squeeze-and-excitation ResNet (SE-ResNet) backbone for channel-aware feature enhancement. The first key component is the Top-k saliency-guided contextual erasure (TSCE) module. Unlike random erasure or hard masking, TSCE uses class activation maps (CAMs) to identify the top-k most influential spatial regions and then applies soft suppression to reduce their dominance. This encourages the network to learn from secondary but complementary regions, thereby alleviating overfitting and improving feature diversity. The second key component is a gated regularization mechanism that imposes an entropy constraint on channel-selection probabilities. This encourages sparse and stable feature-channel selection so that the model consistently relies on a robust set of core channels even under noisy or occluded conditions. During training, features from multiple views are processed by TSCE, weighted by stabilized gates, and fused. A feature-alignment loss then distills the resulting robust multi-view knowledge back into the single-view branch through L2 minimization, enabling efficient single-view inference at test time. The total loss combines identity-classification losses for both branches, the alignment loss, and the gate-regularization loss. **Result** Extensive experiments validate the effectiveness of MGSR-Net. On the Occluded-REID dataset, it achieves state-of-the-art performance with 83.4% mean average precision (mAP) and 92.8% Rank-1 accuracy, surpassing the previous leading multi-view method, multi-view information integration and propagation (MVIIP), by 1.7% mAP and 4.8% Rank-1. On P-DukeMTMC and Market-1501, the model reaches 89.3% Rank-1/77.6% mAP and 95.0% Rank-1/88.0% mAP, respectively, demonstrating strong transferability. Most notably, on the large-scale, multi-scene MSMT17 dataset, MGSR-Net achieves 89.6% Rank-1 and 71.4% mAP, outperforming MVIIP by 5.7% Rank-1 and 10.0% mAP, which highlights its exceptional generalization ability. Ablation studies confirm the contribution of each component: SE-ResNet provides a strong channel-enhanced backbone, TSCE is crucial for feature diversity, gated regularization stabilizes channel selection, and the alignment loss enables effective knowledge distillation from the multi-view branch to the single-view branch. Visualizations further show that MGSR-Net produces more holistic and adaptive attention maps than competing methods, shifting focus to visible informative regions more effectively under heavy occlusion. **Conclusion** This paper proposes MGSR-Net, an integrated framework that combines channel enhancement, saliency-guided soft feature perturbation, and entropy-constrained feature-selection stabilization within a multi-view learning paradigm. It effectively overcomes the limitations of single-view modeling, mitigates overfitting, and promotes robust feature selection under occlusion. Extensive results on both occlusion-specific and large-scale general benchmarks show that MGSR-Net establishes a new state of the art, offering improved discriminability, robustness, and generalization. Its design philosophy also provides a valuable reference for other vision tasks affected by partial information loss. Future work will explore more adaptive multi-view fusion strategies.

Key words: occluded person re-identification; discriminative feature mining; channel attention enhancement; saliency-guided erasure; gated regularization

0 引 言

行人重识别(person re-identification, Re-ID)技术(张永飞 等,2023)在智能监控、公共安全等领域中具有重要应用(Ye等,2022)。然而在真实场景下,行人往往会受到其他目标或障碍物的遮挡,导致可见信息缺失、噪声干扰以及视角错位等问题,从而显著降低Re-ID的性能。遮挡行人重识别(occluded person re-identification, Occluded-Re-ID)旨在当仅有部分区域可见时,依然能够准确完成跨摄像头行人检索,是该领域一项极具挑战性的任务。

为应对遮挡问题(Ning等,2024),现有研究经历了从局部特征建模到数据增强,再到多源信息融合的演进过程。早期方法主要围绕单视图框架展开,这类方法试图从单幅不完整的观测图像中恢复行人特征,其典型策略是依赖外部模型,如姿态估计模型(张红颖 等,2024)等,来定位并对齐人体部位。然而,一旦图像中的关键判别区域被完全遮挡,模型便失去了赖以判断的信息基础。此外,其在特征提取过程中也缺乏对噪声区域的有效抑制机制。

进一步地,研究者提出通过数据增强(Shorten

和 Khoshgoftaar，2019）提升模型的鲁棒性。如随机擦除等增强技术，通过随机屏蔽图像区域模拟遮挡场景，但这种无差别的扰动方式具有盲目性，可能导致关键特征被完全破坏，只能保留噪声区域，与去噪存真的初衷相悖。

为了克服随机增强的盲目性，研究者转而寻求更具引导性的策略——注意力机制。通过让模型自主聚焦于图像中的显著区域，注意力机制能够确保关键特征在训练中得以保留和强化。尽管注意力机制在一定程度上缓解了增强策略的盲目性，但其作用范围仍被限制在单次观测内容之内。 当单幅图像中的关键判别特征因严重遮挡而完全缺失时，无论是随机擦除还是注意力机制，都难以从根本上解决信息不足的问题。并且聚焦机制会使模型容易过度依赖于少数几个高响应区域，从而忽略了其他同样具有判别潜力的辅助特征。

于是，有部分研究者尝试引入多视角信息（Dong 等，2024a）或多模态数据（Wu 等，2025）以弥补单视图的不足，或在特征学习机制层面进行改进（He 等，2021）。这些工作取得了一定进展，但其核心问题在于模型内部难以在融合多视角信息的同时，实现特征挖掘与噪声抑制的自适应平衡以及确保学习过程的稳定性。具体而言，现有方法往往缺乏对通道判别性的精细建模，未能有效缓解显著区域过拟合问题，同时也缺少对特征分布的必要约束。

为解决上述问题，本文提出了一种多视角协同学习框架。主要贡献可归纳为以下 4 个方面：1）提出了一个融合多视角通道注意力、显著性引导擦除与门控正则化的协同学习框架（multi-view gated and soft-regularized network，MGSR-Net），其核心是构建一条从通道增强、注意力重分配到选择稳定的协同机制，以挤压激励残差网络（squeeze-and-excitation residual network，SE-ResNet）实现通道增强，经 Top-k 显著性引导软擦除模块（top-k saliency-guided contextual erasure，TSCE）实现通道驱动的空间抑制，最终凭借门控机制稳定通道选择的过程；2）设计了 Top-k 显著性引导软擦除模块，该模块利用类激活图定位最显著区域并作平滑抑制，从而促使模型挖掘多样化辅助特征，缓解显著区域过拟合问题；3）引入门控正则化机制，借助熵约束（Shu 等，2018）稳定特征选择过程，改善遮挡情况下特征选择不一致的问题，增强模型鲁棒性；4）在 Occluded-REID 与 MSMT17（multi-scene multi-time）数据集上的实验结果显示，本文方法性能显著优于现有主流方案，体现出优越的判别性、鲁棒性及泛化能力。

1 相关工作

遮挡行人重识别的核心挑战在于如何处理部分可见区域内的判别性特征缺失以及由遮挡引入的噪声干扰问题。如引言所述，现有研究虽然多有尝试，但仍存在诸多局限。本节从局部特征建模、遮挡鲁棒性增强与多源信息融合 3 个维度对现有工作进行梳理与分析，在指出现有方法不足的同时，明确本文工作的出发点。

1.1 基于局部特征建模的方法

为缓解遮挡对全局特征造成的干扰，大量研究致力于挖掘行人的局部判别性特征。如引言所述，早期方法多基于单视图框架，且需要通过外部模型实现局部特征的提取与匹配。例如，Che 等人（2023）通过姿态估计工具 OpenPose 提取人体关键点，利用姿态热图增强全局特征，并结合水平切割对齐身体部位的局部特征，以提升重识别准确率。Dung 和 Nakamura（2024）采用预训练分割模型提取人体前景，将其与原始图像水平拼接为双语义解析图像（dual semantic parsing image，DSPI）作为输入，从而增强模型对人体结构信息的感知，提升跨域识别性能。此类方法在一定程度上能够对齐可见部位，但其效果严重受制于外部模型的精度与泛化能力，且整体流程复杂，难以实现端到端优化。

为了降低对外部模型的依赖，内部引导的局部特征学习方法逐渐成为主流。这类方法的核心优势在于将局部特征的发现过程融入主网络中进行端到端学习。如 He 等人（2021）在 TransReID 中引入拼图块模块（jigsaw patch module，JPM），通过在 Transformer 结构内对图像块进行随机平移与重组，主动破坏图像的全局空间结构，迫使模型学习更具判别性的局部关系，从而显著提升遮挡鲁棒性。类似地，Tan 等人（2022）提出随机丢弃特征条带的方法，通过在特征图中随机破坏连续区域，以打破模型对固定判别模式的依赖，这种内部扰动迫使网络必须关注于更广泛的特征，从而学习到更为丰富、更加鲁棒的特征表示。Kim 等人（2022）提出的遮挡校正网络（occlusion correction network，OCNet）则采用预定义

分区策略，将特征图划分为顶部、底部与中心等区域，并通过一个关系自适应模块建模区域间关系、抑制噪声干扰。这些方法虽然摆脱了外部模型，但本质上仍未突破单视图表征的局限，其局部划分或随机丢弃策略缺乏自适应性，无法准确聚焦于判别性最强的关键区域，并可能因随机操作破坏有效信息。

1.2 基于遮挡鲁棒性增强的方法

尽管局部特征建模方法从网络结构层面增强了对遮挡的感知能力，但其性能上限仍受制于原始训练数据的信息完整性与多样性。因此，另一条并行发展的技术路线则从训练策略入手，不改变核心网络结构，而是通过对输入数据施加扰动直接提升模型对遮挡的固有鲁棒性。其中，随机擦除（Zhong等，2020；Dong等，2024b）及其衍生方法通过在训练过程中随机遮挡图像区域模拟遮挡场景，从而提高模型泛化能力。然而这种无差别的随机扰动缺乏针对性，有可能误除关键判别特征，却保留了无关的背景噪声，在一定程度上偏离了模型本来的学习目标。

为增强数据增强的引导性，研究者逐渐引入注意力机制与显著性挖掘。例如，Song等人（2020）提出的上下文交互卷积神经网络（context-interactive convolutional neural network，CI-CNN）利用强化学习动态调整注意力的时空上下文范围。Dong等人（2024a）提出的多视图信息集成与传播框架（multi-view information integration and propagation，MVIIP）利用类激活图定位判别性区域，并融合多视图信息。这类方法赋予模型自主聚焦能力（Vaswani等，2017），但也容易过度依赖少数显著区域，忽略其他具有判别潜力的辅助特征，导致特征多样性不足，限制模型泛化能力。Wu等人（2025）提出的多模态框架虽然通过融合不同模态信息以增强判别性，但在单一RGB模态出现极端遮挡情况时，其内部特征的挖掘深度仍不够充分。

1.3 基于多源信息融合的方法

前述基于单视图的增强策略，无论是随机扰动还是注意力引导，其根本局限在于它们都试图从一幅信息量有限的输入图像中获取更多信息。当单幅图像中的关键判别特征因严重遮挡而缺失时，这类方法便难以从根本上解决问题。

利用多源信息成为应对遮挡问题的关键思路之一。Song等人（2020）利用视频序列中的连续帧，通过建模时序上下文恢复被短暂遮挡的行人特征。这类方法虽然能利用运动信息，但其计算复杂度显著高于图像方法，且严重依赖帧与帧之间的精准对齐，难以满足实时性要求。Wu等人（2025）的工作尝试结合红外、深度等模态与RGB图像，以期在一种模态信息缺失时由其他模态补足。然而，这类方案的实用性会受限于硬件配置，难以大面积普及。

相较之下，利用同一身份在不同摄像头下的多幅静态图像，即多视角学习，成为更具普适性和可行性的技术路径。该思路不依赖特殊硬件或连续帧，仅利用重识别任务本身固有的数据特性。其中，Dong等人（2024a）提出的MVIIP框架是代表性工作，该框架的核心思想是在训练阶段，利用同一行人的多幅不同视角图像进行信息互补。定位模块先通过类激活图定位每幅图像中对识别过程有贡献的区域，然后量化模块根据分类器正确识别特征图的概率为不同图像的特征分配相应权重，最后通过集成模块构建一个综合表征，并通过知识蒸馏将其中的信息传递给基线推理模型。然而，通过深入分析，发现这一框架在处理极端复杂的遮挡场景时，仍存在一些局限性。

首先，它的融合与蒸馏基础对可见可靠区域的依赖过强，默认每幅图像中都能通过定位模块找到可靠的判别区域。但在极端遮挡下，整幅图像内可能不存在任何强判别性区域，导致融合过程变为对弱信号或噪声的加权。

其次，它的特征挖掘过程过于被动，定位模块的设计目标是保留并突出高响应区域，这种被动的选择策略并未主动引导网络去探索那些未被充分激活、但仍具有判别潜力的区域。当所有高响应区域均被遮挡时，模型将陷入无可靠区域可选的困境。

此外，该框架对于“如何选择特征”这一过程本身未施加任何约束。在遮挡造成的噪声干扰下，网络对判别区域的选择会随输入图像可见区域的不同而剧烈波动。这意味着对于同一身份，其多视图融合后的表征可能由完全不同的一组特征主导，导致融合结果缺乏一致性，且蒸馏给单视图模型的目标模糊。这种不稳定选择将会破坏学习特征的泛化能力。

1.4 本文工作的出发点

综上所述，现有研究仍面临三大挑战：1）单视图表征瓶颈：多数方法仍然只进行单幅图像内部的信息挖掘，没有充分利用同一身份下多视角图像的互

补信息(蒋建国 等,2019);2)增强策略的盲目性与过拟合:随机增强方式缺乏针对性,而注意力机制又容易导致模型对少数显著区域过拟合,缺乏引导模型探索多样化特征的有效机制;3)特征稳定性与判别性不足:在复杂遮挡场景下,特征选择过程缺乏稳定性约束,容易引入噪声,导致模型鲁棒性下降。

如1.3节所述,MVIIP等前沿工作通过引入多视角信息试图突破单视图瓶颈,但是"被动地选择并增强显著区域"的核心思想并不能全面解决上述挑战。因此,本文认为,解决遮挡场景下行人重识别的关键在于实现更加多样化与稳定的特征挖掘过程,并在此基础上提出了名为 MGSR-Net 的协同学习框架,如图1所示。

该框架借助SE-ResNet增强通道判别性,通过将多个视图加权融合来构建鲁棒的综合表征,打破了单视图局限;创新性地设计了TSCE模块,利用Top-k显著性引导软擦除主动抑制最显著区域,迫使网络注意力向次显著区域转移,从而挖掘多样化、互补的辅助特征,从根本上解决无可靠区域可用的问题;引入了门控正则化机制,通过对特征通道的选择概率施加熵约束,使模型学会一套稳定的判断标准,从而在面对遮挡干扰时,能始终依赖同样的一组核心特征进行识别,大大增强了模型的抗干扰能力。

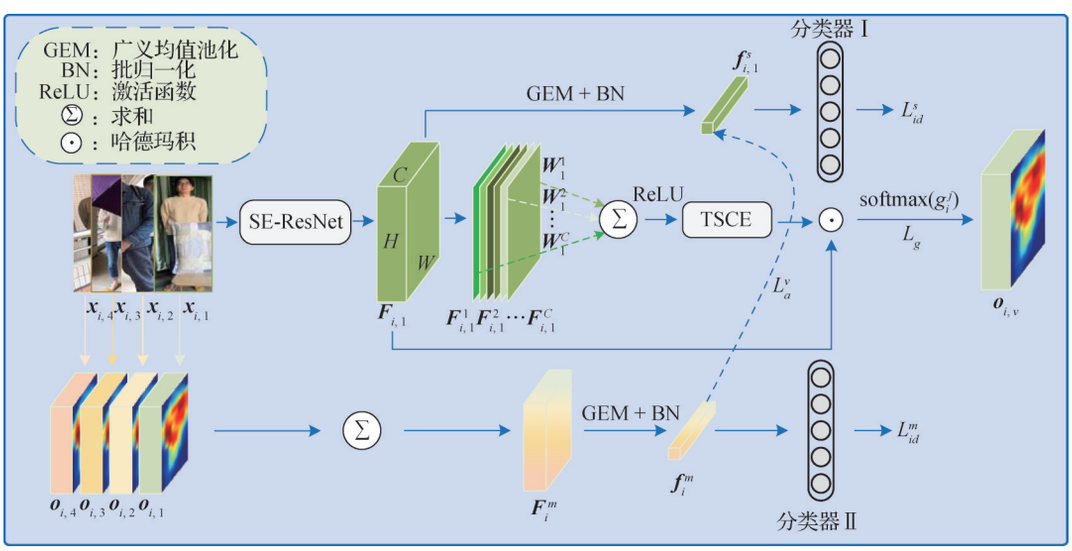


图1　MGSR-Net框架示意图

Fig. 1　MGSR-Net framework diagram

2　本文方法

MGSR-Net 框架包含一个单视角表征分支和一个多视角正则化分支。该框架以SE-ResNet为共享骨干网络,并行地处理同一行人的多个视角图像 $\boldsymbol{x}_{i,1},\boldsymbol{x}_{i,2},\boldsymbol{x}_{i,3},\boldsymbol{x}_{i,4}$,其中下标 i 为样本索引,$v\in 1,2,3,4$ 表示视角索引。单视角分支为每个视角生成一个强判别性的基线特征。多视角分支则依次通过 TSCE 模块和门控正则化机制,生成一个集成且经过正则化的多视角特征,并通过基于 L_2 范数的损失函数将多视角信息传递回单视角分支,该设计使得测试阶段即使只用单幅图像也能获得该行人其他视角的综合表征。以下分节将对各个核心组件进行详细阐述。

2.1　骨干网络SE-ResNet

MGSR-Net 框架采用SE-ResNet(Hu 等,2018)作为模型的骨干网络,以此构建一个强大且鲁棒的特征提取基础。该网络在经典残差网络(residual network,ResNet)(He 等,2016)结构的残差分支中嵌入了一个轻量级的子网络挤压激励(squeeze-and-excitation,SE)模块,作为对标准 ResNet 模块的增强,如图2右侧所示,该模块首先通过全局平均池化(global average pooling,GAP)获取通道级别的全局信息,再利用全连接层(fully connected layer,FC)学习各通道的重要性权重,从而增强判别性通道、抑制噪声通道。这使得模型在面对遮挡时,仍能聚焦于未被遮挡的判别性区域,在提升特征判别力的同时,以极小的计算成本提升了模型的鲁棒性和特征表征

能力。

给定输入图像$\boldsymbol{x}_{i,v}$，共享权重的SE-ResNet骨干网络输出一个三维特征张量$\boldsymbol{F}_{i,v} \in \mathbf{R}^{C\times H\times W}$，式中，$C$表示通道数，$H$、$W$分别表示特征图的高度与宽度。SE模块的核心运算如下：

1）Squeeze操作。通过全局平均池化将每个通道的空间信息压缩为一个全局描述符，记第C个通道的全局描述符为z_c，$z_c \in \mathbf{R}$，具体为

$$z_c = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} u_c(h,w) \tag{1}$$

式中，$u_c(h,w)$表示第C个通道在空间位置(h,w)处的激活值，将所有C个通道的全局描述符拼接，得到压缩后的特征向量$\boldsymbol{z} \in \mathbf{R}^C$，式中，$\boldsymbol{z}=[z_1, z_2, \cdots, z_C]^{\mathrm{T}} \in \mathbf{R}^C$。

2）Excitation操作。通过一个包含降维与还原的两层全连接层结构，生成通道权重向量$\boldsymbol{s} \in \mathbf{R}^C$，具体为

$$\boldsymbol{s} = \sigma(\boldsymbol{W}_2\delta(\boldsymbol{W}_1\boldsymbol{z})) \tag{2}$$

式中，$\boldsymbol{W}_1 \in \mathbf{R}^{\frac{C}{r}\times C}$、$\boldsymbol{W}_2 \in \mathbf{R}^{C\times\frac{C}{r}}$为可学习权重矩阵，$r$为缩减比率，$\delta$表示线性整流单元（rectified linear unit，ReLU）的激活函数，σ表示sigmoid激活函数。随后用该权重向量$\boldsymbol{s}$对原始特征图$\boldsymbol{U}$进行重新校准，具体为

$$\tilde{\boldsymbol{U}} = \boldsymbol{s}\odot\boldsymbol{U} \tag{3}$$

式中，$\odot$表示逐通道的Hadamard积。该操作能够有效增强判别性特征的响应，并抑制不相关或噪声通道的影响。

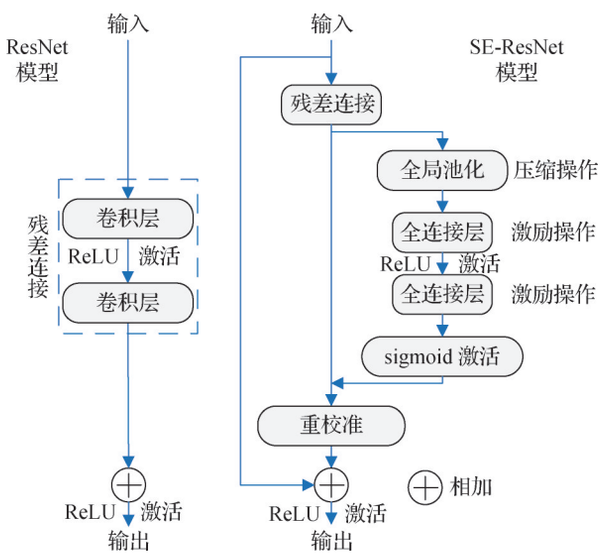


图2　ResNet和SE-ResNet模块的结构示意图

Fig. 2　Schematic diagrams of the ResNet and SE-ResNet modules

2.2　Top-k显著性引导软擦除（TSCE）

SE-ResNet骨干网络通过通道注意力机制为模型提供了强大的特征判别基础，能够自适应地校准通道维度上的特征响应。然而这种通道层面的增强在遮挡场景下会使模型在空间维度上过度聚焦于通道响应最高的局部区域，忽视了其他同样具有判别潜力的身体部位。这种注意力失衡限制了模型的特征多样性，当关键区域被遮挡时，识别性能会显著下降。

为缓解模型对最显著区域的过度依赖问题，并引导网络挖掘多样化的辅助特征，本文设计了Top-k显著性引导软擦除模块（TSCE）。该模块摒弃了直接将区域置零的传统硬擦除方式，而是通过类激活图定位关键区域后再对其进行平滑抑制。这种基于显著性引导的渐进式擦除，实质上是引导模型在每次训练中查漏补缺，通过持续削弱显著特征，逐步强化对多样化辅助特征的学习，最终形成平衡且全面的特征表示。

该模块以骨干网络输出的特征$\boldsymbol{F}_{i,v}$为输入，具体处理流程如下：

1）生成类激活图（class activation map，CAM）（Zhou等，2016）。通过分类器的权重向量$\boldsymbol{W}_{\mathrm{p}} \in \mathbf{R}^C$（p对应目标行人身份pid）与特征图$\boldsymbol{F}_{i,v}$生成类激活图$\boldsymbol{M}_{i,v} \in \mathbf{R}^{H\times W}$，将SE-ResNet在通道维度上学到的判别性知识转化为空间显著性热图，用于标识模型决策所依赖的空间区域。具体为

$$\boldsymbol{M}_{i,v} = ReLU\left(\sum_{c=1}^{C}\boldsymbol{W}_{\mathrm{p}}^{c}\cdot\boldsymbol{F}_{i,v}^{c}\right) \tag{4}$$

随后，将$\boldsymbol{M}_{i,v}$归一化至$[0,1]$区间，为后续的擦除操作提供与身份判别直接相关的引导。

2）确定Top-k阈值与生成软掩码。首先，对于每一幅归一化后的类激活图$\boldsymbol{M}_{i,v} \in [0,1]^{H\times W}$，将其所有像素的显著性值展平为一维向量，记做$\boldsymbol{m} \in \mathrm{R}^{H\times W}$。对$\boldsymbol{m}$进行降序排序，得到有序序列$\boldsymbol{m}\downarrow = [v_1, v_2\cdots, v_N]$，其中$v_1 \geqslant v_2 \geqslant \cdots \geqslant v_N$。

给定预设的擦除比例$\rho \in (0,1)$，计算需要抑制的像素数量$k = H\times W\times\rho$（向下取整），取有序序列$\boldsymbol{m}\downarrow$中第k个元素v_k作为区分高、低响应区域的阈值$T_{i,v}$，无论这幅特征图的整体响应强弱，都只对其中最突出的k个像素区域进行抑制。

根据此阈值，为特征图中的每个空间位置(h,w)

生成一个软掩码$\boldsymbol{\Lambda}_{i,v} \in \mathbf{R}^{H \times W}$。该掩码函数是一个显式的分段函数，由衰减因子$\alpha \in (0,1)$控制，具体为

$$\boldsymbol{\Lambda}_{i,v}(h,w) = \begin{cases} \alpha & \boldsymbol{M}_{i,v}(h,w) \geqslant T_{i,v} \\ 1 & \text{其他} \end{cases} \tag{5}$$

对于类激活响应值大于等于阈值$T_{i,v}$的像素，将其在后续特征中的贡献衰减为原来的α倍，例如设置$\alpha = 0.3$来保留30%的响应；对于其余区域，则全部予以保留。通过这种软抑制，模型被引导去探索那些非最显著但同样具有判别性的特征。同时，保留部分原始信息的策略，确保了梯度回传的稳定性与训练过程的平滑，避免了因信息突然完全缺失而导致的训练振荡。

3）擦除。将软掩码作用于原始特征图上，得到擦除后的特征$\boldsymbol{F}^{e}_{i,v} \in \mathbf{R}^{H \times W \times C}$，具体为

$$\boldsymbol{F}^{e}_{i,v} = \boldsymbol{F}_{i,v} \odot \boldsymbol{\Lambda}_{i,v} \tag{6}$$

该设计实现了对显著区域的自适应抑制。经过TSCE模块处理后，模型被强制降低对原始高响应区域的依赖，从而促使其从图像的其余部分发掘出其他具有互补性的判别特征，显著增强了特征表示的多样性与鲁棒性。

2.3 门控正则化与Gate Mask Loss

TSCE模块通过软擦除有效增强了特征的多样性，但如果辅助特征被大量激活，则会使特征空间复杂化并可能引入冗余。在复杂遮挡场景下，这种特征多样性若缺乏有效约束，模型可能难以在不同视角间收敛到一组稳定且一致的判别性特征，进而影响多视角融合的最终效果。

为此，本框架引入了门控正则化机制。区别于现有许多工作将熵约束应用于分类输出以提升预测置信度，这里将其作用于特征通道的选择过程。其核心是通过对特征向量所生成的门控概率施加熵约束，促使选择分布趋于稀疏稳定，即使得多数特征通道的选择概率逼近0或1，从而在特征构建层面提升选择的确定性。

1）特征压缩与门控概率生成。首先，对经TSCE处理后的特征$\boldsymbol{F}^{e}_{i,v}$进行广义均值池化（generalized mean pooling，GEM）和批归一化（batch normalization，BN），将其压缩为一个特征向量$\boldsymbol{f}^{e}_{i,v} \in \mathbf{R}^{C}$。随后，将该特征向量输入分类器的全连接层，为每个特征通道生成对应的门控值。记第j个特征通道的门控值为$g^{j}_{i,v}$。

对所有C个通道的门控值应用softmax函数，得到归一化的选择概率$p^{j}_{i,v}$，代表第j个特征通道被选择的概率，具体为

$$p^{j}_{i,v} = \frac{\exp\left(g^{j}_{i,v}\right)}{\sum_{k=1}^{C} \exp\left(g^{k}_{i,v}\right)} \tag{7}$$

这一设计使模型能够自适应地学习特征通道的重要性权重，而非依赖固定的规则或分类置信度，为后续的稳定化约束提供了可优化的基础。

2）特征加权与多视角融合。每个视角使用其门控概率对TSCE处理后的特征进行加权，生成该视角的加权表征$\boldsymbol{o}_{i,v} \in \mathbf{R}^{C}$，具体为

$$\boldsymbol{o}_{i,v} = \sum_{j=1}^{C} p^{j}_{i,v} \cdot \boldsymbol{F}^{e,j}_{i,v} \tag{8}$$

式中，$\boldsymbol{F}^{e,j}_{i,v}$表示$\boldsymbol{F}^{e}_{i,v}$的第$j$个通道的特征，这种基于学习权重的特征选择，其目标在于汇聚跨视角下最稳定、最具判别力的特征信息，从而缓解单视图因遮挡造成的特征缺失或噪声干扰。确保了后续的多视角融合中，各视角贡献的都是其最具判别力的特征分量。

3）Gate Mask Loss。为使上述特征选择过程明确且一致，需要约束门控概率$p^{j}_{i,v}$使其趋于稀疏和稳定，因此对其施加基于二值化熵的正则化损失L_{g}，具体为

$$L_{g} = -\frac{1}{C}\sum_{j=1}^{C}\left[\, p^{j}_{i,v} \cdot \log\left(p^{j}_{i,v}\right) + \left(1 - p^{j}_{i,v}\right) \cdot \log\left(1 - p^{j}_{i,v}\right)\right] \tag{9}$$

该损失函数最小化时迫使$p^{j}_{i,v}$趋向于0或1，与常见的作用于分类输出以提升预测置信度的标准熵损失不同，这里的二值化熵损失作用于特征通道的选择概率，其优化目标是消除每个通道在选用与忽略之间的模糊状态，从而稳定特征选择过程本身。

2.4 信息传递与总体优化目标

为将多视角分支的增强效果迁移至高效的单视角推理分支（Hinton等，2015），本框架采用了特征对齐损失进行知识传递，确保单视角分支能够模仿学习经TSCE与门控正则化优化后的鲁棒特征，从而在不引入额外推理成本的前提下，继承模型的判别力与鲁棒性。

对于单视角分支，输入$\boldsymbol{x}_{i,v}$经过骨干网络后，直接通过GEM + BN得到该视角的基线特征向量$\boldsymbol{f}^{s}_{i,v} \in \mathbf{R}^{C}$。

对于多视角分支，其最终输出为融合后的特征$\boldsymbol{F}_i^{\mathrm{m}}$再经过一次GEM + BN后得到的向量$\boldsymbol{f}_i^{\mathrm{m}} \in \mathbf{R}^C$。

通过强制多视角集成表征向单视角基线表征逼近，将前者的有效信息蒸馏至后者。对于样本i的第v个视角，其对齐损失为

$$L_{\mathrm{a}}^v = \left\| \boldsymbol{f}_i^{\mathrm{m}} - \boldsymbol{f}_{i,v}^{\mathrm{s}} \right\|_2 \tag{10}$$

网络的总体训练损失由以下4部分构成，实现了多任务协同优化，具体为

$$L = L_{\mathrm{id}}^{\mathrm{s}} + L_{\mathrm{id}}^{\mathrm{m}} + \lambda_1 L_{\mathrm{a}} + \lambda_2 L_{\mathrm{g}} \tag{11}$$

式中，$L_{\mathrm{id}}^{\mathrm{s}}$为单视角分支的身份分类损失（交叉熵损失），$L_{\mathrm{id}}^{\mathrm{m}}$为多视角分支的身份分类损失（计算于$\boldsymbol{f}_i^{\mathrm{m}}$），$L_{\mathrm{g}}$为门控正则化损失；$\lambda_1$和$\lambda_2$是平衡各项损失的权重超参数；$L_{\mathrm{a}}$为所有视角信息传递损失的平均，具体为

$$L_{\mathrm{a}} = \frac{1}{V}\sum_{v=1}^{V} L_{\mathrm{a}}^v \tag{12}$$

式中，V为视角数。

在推理阶段，不采用多视角分支，仅依靠单视角分支处理查询图像，使用$\boldsymbol{f}_{i,v}^{\mathrm{s}}$进行检索。该方案在维持高效识别性能的基础上，检索效率仍与主流单视角方法保持同一水平。

3 实　验

3.1 数据集

为全面评估MGSR-Net框架的性能与泛化能力，本研究选取了4个具有代表性的数据集进行系统实验。

1）Occluded-REID（Zhuo等，2018）是针对遮挡行人重识别任务的专用数据集，包含2 000幅图像，涵盖200个不同身份。所有图像均采集自真实监控场景，行人目标受到不同程度遮挡。每个身份平均有5幅训练图像和5幅测试图像。该数据集的特点是呈现严重的遮挡情况，为评估模型在信息缺失条件下的鲁棒性提供了理想测试平台。

2）P-DukeMTMC（Zhuo等，2018）是基于DukeMTMC（Duke multi-target multi-camera）构建的半身遮挡数据集，通过合成方式在行人上半身添加随机遮挡块，模拟头部、肩部等关键区域被遮挡的情况。该数据集包含12 927幅训练图像与11 216幅测试图像，共1 299个身份，适用于评估模型对结构化合成遮挡的适应性。

3）Market-1501（Zheng等，2015）是行人重识别领域的经典基准数据集，包含32 668幅图像，涉及1 501个身份，采集自6个摄像头。尽管其非专门针对遮挡设计，但实际图像中包含部分自然遮挡情况，旨在评估模型在更广泛、更通用的行人识别任务中的性能表现。

4）MSMT17（multi-scene multi-time）（Wei等，2018）是目前规模最大、场景最复杂的行人重识别数据集之一，包含126 441幅图像，涉及4 101个身份。该数据集采集自包含15个摄像头视角的校园监控系统，覆盖不同的时间段和天气条件，其真实场景中同样包含了大量的遮挡样本，具备丰富的身份多样性和图像复杂度，能够有效评估模型在大规模复杂场景下的泛化性能。

3.2 实验设置

所有实验均在PyTorch深度学习框架下实现，使用单张NVIDIA RTX 3090 GPU进行训练和测试。骨干网络采用SE-ResNet50，其权重在ImageNet数据集上进行预训练初始化。输入图像统一调整为256 × 128像素。训练阶段采用随机翻转、填充和裁剪等数据增强方法。优化器选用Adam（Kingma和Ba，2015），初始学习率设置为3×10^{-4}，并采用多步长衰减策略（He等，2016）进行调整，在第40和70轮次进行学习率调整。批量大小设置为64，每批次包含8个身份样本，每个身份包含8幅图像，其中选择同一身份的4幅不同视角图像用于多视角信息融合。模型总共训练120个epoch，每5个epoch进行一次性能评估。

TSCE模块中的擦除比例ρ设置为0.2，衰减因子α设为0.3；多任务损失函数中的权重系数设置为：$\lambda_1 = 0.007$，$\lambda_2 = 0.01$。所有实验随机种子均设置为1，以确保结果可复现。

测试阶段采用跨摄像头评估模式，加载训练阶段的最佳模型参数，仅使用单视角分支进行特征提取。评估指标包括平均精度均值（mean average precision，mAP）和累积匹配特性曲线（cumulative matching characteristic curve，CMC）的各阶次识别准确率。

3.3 对比实验

3.3.1　在Occluded-REID数据集的结果

由于Occluded-REID数据集规模较小，单一实

验结果易受随机波动影响。为此，严格遵循其官方划分规则（Zhuo等，2018；Zhao等，2021）进行了10次独立实验，并将平均性能视做本文方法的代表性指标，以确保后续与其他方法对比的可靠性与统计显著性。

如表1所示，Rank-1准确率平均值为92.8%，标准差为1.3%，各次实验结果均在89.8%～95.0%范围内波动。10次独立实验在Occluded-REID数据集上均呈现出良好的稳定性。

随后，在Occluded-REID数据集上，与当前最先进的8种方法进行对比，结果如表2所示。相较于当前较优的多视角方法MVIIP，本文方法的Rank-1准确率从88.0%提升至92.8%，表明MGSR-Net采用的通道注意力增强与门控正则化机制在多视角信息融合与提炼上更具优势。与GSE-Net（graph semantic enhancement network）和GMFSE（3D-guided multi-feature semantically enhanced network）等特征增强模型相比，本文方法在各项指标上均实现全面超越，反映出TSCE模块在引导模型挖掘多样性特征方面的有效性。相比于OGFR（occlusion-guided feature purification learning via reinforced knowledge distillation）和ETND-Net（erasing，transforming，and noising defense network）等专门针对遮挡的模型，本文方法

表1　Occluded-REID数据集上10次实验的结果

Table 1　Results of 10 experiments on the Occluded-REID dataset

/%

实验	Rank-1	Rank-5	Rank-10	mAP
new_0	92.4	97.4	97.8	83.2
new_1	89.8	96.6	98.2	83.2
new_2	93.0	98.4	99.0	83.0
new_3	92.0	98.4	99.4	81.4
new_4	**95.0**	**99.2**	**100.0**	**86.0**
new_5	93.2	98.0	99.8	82.8
new_6	93.8	98.8	99.4	85.0
new_7	92.4	98.8	99.8	84.4
new_8	93.6	98.4	99.2	83.1
new_9	92.8	97.0	98.0	81.7
平均	92.8	98.1	99.1	83.4

注：加粗字体表示各列最优结果。

表2　Occluded-REID数据集上与当前最先进方法的对比

Table 2　Comparison with state-of-the-art methods on the Occluded-REID dataset

/%

方法	Rank-1	Rank-5	Rank-10	mAP
COFD-Net（Li等，2025a）	86.7	–	–	79.5
MVIIP	88.0	95.3	97.7	81.7
OGFR（Zheng等，2025）	86.9	–	–	80.9
ETND-Net（Dong等，2024b）	<u>90.5</u>	<u>96.5</u>	<u>98.5</u>	<u>81.9</u>
GMFSE（Ning等，2025）	82.3	–	–	79.9
GSE-Net（Wei等，2024）	82.4	–	–	78.5
LAF（Li等，2025b）	87.1	–	–	79.5
MAHATMA（Zhang等，2025a）	85.8	–	–	79.5
MGSR-Net	92.4	97.6	97.6	82.4
MGSR-Net*	**92.8**	**98.1**	**99.1**	**83.4**

注：加粗字体表示各列最优结果。下划线字体表示对比方法里的最优值。“*”表示加入门控正则化损失后的完整MGSR-Net模型，“–”表示原论文未展示相关数据。

通过端到端的协同学习框架，在保持高效推理的同时，获得了更高的识别精度。

值得注意的是，基础版本的MGSR-Net在Occluded-REID数据集上已经达到92.4%的Rank-1准确率和82.4%的mAP，优于所有对比方法。在加入门控正则化机制后，完整模型MGSR-Net*的性能得到进一步提升。

3.3.2　在P-DukeMTMC数据集的结果

P-DukeMTMC是通过合成方式模拟半身遮挡的数据集，用于评估模型对合成遮挡的适应性。如表3所示，MGSR-Net*在该数据集上的Rank-1和mAP分别为89.3%和77.6%，较基础版本MGSR-Net分别提升1.3%与3.3%。

与当前主流方法相比，MGSR-Net*在mAP指标上表现最优，显著超越了ISP（identity-guided human semantic parsing approach）、QPM（quality- aware part models）等专门针对部分遮挡设计的方法，并在Rank-1指标上保持领先，仅次于QPM。同时，本文方法明显优于早期局部对齐方法，如PCB（part-based convolutional baseline）、PVPM（pose-guided visible part matching）、PGFA（pose-guided feature alignment）以及融合图结构与注意力机制的RTGAT（rea-

soning and tuning graph attention network)。这一结果表明,MGSR-Net不仅适应真实遮挡,在合成遮挡场景下同样具备强大的竞争力。

表3 P-DukeMTMC数据集上与当前最先进方法的对比
Table 3 Comparison with state-of-the-art methods on the P-DukeMTMC dataset

/%

方法	Rank-1	Rank-5	Rank-10	mAP
PCB(Sun等, 2018)	79.4	87.1	90.0	63.9
PVPM(Gao等, 2020)	85.1	91.3	93.3	69.9
PGFA (Miao等, 2019)	85.7	92.0	94.2	72.4
RTGAT(Huang等,2023)	85.6	91.5	93.4	74.3
ISP(Zhu等, 2020)	89.0	94.1	95.3	74.7
QPM(Wang等,2023)	**90.7**	**94.4**	**95.9**	75.3
MGSR-Net	88.0	93.1	94.2	74.3
MGSR-Net*	89.3	93.9	94.8	**77.6**

注:加粗字体表示各列最优结果。下划线字体表示对比方法里的最优值。

P-DukeMTMC的随机部件遮挡策略,使得模型在每次训练中面临不同的部位缺失。而MGSR-Net中的TSCE模块,其设计初衷正是主动抑制当前最显著的判别区域,迫使网络去探索和依赖其他潜在的辅助特征。这种机制恰好应对了合成数据中随机缺失所带来的不确定性。同时,门控正则化机制确保了这种特征选择过程在不同随机遮挡样本间保持一致性与稳定性。

3.3.3 在Market-1501和MSMT17数据集的结果

Market-1501作为经典通用数据集,用于验证本方法从遮挡任务向常规行人重识别任务的迁移能力。MSMT17是目前规模最大、场景最复杂的行人重识别数据集之一,包含大量自然遮挡样本,适合评估模型在真实复杂场景下的泛化性能。

表4呈现了所提出方法在Market-1501和MSMT17两个权威数据集上的详细结果。其中,完整模型在Market-1501数据集上取得了95.0%的Rank-1精度和88.0%的mAP。其mAP指标为所有对比方法中的最高值,显示出优异的综合检索性能;Rank-1精度也与当前表现最佳的BINet(bidirectional interaction network)模型所得到的95.7%高度接近。基础版本的模型在此数据集上同样表现稳健,获得了94.6%的Rank-1和85.3%的mAP。与同期提出的CENet(cross-erasure enhanced network)、更早的OSNet(omni-scale network)以及基于ViT(vision Transformer)的ViTbase等代表性方法相比,完整模型MGSR-Net*展现出了明显优势。这一结果证明,从专门针对遮挡场景设计的任务框架迁移至常规行人重识别任务时,模型未出现性能损失,反而保持了领先的特征判别能力,体现了其良好的任务适应性与可迁移性。

在规模更大、场景更复杂的MSMT17数据集上,该方法的优势得到了更为显著的体现。完整模型取得了89.6%的Rank-1和71.4%的mAP,相较于其他先进方法实现了大幅度性能提升。

例如,对比同期优秀的CENet模型,完整模型在Rank-1和mAP上分别领先约9.0%和12.6%;与基于Transformer架构的ViTbase模型相比,优势分别达到约7.8%和10.4%。这一优异表现说明了该方法在面对真实世界中普遍存在的自然遮挡、光照变化及复杂背景干扰时依然具备卓越的鲁棒性。基础版本模型在此数据集上也取得了86.9%的Rank-1和66.2%的mAP,其性能已显著超越表4中列出的大多数现有方法,表明该方法的核心架构本身即为处

表4 在Market-1501及MSMT17数据集上与当前最先进方法的对比
Table 4 Comparison with state-of-the-art methods on the Market-1501 and MSMT17 datasets

/%

方法	Market-1501		MSMT17	
	Rank-1	mAP	Rank-1	mAP
CENet(Zhang等,2025b)	95.0	87.3	80.6	58.8
MVIIP(Dong等,2024a)	95.3	87.9	83.9	61.4
DRL-Net(Jia等,2023)	94.7	86.9	78.4	55.3
IANet(Hou等,2019)	94.4	83.1	75.5	46.8
OSNet(Zhou等,2019)	94.8	84.9	78.7	52.9
BINet(Chen等,2021)	**95.7**	87.2	76.1	52.8
DGNet(Zheng等,2019)	94.8	86.0	77.2	52.3
ViTbase(Dosovitskiy等,2021)	94.7	86.8	81.8	61.0
MGSR-Net	94.6	85.3	86.9	66.2
MGSR-Net*	95.0	**88.0**	**89.6**	**71.4**

注:加粗字体表示各列最优结果。下划线字体表示对比方法里的最优值。

理复杂场景提供了强大而有效的泛化基础。

实验充分证明，该方法不仅在标准行人重识别任务上达到了领先水平，而且在高度贴近真实监控需求的复杂场景中展现出了性能优势。

3.4 消融实验

在 Occluded-REID 数据集上（new_0）的消融实验结果如表5所示，以 ResNet50 作为基线模型，逐步添加 SE-ResNet50 骨干网络、TSCE 软抑制擦除模块、门控机制、多视角分支和对齐损失，系统性地评估每个组件的贡献。

首先，将 ResNet50 替换为 SE-ResNet50，观察到 Rank-1 提升了 4.6%，mAP 提升了 6.1%。这说明 SE 模块中的通道注意力机制能够有效增强特征表示能力，对遮挡重识别尤为重要。

表5 在数据集 Occluded-REID 上对 MGSR-Net 各个组件的消融实验

Table 5 Ablation study on the individual components of MGSR-Net on the Occluded-REID dataset

骨干网络	擦除方式	门控机制	多视角分支	对齐损失	Rank-1/%	mAP/%
ResNet50	×	×	×	×	86.8	74.8
SE-ResNet50	×	×	×	×	91.4	80.9
SE-ResNet50	TSCE	×	×	×	**93.4**	81.9
SE-ResNet50	TSCE	√	×	×	92.8	81.8
SE-ResNet50	TSCE	√	√	×	93.0	78.9
SE-ResNet50	硬擦除	√	√	√	92.8	81.2
SE-ResNet50	TSCE	√	√	√	92.8	**83.4**

注：加粗字体表示各列最优结果。√、×分别代表包含、不包含该组件。

在 SE-ResNet50 基础上添加 TSCE 模块后，通过自适应擦除最显著的区域，迫使模型关注被忽略的次要特征，mAP 和 Rank-1 分别提升至 81.9% 和 93.4%。相较于仅用 SE-ResNet50，TSCE 带来了 1.0% 的 mAP 和 2.0% 的 Rank-1 提升。

然后，在 SE-ResNet50 + TSCE 基础上单独添加门控机制时，性能略有下降。说明门控机制作为特征选择器，在单视图训练中容易过度聚焦于某些区域，过滤掉有用的特征信息。

随后，添加多视角分支但未使用对齐损失时，mAP 下降了 2.9%。表明若缺乏跨视角约束可能会导致不同视图学习到相互冲突的特征表示，反而增加了模型的混淆程度。

当加入对齐损失后，性能有了显著回升。与仅使用多视角分支相比，mAP 提升了 4.5%，达到所有实验里最高的 83.4%。这验证了特征对齐机制的重要性，它通过将多视角信息蒸馏到单视图分支，确保了不同视图之间特征表示的一致性，从而实现了信息互补。

为验证软抑制策略相对于传统硬擦除的优越性，在保持门控、多视角分支与对齐损失的完整架构下，将 TSCE 模块替换为硬擦除，即将定位区域特征直接置零。实验结果显示，硬擦除版本的 mAP 为 81.2%，显著低于 TSCE 软抑制的 83.4%。这一对比有力地证明了本文提出的软抑制策略的有效性，适度的衰减可以保留部分梯度信息，避免训练振荡，并引导模型更平滑地学习多样化辅助特征。

同时，通过对比多视角分支无对齐损失和完整 MGSR-Net 两组实验，我们量化了对齐损失的独立贡献，即对齐损失带来了 4.5% 的 mAP 提升，证实了该模块在多视角学习中的必要性。此外，实验设置了专门的 SE-ResNet50 + TSCE +门控的单视图训练基线，完整 MGSR-Net 相较于该基线在 mAP 上提升 1.6%，这一差距展示了多视角正则化 + 蒸馏方案相对于单视图增强方案的增益。

上述结果表明，门控与多视角分支单独添加时效果有限，但二者与对齐损失结合形成完整 MGSR-Net 后，取得了最佳整体性能。在这一架构中，门控提供特征选择能力，多视角分支提供全面监督，二者结合避免了单视图下的过度聚焦，而对齐损失则起到协调与收敛的作用，使不同视图所提取的特征指向统一的语义表示，最终实现多视角信息的有效共享。

3.5 参数分析

为验证TSCE模块中关键参数设计的合理性，本节对衰减因子α、擦除比例ρ以及融合视图数量3个参数进行了敏感性分析，实验结果如图3所示。图3(a)展示了衰减因子α对模型性能的影响。结果表明，当$\alpha = 0.3$时，模型获得最佳的mAP与Rank-1性能，验证了软抑制策略的有效性，即适度的软抑制能够在提升模型对遮挡区域鲁棒性的同时避免过度特征损失；当$\alpha < 0.3$时，特征擦除过度，导致性能下降；当$\alpha > 0.3$时，擦除效果不足，难以有效提升模型对遮挡的适应性。图3(b)分析了擦除比例ρ对性能的影响，考虑到在实际遮挡情况下，行人被遮挡的区域比例通常不会超过50%，所以实验设置了0.1至0.5共5个梯度，可以看到，$\rho = 0.2$时，Rank-1达到峰值92.8%，mAP保持在82.6%的较高水平。该结果表明，擦除20%的高激活区域能够在保留足够判别性特征的前提下有效提升模型鲁棒性。当$\rho > 0.3$后性能显著下降，表明过大的擦除比例会破坏模型的特征提取能力。对于融合视图数量，选择了2、4、8三个具有代表性的配置进行对比分析，这是由于多视图融合的常见设计中，视图数量通常为2的幂次方，且这些数值能够确保训练批次大小的可整除性，避免数据分割不均的问题。如图3(c)所示，4个视图配置下，mAP达到最高值83.2%。综合上述分析，TSCE模块在$\alpha = 0.3$、$\rho = 0.2$、视图数为4的参数配置下实现了最佳性能平衡，为实验提供了可靠的参数基准。

3.6 可视化对比分析

为从视觉层面验证本文方法的有效性，图4通过类激活图可视化对比了本文方法与当前多视角遮挡行人重识别领域的代表性工作在无遮挡和有遮挡场景下的特征响应差异。

在无遮挡场景中，各方法基本都能定位到行人主体，但细节上差异明显。MVIIP与ETND-Net所生成的特征图较为杂乱，注意力难以集中，部分响应甚至扩散到背景区域，说明模型在聚焦关键特征方面仍不够稳定。潜在辅助融合方法LAF(latent-assisted fusion)虽然响应较为集中，但往往只盯着人体的少数几个区域，导致所学特征覆盖范围有限，多样性不足。相比之下，MGSR-Net不仅定位准确，还能均衡地关注从头部到腿部的多个判别区域，反映出模型对行人整体结构具有更完整的理解。

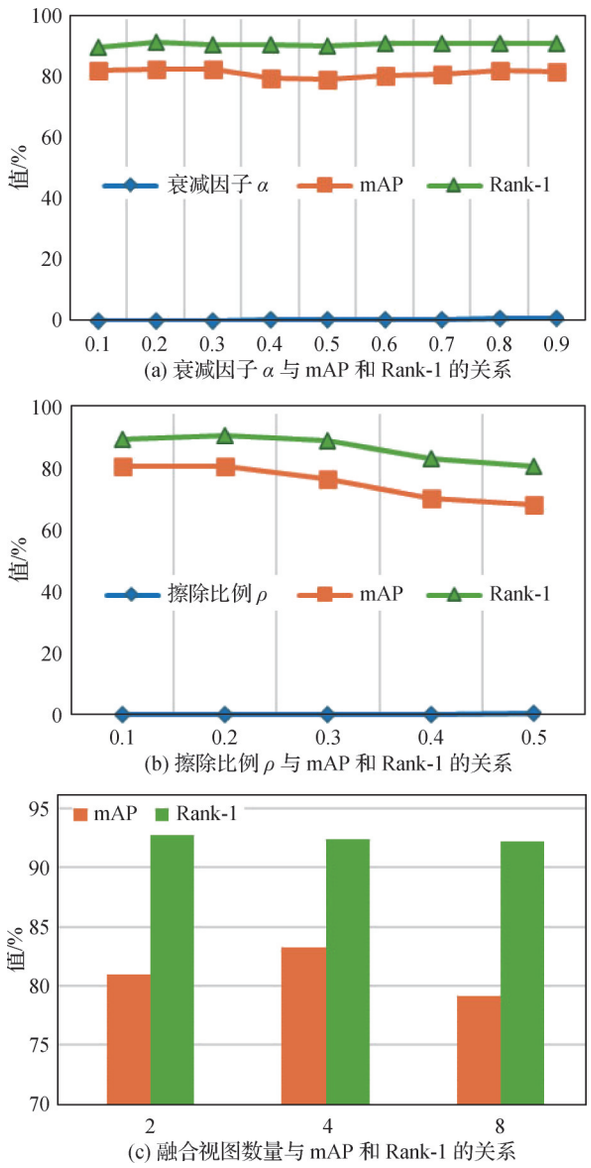


图3 TSCE模块中关键参数敏感性分析

Fig. 3 Sensitivity analysis of TSCE module hyperparameters

((a)relationship between fade factor α and mAP, Rank-1; (b)relationship between erase ratio ρ and mAP, Rank-1; (c)relationship between number of views and mAP, Rank-1)

当场景中出现严重遮挡时，不同方法的优劣对比变得更为分明。MVIIP与ETND-Net的注意力机制无法清晰区分遮挡物与人体本身，导致响应在两者之间来回摆动，难以有效捕捉仍可见的区域。LAF方法保留了一定的响应强度，却未能将注意力调整至其他可见部位。MGSR-Net则表现出优秀的应变能力，当上身被遮挡时，模型能够迅速将注意力转向下半身，如裤子、鞋部等未被遮挡的区域，从而保持稳定

的特征提取能力。这一结果充分验证了本文框架的设计初衷，即通过温和地抑制模型对显著区域的依赖，从而引导其发掘出更多样、更鲁棒的特征表示。



(a) 原图　(b) MVIIP　(c) ETND-Net　(d) LAF　(e) MGSR-Net

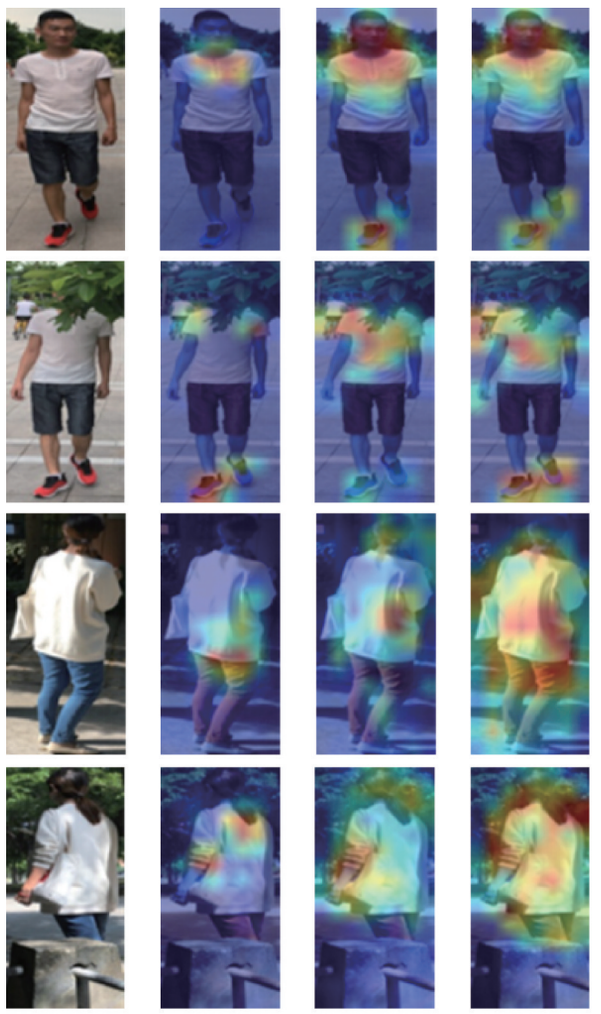
图 4　不同方法在遮挡行人重识别中的特征响应可视化对比

Fig. 4　Feature response visualization comparison of different methods in occluded person re-identification ((a)original images;(b)MVIIP;(c)ETND-Net;(d)LAF;(e)MGSR-Net)

为深入剖析本文方法获得上述优势的内在原因，通过图 5 展示了模型中各核心组件的消融实验可视化结果，清晰揭示了各模块的贡献。

在以 SE-ResNet50 作为骨干网络生成的激活图中，模型能够有效关注人体的显著区域，并准确定位其主要部位。但是其注意力分布相对集中，难以适应关键特征缺失的遮挡场景，这反映出单一注意力机制在实际应用中的局限性。

引入 TSCE 模块后，模型特征多样性有了显著的提升。通过平滑抑制最显著区域，模型注意力成功转移到腿部、手臂等非显著但同样具有判别力的身体部位。这种注意力重分布使模型在遮挡情况下仍能保持稳定的特征响应，验证了渐进式擦除机制在挖掘多样化辅助特征方面的有效性。



(a) 原图　(b) SE-ResNet50　(c) SE-ResNet50 + TSCE　(d) MGSR-Net

图 5　不同模块的消融实验可视化对比

Fig. 5　Visualization comparison of ablation studies for different modules((a)original images;(b)SE-ResNet50;(c)SE-ResNet50 + TSCE;(d)MGSR-Net)

进一步引入门控正则化机制后，完整的模型在保持广泛特征响应的同时，使注意力分布更为集中和精确。模型在复杂遮挡条件下表现出了更好的稳定性，能够准确指向未被遮挡的有效区域，证明了门控机制可以带来更明确、更稳定的特征决策。

这一系列对比直观验证了本文方法的逐模块改进效果，TSCE 模块有效拓展了特征挖掘的多样性，而门控正则化机制则显著提升了特征选择的稳定性，这些模块协同作用，共同奠定了 MGSR-Net 在复杂遮挡环境下保持优秀判别能力的基础。

4 结 论

针对遮挡行人重识别中因信息缺失和噪声干扰导致的特征判别性不足问题，本文提出了一种创新的多视角学习框架MGSR-Net。该框架通过3个核心组件的协同作用，解决了现有方法在单视图表征受限、特征增强盲目性与过拟合，以及特征选择稳定性不足等关键问题。

在特征增强层面，本研究引入的SE-ResNet50骨干网络通过通道注意力机制实现了自适应的重要通道筛选，使模型能够动态聚焦于最具判别力的特征维度。新设计的Top-k显著性引导软擦除模块突破了传统随机擦除的盲目性，通过基于类激活图的智能区域抑制，有效引导模型发掘被忽视的辅助特征，显著提升了特征的多样性；在正则化机制方面，门控正则化的引入通过熵约束优化，确保了特征选择掩码的稀疏性和稳定性，从而显著提升了模型在复杂遮挡场景下的鲁棒性。

在Occluded-REID和MSMT17等多个基准数据集上的实验结果表明，本框架在保持高效单视图推理的同时，显著超越了当前主流方法。消融研究验证了各核心模块的有效性，而可视化分析则从特征响应层面证实了本文方法在增强特征多样性和稳定性方面的优势。

本研究的理论价值在于提出了一种基于多视角协同优化的特征学习范式，为处理部分信息缺失的视觉任务提供了新的思路。在实践层面，该框架的核心思想可推广至其他存在类似挑战的计算机视觉领域。未来研究工作将致力于探索更精细的特征融合机制以替代当前的固定加权策略，更智能地整合多视角信息，以及更高效的知识传递架构，进而优化从多视角分支到单视角分支的信息流，在提升性能的同时降低模型复杂度。

参考文献(References)

Che J, Zhang Y X, Yang Q and He Y T. 2023. Research on person re-identification based on posture guidance and feature alignment. Multimedia Systems, 29(2): 763-770 [DOI: 10.1007/s00530-022-01016-3]

Chen X M, Zheng X T and Lu X Q. 2021. Bidirectional interaction network for person re-identification. IEEE Transactions on Image Processing, 30: 1935-1948 [DOI: 10.1109/TIP.2021.3049943]

Dong N, Yan S L, Tang H, Tang J H and Zhang L Y. 2024a. Multi-view information integration and propagation for occluded person re-identification. Information Fusion, 104: #102201 [DOI: 10.1016/j.inffus.2023.102201]

Dong N, Zhang L Y, Yan S L, Tang H and Tang J H. 2024b. Erasing, transforming, and noising defense network for occluded person re-identification. IEEE Transactions on Circuits and Systems for Video Technology, 34(6): 4458-4472 [DOI: 10.1109/TCSVT.2023.3339167]

Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X H, Unterthiner T, et al. 2021. An image is worth 16 × 16 words: transformers for image recognition at scale//Proceedings of the 9th International Conference on Learning Representations. [s.l.]: OpenReview.net

Dung D A and Nakamura Y. 2024. DSPI — dual semantic parsing image: a robust person representation for person re-identification//Proceedings of the 10th International Conference on Communications and Electronics (ICCE). Danang, Vietnam: IEEE: 613-618 [DOI: 10.1109/ICCE62051.2024.10634704]

Gao S, Wang J Y, Lu H C and Liu Z M. 2020. Pose-guided visible part matching for occluded person ReID//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, USA: IEEE: 11741-11749 [DOI: 10.1109/CVPR42600.2020.01176]

He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]

He S T, Luo H, Wang P C, Wang F, Li H and Jiang W. 2021. TransReID: transformer-based object re-identification//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 14993-15002 [DOI: 10.1109/ICCV48922.2021.01474]

Hinton G, Vinyals O and Dean J. 2015. Distilling the knowledge in a neural network [EB/OL]. [2025-11-14]. https://arxiv.org/pdf/1503.02531.pdf

Hou R B, Ma B P, Chang H, Gu X Q, Shan S G and Chen X L. 2019. Interaction-and-aggregation network for person re-identification//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 9309-9318 [DOI: 10.1109/CVPR.2019.00954]

Hu J, Shen L and Sun G. 2018. Squeeze-and-excitation networks//Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Salt Lake City, USA: IEEE: 7132-7141 [DOI: 10.1109/CVPR.2018.00745]

Huang M Y, Hou C P, Yang Q Y and Wang Z P. 2023. Reasoning and tuning: graph attention network for occluded person re-identification.

IEEE Transactions on Image Processing, 32: 1568-1582 [DOI: 10.1109/TIP.2023.3247159]

Jia M X, Cheng X H, Lu S J and Zhang J. 2023. Learning disentangled representation implicitly via transformer for occluded person re-identification. IEEE Transactions on Multimedia, 25: 1294-1305 [DOI: 10.1109/TMM.2022.3141267]

Jiang J G, Yang N, Qi M B and Chen C Q. 2019. Person re-identification with region block segmentation and fusion. Journal of Image and Graphics, 24(4): 513-522 (蒋建国, 杨宁, 齐美彬, 陈翠群. 2019. 区域块分割与融合的行人再识别. 中国图象图形学报, 24(4): 513-522) [DOI: 10.11834/jig.180370]

Kim M, Cho M A, Lee H, Cho S and Lee S. 2022. Occluded person re-identification via relational adaptive feature correction learning// Proceedings of 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Singapore, Singapore: IEEE: 2719-2723 [DOI: 10.1109/ICASSP43922.2022.9746734]

Kingma D P and Ba J. 2015. Adam: a method for stochastic optimization//Proceedings of the 3rd International Conference on Learning Representations. San Diego, USA: #ICLR

Li A, Shao C R, Zhou L J, Wang Y B and Gao T Y. 2025a. Complete object feature diffusion network for occluded person re-identification. Digital Signal Processing, 159: #104998 [DOI: 10.1016/j. dsp. 2025.104998]

Li M L, Tao Z Y, Lin S and Feng K H. 2025b. LAF: enhancing person re-identification via latent-assisted feature fusion. Alexandria Engineering Journal, 127: 116-128 [DOI: 10.1016/j.aej.2025.04.075]

Miao J X, Wu Y, Liu P, Ding Y H and Yang Y. 2019. Pose-guided feature alignment for occluded person re-identification//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE: 542-551 [DOI: 10.1109/ICCV.2019.00063]

Ning E H, Li W F, Fang J W, Yuan J C, Duan Q H and Wang G. 2025. 3D-guided multi-feature semantic enhancement network for person re-ID. Information Fusion, 117: #102863 [DOI: 10.1016/j.inffus. 2024.102863]

Ning E H, Wang C S, Zhang H, Ning X and Tiwari P. 2024. Occluded person re-identification with deep learning: a survey and perspectives. Expert Systems with Applications, 239: #122419 [DOI: 10. 1016/j.eswa.2023.122419]

Shorten C and Khoshgoftaar T M. 2019. A survey on image data augmentation for deep learning. Journal of Big Data, 6(1): #60 [DOI: 10. 1186/s40537-019-0197-0]

Shu R, Bui H H, Narui H and Ermon S. 2018. A DIRT-T approach to unsupervised domain adaptation//Proceedings of the 6th International Conference on Learning Representations. Vancouver, Canada: ICLR: 1-16

Song W F, Li S, Chang T, Hao A M, Zhao Q P and Qin H. 2020. Context-interactive CNN for person re-identification. IEEE Transactions on Image Processing, 29: 2860-2874 [DOI: 10.1109/TIP. 2019.2953587]

Sun Y F, Zheng L, Yang Y, Tian Q and Wang S J. 2018. Beyond part models: person retrieval with refined part pooling (and a strong convolutional baseline)//Proceedings of the 15th European Conference on Computer Vision (ECCV). Munich, Germany: Springer: 501-518 [DOI: 10.1007/978-3-030-01225-0_30]

Tan H C, Liu X P, Bian Y H, Wang H S and Yin B C. 2022. Incomplete descriptor mining with elastic loss for person re-identification. IEEE Transactions on Circuits and Systems for Video Technology, 32(1): 160-171 [DOI: 10.1109/TCSVT.2021.3061412]

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, et al. 2017. Attention is all you need//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, USA: Curran Associates Inc.: 6000-6010

Wang P F, Ding C X, Shao Z Y, Hong Z B, Zhang S L and Tao D C. 2023. Quality-aware part models for occluded person re-identification. IEEE Transactions on Multimedia, 25: 3154-3165 [DOI: 10.1109/TMM.2022.3156282]

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]

Wei L L, Wang Z X, Lang C Y, Liang L Q, Wang T, Feng S H, et al. 2024. Linkage-based object re-identification via graph learning. IEEE Transactions on Intelligent Transportation Systems, 25(10): 13040-13050 [DOI: 10.1109/TITS.2024.3422286]

Wu J J, Zhong Z, Guo Y R, Hu S J and Hong R C. 2025. Person re-identification with arbitrary modalities: a multi-modal dataset and a unified framework. IEEE Transactions on Information Forensics and Security, 20: 6953-6967 [DOI: 10.1109/TIFS. 2025. 3583887]

Ye M, Shen J B, Lin G J, Xiang T, Shao L and Hoi S C. 2022. Deep learning for person re-identification: a survey and outlook. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(6): 2872-2893 [DOI: 10.1109/TPAMI.2020.3012837]

Zhang G Q, Yang Y, Zheng Y H, Martin G and Wang R L. 2025a. Mask-aware hierarchical aggregation transformer for occluded person re-identification. IEEE Transactions on Circuits and Systems for Video Technology, 35(6): 5821-5832 [DOI: 10.1109/TCSVT. 2025.3531142]

Zhang H Y, Liu T F, Luo Q and Zhang T. 2024. Pose guidance and multi-scale feature fusion for occluded person re-identification. Journal of Image and Graphics, 29(8): 2364-2376 (张红颖, 刘腾飞, 罗谦, 张涛. 2024. 融合姿态引导和多尺度特征的遮挡行人重识别. 中国图象图形学报, 29(8): 2364-2376) [DOI: 10.11834/jig. 230523]

Zhang Y F, Yang H Y, Zhang Y J, Dou Z P, Liao S C, Zheng W S, et al. 2023. Recent progress in person re-ID. Journal of Image and

Graphics, 28(6): 1829-1862 (张永飞, 杨航远, 张雨佳, 豆朝鹏, 廖胜才, 郑伟诗, 等. 2023. 行人再识别技术研究进展. 中国图象图形学报, 28(6): 1829-1862) [DOI: 10.11834/jig.230022]

Zhang Y Z, Yang Y H, Kang W L and Zhen J W. 2025b. Cross-erasure enhanced network for occluded person re-identification. Pattern Recognition Letters, 193: 108-114 [DOI: 10.1016/j.patrec.2025.04.015]

Zhao C R, Lyu X B, Dou S G, Zhang S S, Wu J and Wang L. 2021. Incremental generative occlusion adversarial suppression network for person ReID. IEEE Transactions on Image Processing, 30: 4212-4224 [DOI: 10.1109/TIP.2021.3070182]

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 (ICCV). Santiago, Chile: IEEE: 1116-1124 [DOI: 10.1109/ICCV.2015.133]

Zheng Y F, Wang W J, Gan W J and Liu J W. 2025. Occlusion-guided feature purification learning via reinforced knowledge distillation for occluded person re-identification [EB/OL]. [2025-11-14]. https://arxiv.org/pdf/2507.08520.pdf

Zheng Z D, Yang X D, Yu Z D, Zheng L, Yang Y and Kautz J. 2019. Joint discriminative and generative learning for person re-identification//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 2133-2142 [DOI: 10.1109/CVPR.2019.00224]

Zhong Z, Zheng L, Kang G L, Li S Z and Yang Y. 2020. Random erasing data augmentation//Proceedings of the 34th AAAI Conference on Artificial Intelligence. New York, USA: AAAI: 13001-13008 [DOI: 10.1609/aaai.v34i07.7000]

Zhou B L, Khosla A, Lapedriza A, Oliva A and Torralba A. 2016. Learning deep features for discriminative localization//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 2921-2929 [DOI: 10.1109/CVPR.2016.319]

Zhou K Y, Yang Y X, Cavallaro A and Xiang T. 2019. Omni-scale feature learning for person re-identification//Proceedings of 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea (South): IEEE: 3701-3711 [DOI: 10.1109/ICCV.2019.00380]

Zhu K, Guo H Y, Liu Z W, Tang M and Wang J Q. 2020. Identity-guided human semantic parsing for person re-identification//Proceedings of the 16th European Conference on Computer Vision. Glasgow, UK: Springer: 346-363 [DOI: 10.1007/978-3-030-58580-8_21]

Zhuo J X, Chen Z Y, Lai J H and Wang G C. 2018. Occluded person re-identification//Proceedings of 2018 IEEE International Conference on Multimedia and Expo (ICME). San Diego, USA: IEEE: 1-6 [DOI: 10.1109/ICME.2018.8486568]

作者简介

叶学义,男,副教授,硕士生导师,主要研究方向为图像处理、模式识别和信息安全。E-mail: xueyiye@hdu.edu.cn

薛智权,男,硕士研究生,主要研究方向为图像处理。E-mail: a879764330@163.com

蒋德琦,男,硕士研究生,主要研究方向为图像处理。E-mail:232080107@hdu.edu.cn

戴渝琛,男,硕士研究生,主要研究方向为步态识别。E-mail:232080133@hdu.edu.cn

王佳欣,女,硕士研究生,主要研究方向为图像处理。E-mail:wjx15168134014@163.com