Computer Science and Application 计算机科学与应用, 2026, 16(4), 439-452
Published Online April 2026 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2026.164143

Hans汉斯

双分支网络下监督通道注意力机制的行人重识别

魏子仪[1]，杨海波[1,2]，芦俊池[1*]

[1]沈阳工业大学信息科学与工程学院，辽宁 沈阳
[2]先进计算与信创技术重点实验室，辽宁 沈阳

收稿日期：2026年3月24日；录用日期：2026年4月22日；发布日期：2026年4月30日

摘　要

行人重识别(ReID)技术旨在跨摄像头图像或视频序列识别特定个体，针对不同图像中行人比例各异的核心挑战，本文提出双分支监督通道注意力模型(DSAM)。该方法在ResNet-50的瓶颈模块嵌入无参监督注意力模块，结合分类层权重矩阵作为监督信号，通过均值与方差动态调整通道权重，有效抑制背景干扰、聚焦关键特征区域。基于ResNet-50构建双分支网络，分别输入原始图像与裁剪图像以增强主体特征显著性；同时在conv3、conv4层引入CBAM模块，通过特征融合模块(FFM)将其增强特征与conv5高层语义特征跨尺度整合，实现低层细节与高层语义的互补，提升模型对尺度变化、视角差异及光照波动的鲁棒性。在MSMT17、Market-1501和DukeMTMC-ReID数据集上的实验表明，所提方法的mAP与Rank-1指标均优于现有方法。

关键词

行人重识别，特征融合，双分支，监督注意力

Supervised Channel Attention Learning in Dual-Branch Networks for Person Re-Identification

Ziyi Wei[1], Haibo Yang[1,2], Junchi Lu[1*]

[1]School of Information Science and Engineering, Shenyang University of Technology, Shenyang Liaoning
[2]Key Laboratory of Advanced Computing and Application Innovation Technology, Shenyang Liaoning

Received: March 24, 2026; accepted: April 22, 2026; published: April 30, 2026

*通讯作者。

文章引用: 魏子仪, 杨海波, 芦俊池. 双分支网络下监督通道注意力机制的行人重识别[J]. 计算机科学与应用, 2026, 16(4): 439-452. DOI: 10.12677/csa.2026.164143

Abstract

Person re-identification (ReID) technology aims to identify and track specific individuals through cross-camera images or video sequences. In order to solve the challenge of the different proportions of persons in different images, this paper proposes a ReID method based on a Dual-Branch Supervised Channel Attention Model (DSAM). DSAM embeds a parameter-free supervised attention module into the bottleneck blocks of ResNet-50 and leverages the classification-layer weight matrix as a supervisory signal. Channel-wise weights are dynamically adjusted via channel mean and variance, effectively suppressing background noise and focusing on salient person regions. Built on ResNet-50, DSAM adopts a dual-branch architecture that feeds both original and randomly resized images to enhance foreground distinctiveness. CBAM (Convolutional Block Attention Modules) are inserted after conv3 and conv4, and their augmented features are cross-scale fused with high-level semantic features from conv5 through a Feature Fusion Module (FFM). This complementary integration of low-level details and high-level semantics improves robustness against scale variations, viewpoint changes, and illumination fluctuations. Experiments on MSMT17, Market-1501 and DukeMTMC-ReID datasets show that the proposed method is superior to the existing methods in mAP and Rank-1 metrics.

Keywords

Person Re-Identification, Feature Fusion, Dual-Branch, Supervised Attention

1. 引言

行人重识别(ReID) [1]技术旨在利用图像或视频序列在非重叠的摄像头监控区域内识别并跟踪特定个体。随着城市化进程的加快和公共安全需求的增长，ReID 在智能视频监控和安全系统中变得愈发重要。随着深度学习技术的显著进步，ReID 已取得重大进展。由于人与摄像头的距离不同，同一目标在不同摄像头中的尺寸存在显著差异。即使经过统一缩放，目标在图像中的比例仍有明显波动。这种尺度变化使得模型难以稳定提取具有判别性的特征，尤其是在遮挡或低分辨率场景中。

早期研究通过多尺度特征融合增强模型对尺度变化的适应性。例如，Wang 等人[2]设计的 MGN 通过两个局部特征提取分支将特征分成不同的条纹。Bai 等人[3]通过将行人图像分割成多个部分来解决多尺度信息整合问题。这些部分通过 LSTM [4]作为局部序列进行处理。Zhang 等人[5]引入了区域感知全局注意力。这种方法对特征节点之间的关系进行建模，通过改进语义推理，显著增强了模型处理跨摄像头尺度变化的能力。

现有方法通过全局注意力关系建模与多尺度特征融合，提升了模型对人物特征的表征能力。然而，人物与相机之间的距离变化会导致人物在图像中的占比产生差异，这一问题影响了模型的泛化能力。具体而言，当所有图像被等比缩放到同一尺寸时，目标行人在画面中的占比会因场景差异而显著不同，导致其视觉尺寸在多张图像间剧烈波动。这种尺寸变化对基于深度学习的模型构成挑战——网络必须能够在不同尺度下仍准确捕获并适应行人特征，以维持识别的精度与鲁棒性。为此，我们提出一种面向 ReID 的双分支网络模型。

为进一步改善由视角变化引发的尺度给 ReID 带来挑战，我们提出一种双分支网络模型。

本文所提模型的主要贡献如下：

(1) 提出双分支监督通道注意力模型(Dual-Branch Supervised Channel Attention Model，简称 DSAM)。该模型通过双分支结构适配不同尺度的图像，以精准实现 ReID 任务。

(2) 提出瓶颈无参监督注意力模块(Bottleneck Parameter-free Supervised Attention Module，简称 BSAM)。该模块结合分类层权重矩阵作为监督信号，通过均值与方差动态调整通道权重，同时避免了参数冗余对模型泛化能力的负面影响。

(3) 设计特征融合模块(Feature Fusion Module)，将通过注意力模块的特征的中层细节与高层语义互补融合，提升模型对尺度、视角变化的鲁棒性。

(4) 设计了一种基于注意力图的损失函数。该函数通过衡量原始图像注意力图与增强后图像注意力图之间的差异，增强了模型对关键区域的聚焦能力，进而提升了模型的鲁棒性。

2. 网络模型

2.1. 双分支监督注意力模型

在我们的网络架构设计中，引入了 DSAM 模型，具体如图 1 所示。

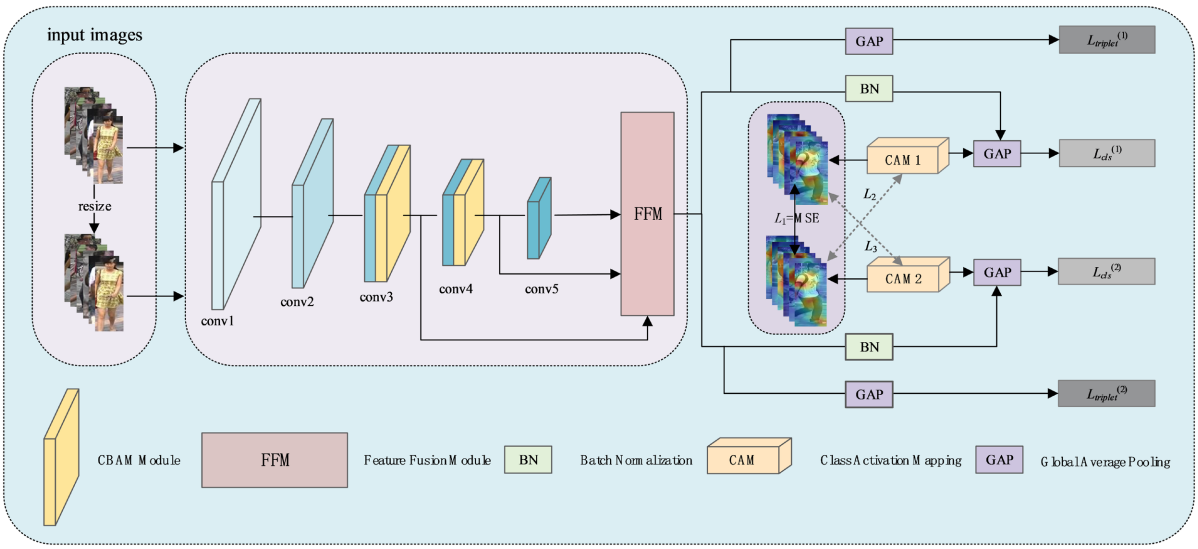

Figure 1. The DSAM model
图 1. DSAM 模型

该模型以 ResNet-50 [6]为基础架构，首先对输入图像执行 resize 操作生成多尺度输入，把两种输入图像经过网络作为两个分支，其中 conv3 与 conv4 模块后嵌入 CBAM [7]注意力模块，通过通道与空间注意力的联合加权强化关键特征的表达能力；为了避免信息的重复使用，仅选取主干网络中三个尺度输出(3、4、5)的行人特征经特征融合模块(FFM)完成信息整合。双分支分别通过融合特征直接通过全局平均池化(GAP)生成全局特征向量，由 $L_{triplet}^{(1)}$ 与 $L_{triplet}^{(2)}$ 分别监督其度量学习能力；另一方面，融合特征先经批量归一化(BN)稳定分布，经 GAP 转化为细粒度特征向量后，由 $L_{cls}^{(1)}$ 与 $L_{cls}^{(2)}$ 监督其身份分类能力。双分支得到不同尺度的子特征图，随后通过两个独立的类别激活映射模块定位行人判别性区域，计算两分支均方差损失。

2.2. 瓶颈无参监督注意力模块

针对常规的残差网络无法为图像的关键部分分配较高权重的问题，本文引入瓶颈无参监督注意力模块(BSAM)，如图 2 所示，以此强化模型对关键特征的表达能力。

瓶颈无参监督注意力模块融合了无参通道注意力的 Bottleneck 模块，在标准 Bottleneck 的两个 1 × 1 卷积和一个 3 × 3 卷积残差分支上，每次卷积后依次接入 BN 层与 ReLU 层，以缓解梯度消失并加速收敛；无参通道注意力(SCAM)被置于残差相加之前，在不引入额外参数的前提下对通道权重进行重分配，借助 CAM [8]提供的监督信号增强注意力区域的判别力，从而突出对当前任务贡献更大的特征，其具体实现如图 3 所示。

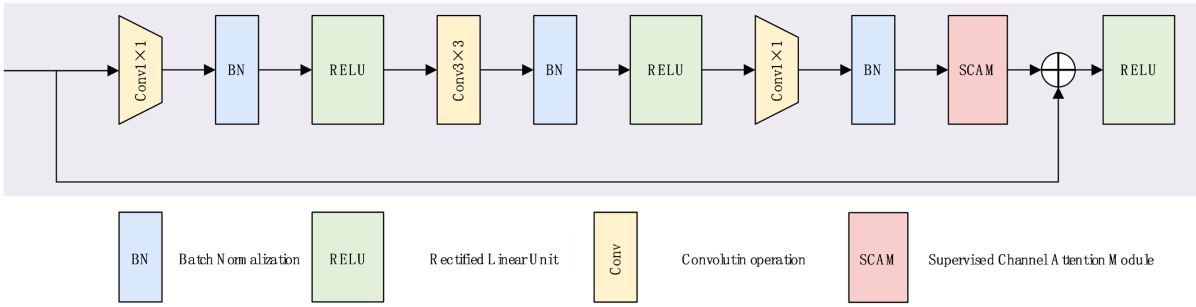

Figure 2. Bottleneck parameter-free supervised attention module
图 2. 瓶颈无参监督注意力模块

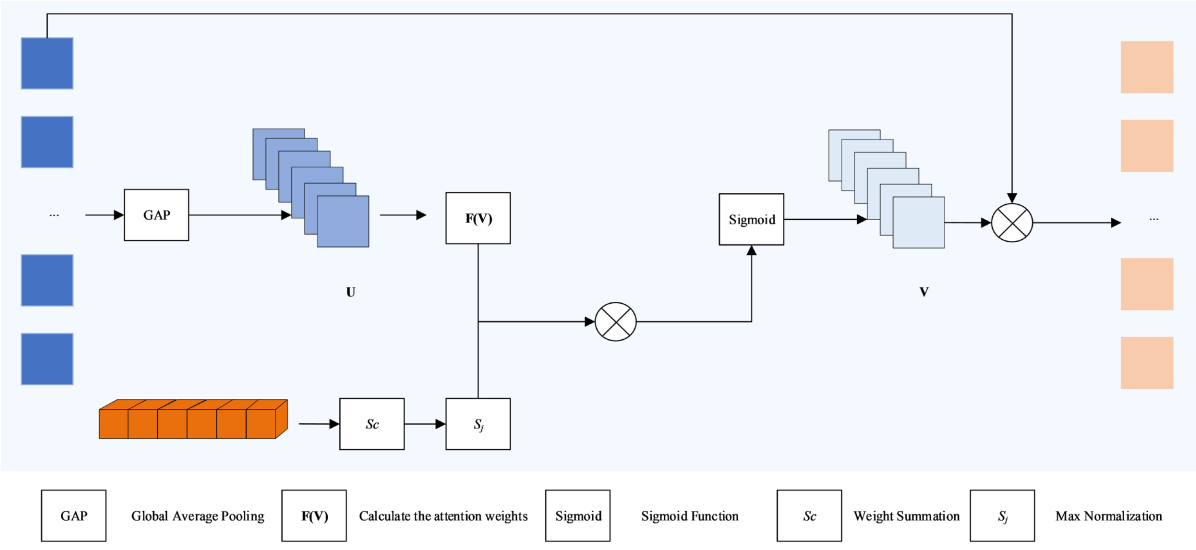

Figure 3. Supervised channel attention module
图 3. 监督通道注意力模块

输入特征图 $\mathbf{X}\in \mathrm{R}^{C\times H\times W}$，首先通过 GAP 层，将空间维度压缩，得到通道级特征向量 $\mathbf{V}\in \mathrm{R}^{C\times 1\times 1}$。随后，依据统计量 $\mathbf{U}$ 的均值 μ 和方差 σ^2 计算初始注意力权重：

$$\mathbf{V}_j=\frac{\left(\mathbf{U}_j-\mu\right)^2+2\left(\sigma^2+\lambda\right)}{4\left(\sigma^2+\lambda\right)},\ j=1,2,\cdots,C \tag{1}$$

其中，λ 为用于稳定方差的小常数(通常设为 1e−4)。$\left(\mathbf{U}_j-\mu\right)^2$ 衡量第 j 个通道与全局分布的偏离度，偏离越大则权重越高，以突出判别性显著的通道特征。σ^2 为通道特征的全局方差，分母 $4\left(\sigma^2+\lambda\right)$ 实现权重归一化，避免权重失衡；分子中的 $2\left(\sigma^2+\lambda\right)$ 提供基础权重，确保低差异通道不被完全抑制，同时正则项 λ 防止方差为 0 时，差异较小的通道特征被完全抑制。该公式通过建模通道内特征分布差异来量化每个通道的重要性，且无需任何可训练参数。

之后，我们引入分类层权重矩阵作为监督信号。其物理含义为：$\mathbf{W}_{c,n}$ 代表第 c 个通道对第 n 类预测结果的贡献权重。对于每个通道 c，通过对所有类别下的贡献权重取绝对值后求和，计算该通道的全局重要性，公式如：

$$S_c=\sum_{n=1}^{N}\left|\mathbf{W}_{c,n}\right| \tag{2}$$

绝对值求和可避免正负权重之间的相互抵消，直接反映通道的全局重要性。将通道全局重要性指标归一化到[0, 1]区间：

$$S_j=\frac{S_c}{\max\left(S\right)},S=S_1,S_2,\cdots,S_c \tag{3}$$

归一化操作可消除通道间的尺度差异，确保注意力权重的均衡调整。将归一化后的信号 S_j 与原始注意力权重进行通道维度的逐元素相乘，生成监督增强型注意力权重：

$$\mathbf{V}=\mathbf{V}_j\cdot S_j \tag{4}$$

经过改进的注意力权重通过 Sigmoid 函数激活后，与输入特征图 $\mathbf{X}$ 进行通道维度的逐元素相乘，得到最终输出：

$$\mathbf{Y}=\mathbf{X}\cdot\mathrm{Sigmoid}\left(\mathbf{V}\right) \tag{5}$$

2.3. 注意力模块

同一摄像头常捕捉到衣着与动作高度相似的不同行人，若网络聚焦于摄像头固定的背景，极易造成误识。为此，主干网络引入 CBAM 注意力，使网络关注行人关键区域，抑制背景噪声，从而提取更具判别力的特征。CBAM 为轻量级前馈模块，如图 4 所示，它顺序集成通道注意力和空间注意力来增强卷积神经网络的特征表示能力。

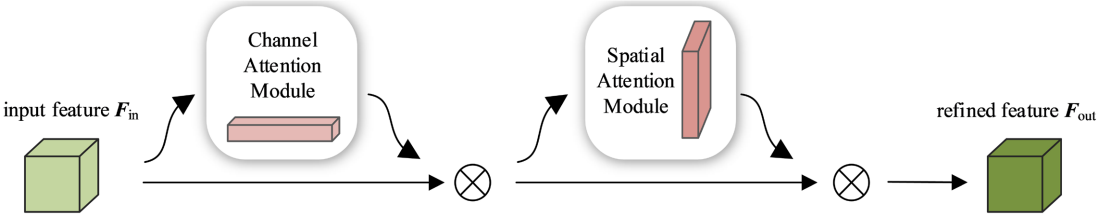

Figure 4. The CBAM structure
图 4. CBAM 结构

(1) 通道注意力

通道注意力机制通过重新加权各输入通道，使富含目标区域的关键通道获得更高响应。如图 5 所示，输入特征首先进行平均池化和最大池化的并行操作，以减少有效信息的丢失，生成 2 个大小为 1 × 1 × c 的不同特征图 $\boldsymbol{F}_{\mathrm{avg}}^{\mathrm{c}}$ 和 $\boldsymbol{F}_{\mathrm{max}}^{\mathrm{c}}$；随后，将特征图送入 2 层共享神经网络(sharedMLP)，执行元素求和及激活操

作，最终生成通道注意力图 $\boldsymbol{M}_{\mathrm{c}}$。通道注意力计算为：

$$\begin{aligned}\boldsymbol{M}_{\mathrm{c}}\left(\boldsymbol{F}_{\mathrm{in}}\right)&=\sigma\left(\mathrm{MLP}\left(\mathrm{AvgPool}\left(\boldsymbol{F}_{\mathrm{in}}\right)\right)+\mathrm{MLP}\left(\mathrm{MaxPool}\left(\boldsymbol{F}_{\mathrm{in}}\right)\right)\right)\\&=\sigma\left(\boldsymbol{W}_1\left(\boldsymbol{W}_0\left(\boldsymbol{F}_{\mathrm{avg}}^{\mathrm{c}}\right)\right)+\boldsymbol{W}_1\left(\boldsymbol{W}_0\left(\boldsymbol{F}_{\mathrm{max}}^{\mathrm{c}}\right)\right)\right)\end{aligned}\tag{6}$$

式中：σ 为 Sigmoid 激活函数；$\boldsymbol{W}_0\in\mathrm{R}^{C/r\times C}$ 和 $\boldsymbol{W}_1\in\mathrm{R}^{C\times C/r}$ 为 MLP 共享权重。

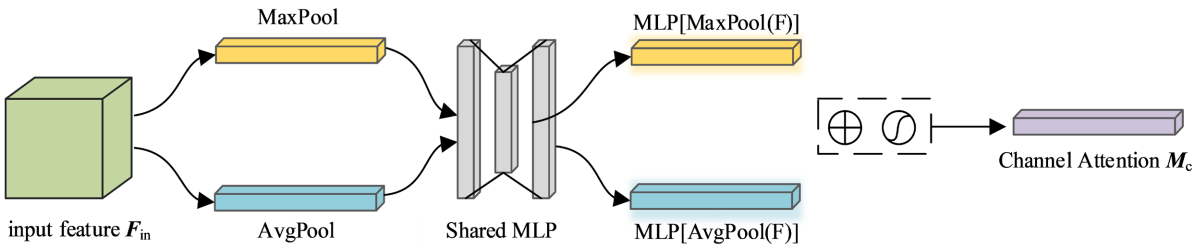

Figure 5. Channel attention
图 5. 通道注意力模块

(2) 空间注意力

不同于通道注意力，空间注意力“在哪里”的位置信息，与前者互补。空间注意力将通道注意力特征图 $\boldsymbol{M}_{\mathrm{c}}$ 与原始输入特征图 $\boldsymbol{F}_{\mathrm{in}}$ 相乘的结果作为输入，如图 6 所示。首先进行最大池化和平均池化的通道轴操作，分别生成 2 个不同的 $1\times H\times W$ 特征图 $\boldsymbol{F}_{\mathrm{max}}^{\mathrm{s}}$ 和 $\boldsymbol{F}_{\mathrm{avg}}^{\mathrm{s}}$，利用 7×7 卷积层连接、卷积突出区域，再由激活层得空间注意力图 $\boldsymbol{M}_{\mathrm{s}}$，将其与原始输入特征图 $\boldsymbol{F}'$ 相乘生成最终的特征图 $\boldsymbol{F}_{\mathrm{out}}$。空间注意力计算为：

$$\boldsymbol{M}_{\mathrm{s}}\left(\boldsymbol{F}'\right)=\sigma\left(f^{7\times7}\left(\left[\mathrm{AvgPool}\left(\boldsymbol{F}'\right);\mathrm{MaxPool}\left(\boldsymbol{F}'\right)\right]\right)\right)=\sigma\left(f^{7\times7}\left(\left[\boldsymbol{F}_{\mathrm{avg}}^{\mathrm{s}};\boldsymbol{F}_{\mathrm{max}}^{\mathrm{s}}\right]\right)\right)\tag{7}$$

式中：$f^{7\times7}$ 是滤波器大小为 7×7 的卷积运算。

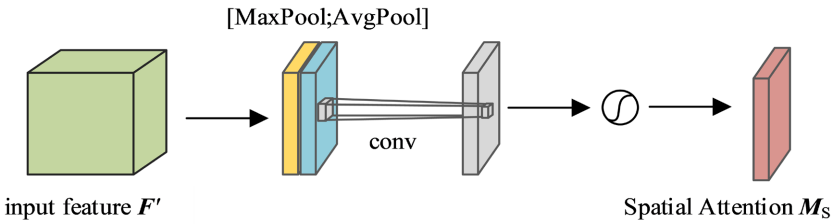

Figure 6. Spatial attention module
图 6. 空间注意力模块

2.4. 特征融合模块

注意力融合分支旨在利用 DSAM 中不同层级的显著特征，注意力融合模块的结构如图 7 所示。

输入特征图 $\boldsymbol{F}_3$、$\boldsymbol{F}_4$ 分别由注意力模块 CBAM 生成，$\boldsymbol{F}_5$ 则是 DSAM 最后一个卷积层的输出。为更好地利用这些显著特征，将 $\boldsymbol{F}_5$ 分别与 $\boldsymbol{F}_3$ 和 $\boldsymbol{F}_4$ 进行拼接；为匹配 $\boldsymbol{F}_3$ 的尺寸，通过双线性插值对 $\boldsymbol{F}_5$ 进行了尺寸调整。拼接完成后，对复合特征图 $\boldsymbol{F}_{5\text{-}3}$ 和 $\boldsymbol{F}_{5\text{-}4}$ 依次执行全局池化操作，及由 1 × 1 卷积块、BN 层和 ReLU 激活函数构成的降维层进行处理。

$$\boldsymbol{F}_{5\text{-}3}=concat\left(\boldsymbol{F}_5,\boldsymbol{F}_3\right)\tag{8}$$

$$\boldsymbol{F}_{5\text{-}4} = concat\left(\boldsymbol{F}_5, \boldsymbol{F}_4\right) \tag{9}$$

$$f_{5a} = concat\left(R_{35}\left(G\left(\boldsymbol{F}_{5\text{-}3}\right)\right), R_{45}\left(G\left(\boldsymbol{F}_{5\text{-}4}\right)\right)\right) \tag{10}$$

其中，*concat*、*G* 和 *R* 分别代表拼接操作、全局池化操作和降维层；f_{5a} 代表注意力融合分支的最终输出。

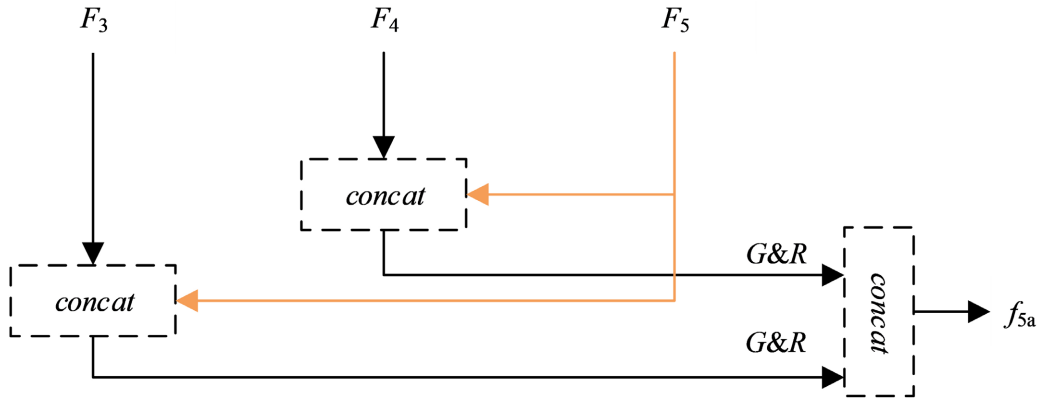

Figure 7. Feature fusion module
图 7. FFM

2.5. 损失函数

每个分支的输出均通过全连接层进行分类，且我们对每个分支的分类结果采用标准交叉熵损失。采用 $(\cdot)^{(1,2)}$ 来区分两个分支的损失。其交叉熵损失函数定义如下：

$$L_{cls}^{(1,2)} = \frac{1}{N}\sum_{i=1}^{N} CE\left(y_i, \hat{y}_i\right) \tag{11}$$

N 表示批次(batch)中的样本数量，y_i 为第 *i* 个样本的真实标签，该标签通常采用独热编码形式；$\hat{y}_i$ 则表示第 *i* 个样本的预测概率分布。

我们对每个分支的输出特征应用三元组损失其公式定义如下：

$$L_{triplet}^{(1,2)} = \max\left(d\left(f\left(x_a\right), f\left(x_p\right)\right) - d\left(f\left(x_a\right), f\left(x_n\right)\right) + \alpha, 0\right) \tag{12}$$

其中 x_a 表示输入样本，x_p 代表与锚点样本属于同一身份的样本，x_n 代表与锚点样本属于不同身份的样本。$f(\cdot)$ 表示将样本映射到特征空间的函数，$(\cdot,\cdot)$ 用于计算两个特征向量之间的余弦距离。α 为边界值，$\max(\cdot,0)$ 是返回最大值的操作。

我们在原始损失的基础上，为双分支增加了一项注意力一致性约束，计算两个分支生成的热力图之间的均方偏差损失。我们对来自最后一个卷积层的特征图 $\boldsymbol{F} \in \mathrm{R}^{C\times H\times W}$，(*C*，*H*，and *W* 分别表示特征图的通道数、高度和宽度)应用 GAP 操作。

经池化后的特征随后被传入最终输出层，其权重矩阵 $\mathbf{W} \in \mathrm{R}^{L\times C}$ (其中 *L* 为标签数量)，用于完成分类任务。注意力热力图则通过 CAM，对所有通道特征进行线性组合计算得到：

$$\boldsymbol{M}_j\left(m,n\right) = \sum_{k=1}^{C} \mathbf{W}\left(j,k\right)\boldsymbol{F}_k\left(m,n\right) \tag{13}$$

其中，$\boldsymbol{M}_j\left(m,n\right)$ 表示标签 *j* 在空间位置 (m,n) 处的注意力热力图，$\mathbf{W}\left(j,k\right)$ 代表特征图第 *k* 个通道与标签 *j* 相关联的权重，$\boldsymbol{F}_k\left(m,n\right)$ 表示最后一个卷积层第 *k* 个通道在空间位置 (m,n) 处的特征图。我们用 $\boldsymbol{M}_{heat}$ 表

示两个分支中的注意力热力图，并引入如下 CAM 损失：

$$L_1 = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \left(\boldsymbol{M}_{heat}^{(1)} - \boldsymbol{M}_{heat}^{(2)} \right)^2 \tag{14}$$

我们采用交叉监督机制，使特征图中的高激活区域与热力图预测的关键区域保持一致。在每个分支内部，计算热力图与注意力图之间的均方偏差损失，一致性损失定义如下：

$$\boldsymbol{F}_{att} = \boldsymbol{F} \odot \frac{1}{C} \sum_{c=1}^{C} \boldsymbol{F}_{b,c,h,w} \tag{15}$$

$$L_2 = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \left(\boldsymbol{M}_{heat}^{(1)} - \boldsymbol{F}_{att}^{(2)} \right)^2 \tag{16}$$

$$L_3 = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \left(\boldsymbol{M}_{heat}^{(2)} - \boldsymbol{F}_{att}^{(1)} \right)^2 \tag{17}$$

$\boldsymbol{F}_{b,c,h,w}$ 表示输入特征图中第 b 个样本，第 c 个通道，空间位置 (h,w) 处的特征图，符号⊙表示 Hadamard 积逐元素相乘，$\boldsymbol{F}_{att}$ 为最终得到的特征图。在注意力相关的损失函数中，我们引入可调节参数 λ，该参数用于平衡损失函数中不同部分的权重占比。ReID 的总损失可表示为：

$$L_{all} = L_{cls}^{(1,2)} + L_{triplet}^{(1,2)} + \lambda \left(L_1 + L_2 + L_3 \right) \tag{18}$$

3. 实验分析

3.1. 数据集及评价指标

为验证所提模型的有效性，我们所做实验全部在三个数据集，包括 MSMT17 [9]、Market-1501 [10]和 DukeMTMC-ReID [11]，数据集的详细信息如表 1。我们遵循 ReID 的标准评估协议，并将 Rank-1 准确率和查询样本的平均精度均值(mean Average Precision, mAP)作为评价指标。

Table 1. The details of datasets
表 1. 数据集详情

Datesets	Cameras	TrainIDs	Train	TestIDs	Query	Gallery
MSMT17	15	1041	30,602	1041	10,516	18,184
Market-1501	6	751	12,396	750	3368	19,732
DukeMTMC-ReID	8	702	16,522	702	2228	17,661

3.2. 实验设置

本研究的实验环境配置如下：采用 NVIDIARTX3090 显卡，操作系统为 Ubuntu20.04 所有模型均基于 Python3.8 编程语言与 PyTorch1.9 深度学习框架实现。在训练阶段，图像数据被裁剪至 384 × 128 像素大小，并采用数据增强策略。模型总计训练 50 轮，训练阶段的批量大小设为 64。为确保实验结果的可复现性，实验设置随机种子为 42，并重复进行 5 次实验。测试阶段的批量大小设为 128。

3.3. 与代表性方法的对比

为验证所提方法的有效性，我们将其与 ReID 领域的代表性方法进行对比实验，对所提方法与其他多

种对比方法的性能进行了评估，如表 2 所示。

本文提出的方法在三个 ReID 数据集上均取得了优异性能：在 MSMT17 数据集上，Rank-1 达 86.5%，mAP 达 70.6%；在 Market-1501 数据集上，Rank-1 和 mAP 分别达到 96.5%和 92.8%；在 DukeMTMC-ReID 数据集上，Rank-1 和 mAP 则为 92.4%和 88.1%。在采用 ResNet-50 作为骨干网络的方法中，尽管 NFormer 在 Market-1501 数据集上的 mAP 达到 93.0%，但我们的方法在其他所有指标上均实现了超越。

与轻量级网络(如 OSNet、CDNet)以及专为 ReID 任务设计的网络(如 DPFL)相比，我们的方法在所有评估指标上均展现出显著优势。

与 Transformer 相比，例如 PASS 在 MSMT17 数据集上的 Rank-1 与我们的方法相当，我们的方法在三个数据集上的综合性能仍处于最优水平。

总体而言，与其他代表性方法相比，本文提出的方法在多个数据集上均对评价指标有一定提升，充分证明了其在 ReID 任务中的有效性。

Table 2. Comparison with the representative methods on three datasets
表 2. 在三个数据集上与代表性方法的对比

	backbone	MSMT17		Market-1501		DukeMTMC-ReID	
		mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
DPFL [12]	DPFL	-	-	72.6	88.6	60.6	79.2
OSNet [13]	OSNet	52.9	78.7	84.9	94.8	73.5	88.6
CDNet [14]	CDNet	54.7	78.9	86.0	95.1	76.8	88.6
PASS [15]	Transformer	69.1	86.5	92.2	96.3	-	-
TransCL [16]	Transformer	41.3	68.6	82.9	93.0	-	-
Dip [17]	Transformer	67.5	84.6	90.3	95.7	82.8	91.2
IIDCL [18]	Transformer	55.0	80.4	89.9	95.4	-	-
TMGF [19]	Transformer	58.2	83.3	89.5	95.5	-	-
MLFN [20]	ResNet-50	-	-	74.3	90.0	62.8	81.0
PCB [21]	ResNet-50	40.4	68.2	77.4	92.3	78.4	88.7
AANet [22]	ResNet-50	-	-	82.5	93.9	72.6	86.4
CAL [23]	ResNet-50	56.2	79.5	87.0	94.5	76.4	87.2
APNet-S [24]	ResNet-50	59.0	80.8	89.0	96.1	78.8	89.3
NFormer [25]	ResNet-50	62.2	80.8	93.0	95.7	85.7	90.6
DCCT [26]	ResNet-50	41.8	68.7	86.3	94.4	-	-
LESL [27]	ResNet-50	42.6	68.2	83.4	92.9	72.7	83.4
MDL [28]	ResNet-50	40.9	69.1	89.1	95.1	-	-
ours	DSAM	**70.6**	**86.5**	**92.8**	**96.5**	**88.1**	**92.4**

3.4. 消融实验

为明确所提方法中 dual-branch、BSAM 模块、CBAM 模块和 FFM 模块对算法性能的影响，我们在 MSMT17、Market-1501 和 DukeMTMC-ReID 三个数据集上设计了消融实验。

3.4.1. Dual-Branch 的有效性验证

如表 3 中 Index1 所示，baseline 在 MSMT17 的 mAP 仅为 48.4%、Rank-1 为 77.2%；引入双分支 Index3 中，MSMT17 的 mAP 提升至 64.8%、Rank-1 提升至 83.1%，Market-1501 的 mAP 从 82.8%升至 91.4%、Rank-1 从 94.3%升至 95.7%。

在 Index2 中，仅引入 BSAM 时，性能提升极其有限：Market-1501 的 mAP 仅达 90.0%，DukeMTMC-ReID 的 Rank-1 仅 87.2%。这是因为缺失双分支后，模型仅依赖单一尺度输入，无法覆盖行人近景或远景的尺度差异，导致特征对尺度变化的鲁棒性不足，最终限制了后续模块的优化空间，因此后续我们在双分支结构上进行消融实验。

Table 3. The ablation studies of our proposed method on three datasets
表 3. 所提方法在三个数据集上的消融实验

实验	Dual-Branch	BSAM	CBAM	FFM	MSMT17		Market-1501		DukeMTMC-ReID	
					mAP	Rank-1	mAP	Rank-1	mAP	Rank-1
1	×	×	×	×	48.4	77.2	82.8	94.3	72.8	86.5
2	×	√	×	×	55.9	79.2	90.0	94.9	79.0	87.2
3	√	×	×	×	64.8	83.1	91.4	95.7	84.6	90.3
4	√	×	√	×	67.2	84.8	92.0	96.2	86.3	91.4
5	√	×	√	√	68.9	85.8	92.4	96.4	87.5	91.9
6	√	√	√	×	69.3	85.2	92.5	96.0	87.0	92.0
7	√	√	√	√	**70.6**	**86.5**	**92.8**	**96.5**	**88.1**	**92.4**

3.4.2. 注意力模块的有效性验证

实验中两种注意力模块的消融结果清晰展现了不同注意力机制的特征优化能力。Index2 中性能得到提升 Market-1501 的 mAP 与 Rank-1 提升 0.3%和 0.2%，DukeMTMC-ReID 对应提升 0.2%和 0.2%，这源于 BSAM 无参监督的轻量型设计，可提前抑制背景冗余信息保留行人主体的基础特征信号。Index4 实验性能实现显著跃升：MSMT17 的 mAP 与 Rank-1 提升到 67.2%和 84.8%，Market-1501 的 mAP 突破 92.0%，这是因为 CBAM 通过通道注意力与空间注意力的串行优化，能强化关键特征通道的响应，有效提升了特征的判别能力。在 Index6 中较 Index4，在复杂数据集 MSMT17 上的 mAP 从 67.2%提升至 69.3%，进一步说明 BSAM 的前置降噪让 CBAM 无需在冗余特征上消耗计算资源，可更聚焦于行人判别性区域的精准优化。

3.4.3. 特征融合的有效性验证

对比 Index4 与 Index5 的性能：MSMT17 的 mAP 从 69.3%提升 1.3%，Market-1501 的 mAP 从 92.5%升至 92.8%，DukeMTMC-ReID 的 mAP 与 Rank-1 分别提升 1.1%和 0.4%。这是因为 FFM 聚合了主干网

络的低层细节与高层语义特征，弥补了单一层级特征的信息短板，尤其在 MSMT17 这类多场景数据集上，跨层互补性大幅增强了特征对光照视角变化的鲁棒性。

3.4.4. 参数 λ 的最优值探索

为确定总损失函数中的 λ 贡献权重，我们测试了不同值对模型性能的影响，如图 8 所示，表明 λ = 0.01 是最优参数。这一结果说明：通过在注意力相关损失与其他损失项之间建立合理平衡，既能有效增强模型对人体关键区域的关注，又能提升类内特征一致性。当 λ = 0.5 时，模型性能出现显著退化，这是因为过大的值破坏了损失函数的平衡，导致模型以牺牲基础特征学习为代价，过度拟合注意力对齐目标。这一现象印证了：在深度神经网络中，超参数调优对平衡注意力机制与特征表示学习具有至关重要的作用。

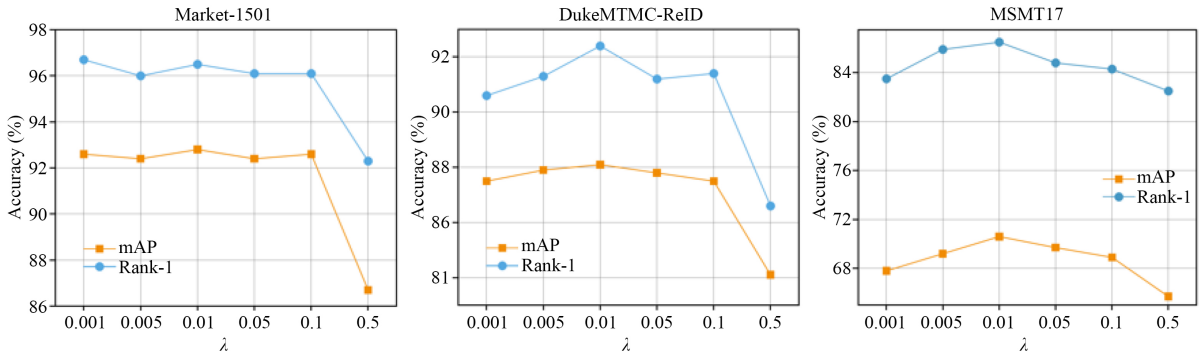

Figure 8. The results λ of on three datasets
图 8. 参数 λ 在三个数据集上的结果

3.5. 可视化分析

3.5.1. 热力图可视化

为验证所提方法的有效性，我们在 Market-1501 数据集上对网络不同阶段的特征与注意力分布进行可视化，具体结果如图 9 所示。

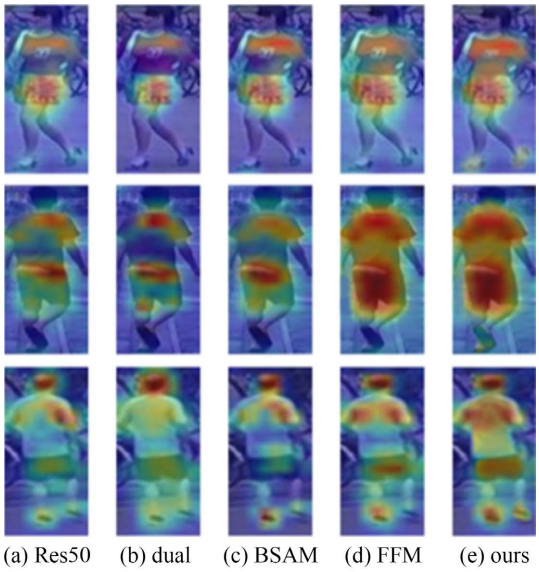

Figure 9. Visualization of feature maps based on different stages
图 9. 基于不同网络阶段的特征图可视化

与 ResNet-50 相比，本文提出的 BSAM 模块更关注行人身体的头部、躯干。通过双分支结构的设计，行人的关键特征区域不仅得到强化，覆盖范围也进一步扩大，最终使网络的注意力更集中于身份判别相关区域。图中，最后一列图像为所提模型的注意力热力图：对于第一位行人，热力图清晰捕捉到上衣细节；对于第二位行人，躯干信息更明确，裤子细节也更突出；对于第三位行人，身体轮廓信息更清晰，同时还能关注到鞋子细节。

3.5.2. 检索结果可视化

此外，我们分别采用 DSAM、ViT 和 ResNet-50 三种模型提取行人特征，再根据特征相似度得分筛选出前 10 个检索结果。其中，正确匹配的结果用绿色方框标记，错误匹配的结果用红色方框突出显示。

图 10 所示结果表明，DSAM 模型能够快速检测出所具有相同身份的目标；且对于查询样本而言，其匹配结果的质量显著优于 ResNet-50 模型与 ViT 模型。

Figure 10. Top 10 visualization results of the rank-list Market-1501
图 10. Market-1501 数据集排序列表的前 10 个可视化结果

4. 结论

本文提出一种创新性的双分支监督注意力模型(DSAM)。该模型通过融合双分支结构与监督注意力机制，显著提升了 ReID 任务的识别性能。实验结果表明，DSAM 在多个标准数据集上均取得优异性能，充分验证了其有效性与鲁棒性。文中详细介绍了模型各核心组件的作用，包括双分支结构、瓶颈无参监督注意力模块、特征融合模块及对应的损失函数。通过消融实验，我们验证了每个模块对模型性能的独立贡献与协同价值。总体而言，DSAM 为 ReID 领域的技术发展提供了新的思路与方法，为后续相关研究提供了可借鉴的框架。

参考文献

[1] Chen, T., Ding, S., Xie, J., Yuan, Y., Chen, W., Yang, Y., *et al*. (2019) ABD-Net: Attentive but Diverse Person Re-identification. 2019 *IEEE/CVF International Conference on Computer Vision* (*ICCV*), Seoul, 27 October-2 November 2019, 8351-8361. https://doi.org/10.1109/iccv.2019.00844

[2] Wang, G., Yuan, Y., Chen, X., Li, J. and Zhou, X. (2018) Learning Discriminative Features with Multiple Granularities for Person Re-Identification. *Proceedings of the 26th ACM International Conference on Multimedia*, Seoul, 22-26 October 2018, 274-282. https://doi.org/10.1145/3240508.3240552

[3] Bai, X., Yang, M., Huang, T., Dou, Z., Yu, R. and Xu, Y. (2020) Deep-Person: Learning Discriminative Deep Features for Person Re-Identification. *Pattern Recognition*, **98**, Article ID: 107036. https://doi.org/10.1016/j.patcog.2019.107036

[4] Vinyals, O., Toshev, A., Bengio, S. and Erhan, D. (2015) Show and Tell: A Neural Image Caption Generator. 2015 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Boston, 7-12 June 2015, 3156-3164. https://doi.org/10.1109/cvpr.2015.7298935

[5] Zhang, Z., Lan, C., Zeng, W., Jin, X. and Chen, Z. (2020) Relation-Aware Global Attention for Person Re-Identification. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 13-19 June 2020, 3186-3195. https://doi.org/10.1109/cvpr42600.2020.00325

[6] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. 2016 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Las Vegas, 27-30 June 2016, 770-778. https://doi.org/10.1109/cvpr.2016.90

[7] Woo, S., Park, J., Lee, J.Y. and Kweon, I.S. (2018) CBAM: Convolutional Block Attention Module. In: Ferrari, V., *et al*., Eds., *Computer Vision—ECCV 2018*, Springer International Publishing, 3-19. https://doi.org/10.1007/978-3-030-01234-2_1

[8] Tan, Z., Zhang, G., Tan, Z., Tiwari, P., Wang, Y. and Yang, Y. (2025) Cam2former: Fusion of Camera-Specific Class Activation Map Matters for Occluded Person Re-Identification. *Information Fusion*, **120**, Article ID: 103011. https://doi.org/10.1016/j.inffus.2025.103011

[9] Wei, L., Zhang, S., Gao, W. and Tian, Q. (2018) Person Transfer GAN to Bridge Domain Gap for Person Re-Identification. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 79-88. https://doi.org/10.1109/cvpr.2018.00016

[10] Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J. and Tian, Q. (2015) Scalable Person Re-Identification: A Benchmark. 2015 *IEEE International Conference on Computer Vision* (*ICCV*), Santiago, 7-13 December 2015, 1116-1124. https://doi.org/10.1109/iccv.2015.133

[11] Ristani, E., Solera, F., Zou, R., Cucchiara, R. and Tomasi, C. (2016) Performance Measures and a Data Set for Multi-Target, Multi-Camera Tracking. In: Hua, G. and Jégou, H., Eds., *Computer Vision—ECCV 2016 Workshops*, Springer International Publishing, 17-35. https://doi.org/10.1007/978-3-319-48881-3_2

[12] Chen, Y., Zhu, X. and Gong, S. (2017) Person Re-Identification by Deep Learning Multi-Scale Representations. 2017 *IEEE International Conference on Computer Vision Workshops* (*ICCVW*), Venice, 22-29 October 2017, 2590-2600. https://doi.org/10.1109/iccvw.2017.304

[13] Zhou, K., Yang, Y., Cavallaro, A. and Xiang, T. (2019) Omni-Scale Feature Learning for Person Re-Identification. 2019 *IEEE/CVF International Conference on Computer Vision* (*ICCV*), Seoul, 27 October-2 November 2019, 3702-3712. https://doi.org/10.1109/iccv.2019.00380

[14] Li, H., Wu, G. and Zheng, W. (2021) Combined Depth Space Based Architecture Search for Person Re-Identification. 2021 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Nashville, 20-25 June 2021, 6729-6738. https://doi.org/10.1109/cvpr46437.2021.00666

[15] Zhu, K., Guo, H., Yan, T., Zhu, Y., Wang, J. and Tang, M. (2022) PASS: Part-Aware Self-Supervised Pre-Training for Person Re-Identification. In: Avidan, S., *et al*., Eds., *Computer Vision—ECCV* 2022, Springer, 198-214. https://doi.org/10.1007/978-3-031-19781-9_12

[16] Tao, Y., Zhang, J., Chen, T., Wang, Y. and Zhu, Y. (2022) Transformer-Based Contrastive Learning for Unsupervised Person Re-Identification. 2022 *International Joint Conference on Neural Networks* (*IJCNN*), Padua, 18-23 July 2022, 1-9. https://doi.org/10.1109/ijcnn55064.2022.9892516

[17] Li, D., Chen, S., Zhong, Y. and Ma, L. (2023) Dip: Learning Discriminative Implicit Parts for Person Re-Identification.

[18] Xiong, M., Hu, K., Lyu, Z., Fang, F., Wang, Z., Hu, R., *et al*. (2024) Inter-Camera Identity Discrimination for Unsupervised Person Re-Identification. *ACM Transactions on Multimedia Computing, Communications, and Applications*, **20**, 1-18. https://doi.org/10.1145/3652858

[19] Li, J., Wang, M. and Gong, X. (2023) Transformer Based Multi-Grained Features for Unsupervised Person Re-Identification. 2023 *IEEE/CVF Winter Conference on Applications of Computer Vision Workshops* (*WACVW*), Waikoloa, 3-7 January 2023, 42-50. https://doi.org/10.1109/wacvw58289.2023.00009

[20] Chang, X., Hospedales, T.M. and Xiang, T. (2018) Multi-Level Factorisation Net for Person Re-Identification. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 2109-2117. https://doi.org/10.1109/cvpr.2018.00225

[21] Han, J., Yao, X., Cheng, G., Feng, X. and Xu, D. (2022) P-CNN: Part-Based Convolutional Neural Networks for Fine-Grained Visual Categorization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **44**, 579-590. https://doi.org/10.1109/tpami.2019.2933510

[22] Tay, C.-P., Roy, S. and Yap, K.-H. (2019) AANet: Attribute Attention Network for Person Re-Identifications. 2019

IEEE/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Long Beach, 15-20 June 2019, 7134-7143. https://doi.org/10.1109/cvpr.2019.00730

[23] Rao, Y., Chen, G., Lu, J. and Zhou, J. (2021) Counterfactual Attention Learning for Fine-Grained Visual Categorization and Re-Identification. 2021 *IEEE*/*CVF International Conference on Computer Vision* (*ICCV*), Montreal, 10-17 October 2021, 1025-1034. https://doi.org/10.1109/iccv48922.2021.00106

[24] Chen, G., Gu, T., Lu, J., Bao, J.-A. and Zhou, J. (2022) Person Re-Identification via Attention Pyramid. *IEEE Transactions on Image Processing*, **31**, 7167-7179.

[25] Wang, H., Shen, J., Liu, Y., Gao, Y. and Gavves, E. (2022) NFormer: Robust Person Re-Identification with Neighbor Transformer. 2022 *IEEE*/*CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), New Orleans, 18-24 June 2022, 7297-7307. https://doi.org/10.1109/cvpr52688.2022.00715

[26] Chen, Z., Cui, Z., Zhang, C., Zhou, J. and Liu, Y. (2023) Dual Clustering Co-Teaching with Consistent Sample Mining for Unsupervised Person Re-Identification. *IEEE Transactions on Circuits and Systems for Video Technology*, **33**, 5908-5920. https://doi.org/10.1109/tcsvt.2023.3261898

[27] Bertocco, G.C., Theophilo, A., Andaló, F. and De Rezende Rocha, A. (2023) Leveraging Ensembles and Self-Supervised Learning for Fully-Unsupervised Person Re-Identification and Text Authorship Attribution. *IEEE Transactions on Information Forensics and Security*, **18**, 3876-3890. https://doi.org/10.1109/tifs.2023.3289448

[28] Ji, F., Zhang, B., Chao, L., Guo, H. and Li, J. (2024) MDL: Multi-Granularity Distribution Features Learning for Unsupervised Person Re-Identification. 2024 *IEEE* 30*th International Conference on Parallel and Distributed Systems* (*ICPADS*), Belgrade, 10-14 October 2024, 512-519. https://doi.org/10.1109/icpads63350.2024.00073