E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP391.41　　**文献标识码**:A　　**文章编号**:1006-8961(2024)08-2364-13

论文引用格式: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(08):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(08):2364-2376)[DOI:10.11834/jig.230523]

融合姿态引导和多尺度特征的遮挡行人重识别

张红颖[1*],刘腾飞[1],罗谦[2],张涛[2]

1. 中国民航大学电子信息与自动化学院,天津　300300; 2. 民航成都电子技术有限责任公司,成都　610041

摘　要:**目的**　在行人重识别任务中,行人外观特征会因为遮挡发生变化,从而降低行人特征的辨别性,仅基于可视部分的传统方法仍会识别错误。针对此问题,提出了一种融合姿态引导和多尺度特征的遮挡行人重识别方法。**方法**　首先,构建了一种特征修复模块,根据遮挡部位邻近信息恢复特征空间中被遮挡区域的语义信息,实现缺失部位特征的修补。然后,为了从修复的图像中提取有效的姿态信息,设计了一种姿态引导模块,通过姿态估计引导特征提取,实现更加精准的行人匹配。最后,搭建了特征增强模块,并融合显著性区域检测方法增强有效的身体部位特征,同时消除背景信息造成的干扰。**结果**　在3个公开的数据集上进行了对比实验和消融实验,在Market1501、DukeMTMC-reID(Duke multi-tracking multi-camera re-identification)和Occluded-DukeMTMC(occluded Duke multi-tracking multi-camera re-identification)数据集上的平均精度均值(mean average precision, mAP)和首次命中率(rank-1 accuracy, Rank-1)分别为88.8%和95.5%、79.2%和89.3%、51.7%和60.3%。对比实验结果表明提出的融合算法提高了行人匹配的准确率,具有较好的竞争优势。**结论**　本文所提的姿态引导和多尺度融合方法,修复了因遮挡而缺失的部位特征,结合姿态信息融合了不同粒度的图像特征,提高了模型的识别准确率,能有效缓解遮挡导致的误识别现象,验证了方法的有效性。

关键词:行人重识别(ReID);遮挡;姿态引导;特征融合;特征修补

Pose guidance and multi-scale feature fusion for occluded person re-identification

Zhang Hongying[1*], Liu Tengfei[1], Luo Qian[2], Zhang Tao[2]

1. College of Electronic Information and Automation, Civil Aviation University of China, Tianjin　300300, China;
2. Civil Aviation Electronic Technology Co., Ltd., Chengdu　610041, China

Abstract: **Objective** Person re-identification (ReID) is an important task in computer vision, and it aims to accurately identify and associate the same person between multiple visual surveillance cameras by extracting and matching features of pedestrian under different scenarios. Occluded person ReID is a challenging and specialized task in the existing person ReID problems. In real-world settings, occlusion is a common issue, and it impacts the practical application of person ReID technique to a certain extent. Recently, occluded person ReID has gradually attracted the attention of many researchers, and several methods have been proposed to address the issue of occlusion, which achieve impressive results. Currently, these methods primarily focus on the visible regions in images. Concretely, it first locates the visible regions in the

收稿日期:2023-08-11;**修回日期**:2023-10-24;**预印本日期**:2023-11-01

*通信作者:张红颖　carole_zhang0716@163.com

基金项目:国家自然科学基金民航联合研究基金重点支持项目(U2133211)

Supported by: National Natural Science Foundation of China Civil Aviation Joint Research Fund(U2133211)

image and then specially designs a model to extract discerning feature information from these regions, which achieves accurate person matching. These methods typically remove features coming from the occluded areas and then exploit discriminative features from the non-occluded regions for matching. Although these methods achieve impressive results, the influence of occluded regions and background interference in images are ignored, which results in the aforementioned solutions failing to effectively address the misclassification issue resulting from similar appearances in non-occluded regions. Consequently, merely relying on visible regions for subsequent recognition task leads to a sharp performance drop of the model, and the interference coming from image backgrounds also affects the further improvement in recognition accuracy. Some methods have been proposed to recover the occluded regions in images for overcoming the abovementioned issues. Specifically, these methods restore the occluded parts by utilizing the unobstructed image information at the image level. However, the restoration approaches may cause image distortion and introduce an excessive number of parameters. **Method** We propose a person ReID method based on pose guidance and multi-scale feature fusion to alleviate the aforementioned issues. This method can enhance the feature representation capability of the model and obtain more discriminative features. First, a feature restoration module is constructed to restore the occluded image features at the feature level while effectively reducing the parameters of the model. The module uses spatial contextual information from the non-occluded regions to predict the features of adjacent occluded regions, which restores the semantic information of the occluded regions in the feature space. The feature restoration module mainly consists of two subparts: the adaptive region division unit and the feature restoration one. The adaptive region division unit divides the image into six regions adaptively according to the predicted localization points to facilitate the clustering of similar feature information in different regions. The adaptive division in the module could effectively alleviate the misalignment caused by fixed division methods, and it could achieve more accurate position alignment. The feature restoration unit comprises of an encoder and a decoder. The encoder encodes the feature information coming from the divided regions of the image with similar appearances or close positions into a cluster. Meanwhile, the decoder assigns the cluster information to the occluded body parts in the image, which completes the feature restoration of missing body parts. Second, a pose estimation network is employed to extract pedestrian pose information. The pose estimation network is responsible for guiding the generation of keypoint heatmaps for the restored complete image features. Then, it implements the prediction of body keypoints with the heatmaps to obtain pose information. The pretrained pose estimation guidance model performs fusion learning on the global non-occluded regions and the restored regions to obtain more distinctive pedestrian feature information for more accurate pedestrian matching. Finally, a feature enhancement module is proposed to extract salient features from the image for eliminating the interference coming from background information while enhancing the learning capability for effective information. This module not only makes the network pay close attention to the valid semantic information in the feature maps but also reduces the interference coming from background noises, which could effectively alleviate the failure of feature learning caused by occlusion. **Result** We conducted several comparative experiments and ablation experiments on three publicly available datasets to validate the effectiveness of our method. We employed mean average precision (mAP) and Rank-1 accuracy as our evaluation metrics. Experiment results demonstrate that our method achieves mAP and Rank-1 of 88.8% and 95.5% on the Market1501 dataset, respectively. The mAP and Rank-1 are 79.2% and 89.3%, respectively, on the Duke multi-tracking multi-camera ReID (DukeMTMC-reID) dataset. On the occluded Duke multi-tracking multi-camera re-recognition (Occluded-DukeMTMC) dataset, the mAP and Rank-1 can reach 51.7% and 60.3%, respectively. Moreover, our method outperforms the PGMA-Net by 0.4% in mAP on the Market1501 dataset, by 0.8% in mAP and 0.7% in Rank-1 on the DukeMTMC-reID dataset, and by 1.2% in mAP on the Occluded-DukeMTMC dataset. At the same time, the ablation experiments confirm the effectiveness of the three proposed modules. **Conclusion** Our proposed method, pose-guided and multi-scale feature fusion (PGMF), could effectively recover the features of missing body parts, alleviate the issue of background interference, and achieve accurate pedestrian matching. Therefore, the proposed model effectively alleviates the misidentification caused by occlusion, improves the accuracy of person ReID, and exhibits robustness.

Key words: person re-identification (ReID); occlusion; pose guidance; feature fusion; feature restoration

0 引 言

行人重识别(person re-identification, ReID)是计算机视觉领域的一项重要研究,旨在在不同的监控视频中识别和跟踪同一个行人,在视频监控、智能交通、安防等领域发挥着基础性的作用(李擎 等, 2022; Peng等, 2022; Zahra等, 2023)。近年来,行人重识别问题得到了研究人员的高度关注,而遮挡行人重识别是现有行人重识别问题的一种特殊和挑战场景(张永飞 等, 2023)。因此,解决遮挡问题成为行人重识别领域的研究重点之一。

为了解决行人重识别中的遮挡问题,姿态估计、注意力机制和图像修复等方法被应用于遮挡行人重识别。Miao等人(2019)提出了一种姿态引导特征对齐方法,利用姿态信息从遮挡图像中获取具有判别性的信息。该方法舍弃了对图像中被遮挡部分的处理,仅考虑未遮挡的关键点特征之间的距离,从而达到特征对齐的目的。Gao等人(2020)提出了一种基于位姿引导的可视部分匹配方法,该方法在不考虑遮挡部位的情况下,联合学习带有位姿引导注意力的判别特征,在端到端的框架下挖掘可见部位特征。Yang等人(2022)提出了一种姿态驱动的注意力融合机制方法,通过结合人体姿态信息和图像中的空间语义信息,将注意力机制感兴趣的区域与未遮挡的空间语义信息进行匹配和融合,以消除遮挡对人员再识别的影响。然而这些方法的关注点仅限于图像中未被遮挡的区域,并没有对被遮挡部分的特征进行恢复,这限制了特征辨别能力的提升。由于没有考虑到遮挡区域的特征信息,这些方法无法有效提高图像特征的区分度。Tian等人(2023)提出了一种注意力机制方法来解决遮挡问题,该方法通过提取图像的局部特征,并使用多头注意力机制来对全局特征进行建模,不同注意力头之间进行信息交互,从而提高模型的保证能力。但是注意力机制建模全局特征不仅无法修复缺失的遮挡部位特征,还容易引入噪声干扰。Wang等人(2022)提出了一种差分注意力孪生网络,通过将遮挡特征从原始图像中减除得到非遮挡区域的差异特征,并集成注意力机制,以减轻对非遮挡区域特征的不准确探索。但是通过去除遮挡区域的特征,挖掘非遮挡区域的判别特征实现匹配的方法并没有充分解决当存在外观相似性的非遮挡区域时可能出现的混淆问题。Hou等人(2019)通过使用卷积神经网络直接对遮挡图像进行修复,但是这种方法会修改原始图像的内容,导致图像信息丢失或失真,会对原始图像造成不可逆的改变,同时也增加了模型的复杂性和参数的数量。因此,仅使用可见部位提取特征的方法,无法有效地解决当非遮挡区域存在相似外观时可能导致的判别错误问题,而直接对图像进行修复不仅会造成图像的失真还会增加模型的参数量。

针对以上问题,本文提出了一种融合姿态引导和多尺度特征的方法,通过修复遮挡部位的特征同时学习行人的姿态信息,以提高网络模型的特征表达能力,使其获得更加具有判别性的特征。

本文主要贡献如下:1)针对因遮挡物导致的特征缺失问题,提出了一种特征修补模块,该模块使用非遮挡区域的空间上下文来预测相邻遮挡区域的特征,实现缺失特征的补全;2)在姿态估计引导模块中,预训练的姿态引导模型对全局的非遮挡区域和修补区域的特征进行融合学习,实现更加精确的行人特征匹配;3)提出了一种特征增强模块,通过显著性区域检测方法自适应地提取图像中具有区分性的显著性特征,并降低背景噪声带来的干扰,有效缓解因遮挡导致的特征学习失败问题。

1 融合姿态引导和多尺度特征网络

为了缓解因遮挡导致的误识别问题,本文提出了融合姿态引导和多尺度特征的行人重识别网络(pose guidance and multi-scale features fusion network, PGMF),整体结构如图1所示。

PGMF网络模型采用ResNet50(residual network 50)作为骨干网,共包含空间特征修补模块(spatial feature patching module, SFPM)、姿态估计模块(pose estimation module, PEM)和特征增强模块(feature enhancement block, FEB)3个模块。首先,考虑到现实场景中人体下半部分被遮挡的频率最高,对人体下半部分进行了更为细致的划分,为了降低网络模型的复杂度,减少参数量的同时尽可能学习更多的特征,空间特征修补模块中的动态调整分区单元将图像分割成6个子区域,并生成掩膜进行位置对齐。特征修复单元对子区域特征进行处理,编码器将6个子区域映射为3个聚类,解码器将聚类信息

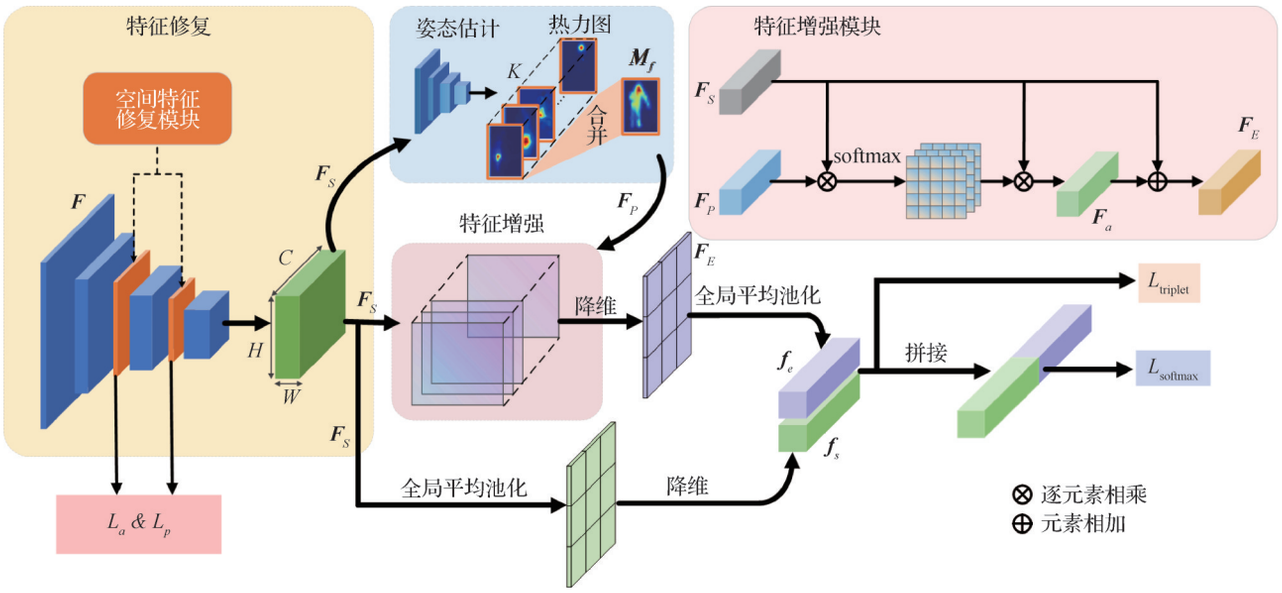

图 1　融合姿态引导和多尺度特征的网络结构

Fig. 1　The structure of pose guidance and multi-scale features fusion network

传播到被遮挡的区域，实现对遮挡区域特征的补全。然后，姿态引导模块使用修补后的特征图生成关键点热力图，将它们合并成一个包含完整人体关键点信息的热力图，并与原特征图逐元素相乘，获得包含完整人体姿态信息的特征图。最后，特征增强模块通过显著性区域检测方法增强来自姿态估计模块的特征，提高网络的特征提取能力。

1.1　空间特征修复模块

1.1.1　动态调整分区单元

为了使用图像中相邻部位的信息修复被遮挡部位的特征，同时抑制无关信息的干扰，需要将图像细化分为不同的区域。采用固定分区的方式划分图像，分析每个区域是否包含遮挡的方法容易造成空间位置错位(Zheng 等，2019a)。所以，本文提出一种动态调整分区单元，通过在特征图 $\boldsymbol{F}$ 上预测定位点的位置，将图像分成 6 个区域，细化行人身体部位，同时为每个区域生成相应的掩码，实现在遮挡情况下的位置对齐，如图 2 所示。

首先，使用一个 1 × 1 卷积降低特征图的通道维度，然后在垂直方向上分别以肩部、臀部和膝盖作为图像划分标签，使用姿态估计模型(Liang 等，2019)获得人体关键点位置信息。利用获得的位置信息，沿水平方向对特征图执行平均池化操作，获取按列划分的描述符。随后，将经过展平处理的列描述符输入到线性层中，使用 softmax 函数生成相应的概率分布 p^i，最后，取 p^i 中最大值的索引得到定位点 $a^i(i = 1, 2, 3)$。相似地，沿垂直方向对特征图进行平均池化操作以获得行描述符，对行描述符采用上述方法得到概率分布 p^4，进而得到 a^4。将获得的 4 个

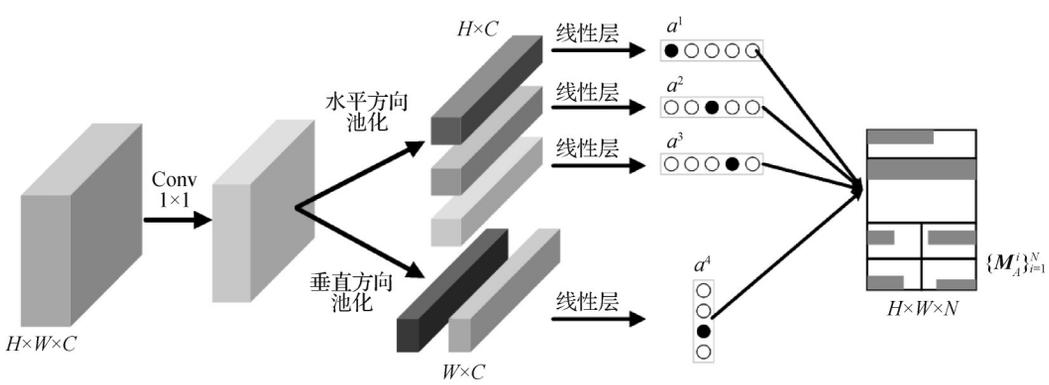

图 2　动态调整分区单元

Fig. 2　Dynamic adjustment of partition unit

最大索引值作为划分图像时的定位点$(a^1, 0)$,$(a^2, 0)$,$(a^3, 0)$,$(0, a^4)$,根据获得的定位点将特征图分成6个区域$\{\boldsymbol{A}^i\}_{i=1}^N$。具体为

$$P_{(h,w)} \in \begin{cases} \boldsymbol{A}^1 & h \leqslant a^1 \\ \boldsymbol{A}^2 & a^1 < h \leqslant a^2 \\ \boldsymbol{A}^3 & a^2 < h \leqslant a^3, w \leqslant a^4 \\ \boldsymbol{A}^4 & a^2 < h \leqslant a^3, w > a^4 \\ \boldsymbol{A}^5 & h > a^3, w \leqslant a^4 \\ \boldsymbol{A}^6 & h > a^3, w > a^4 \end{cases} \tag{1}$$

式中,(h,w)表示每个像素$P_{(h,w)}$在特征图上的空间位置。

通过将每个区域$\boldsymbol{A}^i$中的像素值设置为1,获得每个区域的掩码$\boldsymbol{M}_A^i$用于空间位置对齐。最后,对特征图$\boldsymbol{F}$使用softmax函数得到前景概率图$\boldsymbol{q}$,利用掩码$\boldsymbol{M}_A^i$和$\boldsymbol{q}$为每个划分的区域生成具有区分性的区域特征$\boldsymbol{f}^i$,以便修复被遮挡区域的特征。

$$\boldsymbol{f}^i = \frac{\boldsymbol{M}^i \otimes \boldsymbol{q}}{\left\| \boldsymbol{M}^i \otimes \boldsymbol{q} \right\|_1} \times \boldsymbol{F} \tag{2}$$

式中,$\boldsymbol{f}^i$表示第i个区域的特征向量;$\otimes$表示按元素相乘;$\|\cdot\|_1$表示L_1范数。

1.1.2 特征修复单元

为了避免在图像修复过程中造成图像失真,同时减少模型的参数量,针对图像空间特征进行修复的方法能很好地解决以上问题(Hou等,2022)。空间特征修复模块利用获得的区域特征信息$\boldsymbol{f}^i$来修补被遮挡区域的空间特征,该模块由一个区域编码器和一个区域解码器组成。区域编码器能够将具有相似外观或相邻位置的不同区域信息分配到同一簇中,实现对相关联区域特征的聚类;区域解码器通过解码聚合的特征信息修补图像中缺失的部位特征。该方法结合外观相似性和位置相邻性信息,充分利用了不同模态的信息,其中外观相似性反映了数据点之间的内容相似性,而位置相邻性则考虑了空间位置关系,结合两者的优势可以提高聚类的准确性,如图3所示。

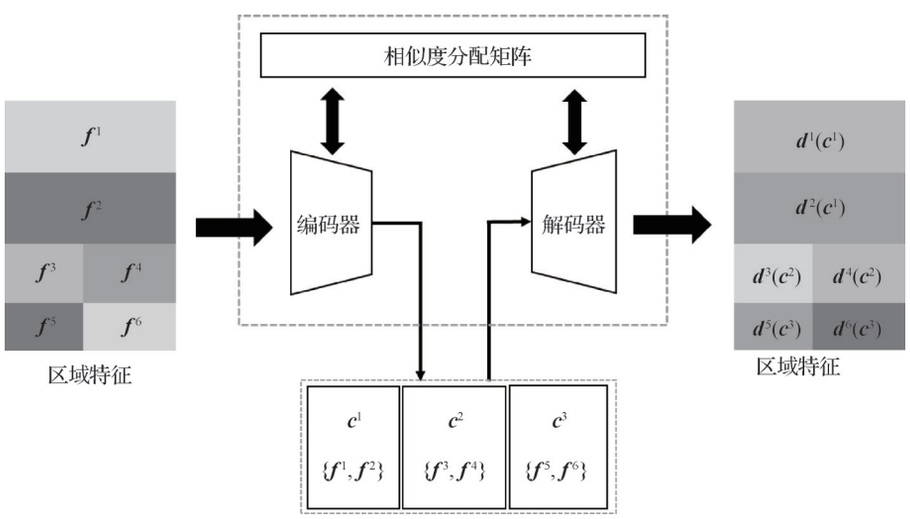

图3 空间特征修补模块图

Fig. 3 Spatial feature patching module diagram

在编码阶段,相似度分配矩阵$\boldsymbol{S}$将不同区域具有相似外观或相邻位置的区域特征$\{\boldsymbol{f}^i\}_{i=1}^N$聚合为一组聚类($k < 6$),如图4所示。

$$\boldsymbol{c}^k = \boldsymbol{S}_{ik}\{\boldsymbol{A}^i\} \tag{3}$$

式中,$\boldsymbol{S}_{ik}$表示将区域特征$\boldsymbol{f}^i$分配给聚类$\boldsymbol{c}^k$的概率。

相似度分配矩阵$\boldsymbol{S}$由外观相似矩阵$\boldsymbol{S}^A$和位置相邻矩阵$\boldsymbol{S}^P$组成。具体为

$$\boldsymbol{S} = \frac{\boldsymbol{S}^A + \boldsymbol{S}^P}{2} \tag{4}$$

受图聚类算法中相似矩阵的启发,构建了相似度分配矩阵$\boldsymbol{S}$,将其第k列的值进行归一化后作为权重对区域特征$\boldsymbol{f}$进行加权求和得到每个聚类。该方法可以实现自适应的特征加权,使得相似的区域被赋予更高的权重,而不相似的区域将受到较低的影响,有助于提高模型对数据内在结构上的感知能力。具体来说,首先将每个区域的外观特征$\boldsymbol{f}$和位置编码$\boldsymbol{L}$作为输入,通过卷积操作得到相似度分值矩阵。随后对该矩阵使用softmax函数得到相似度分配矩

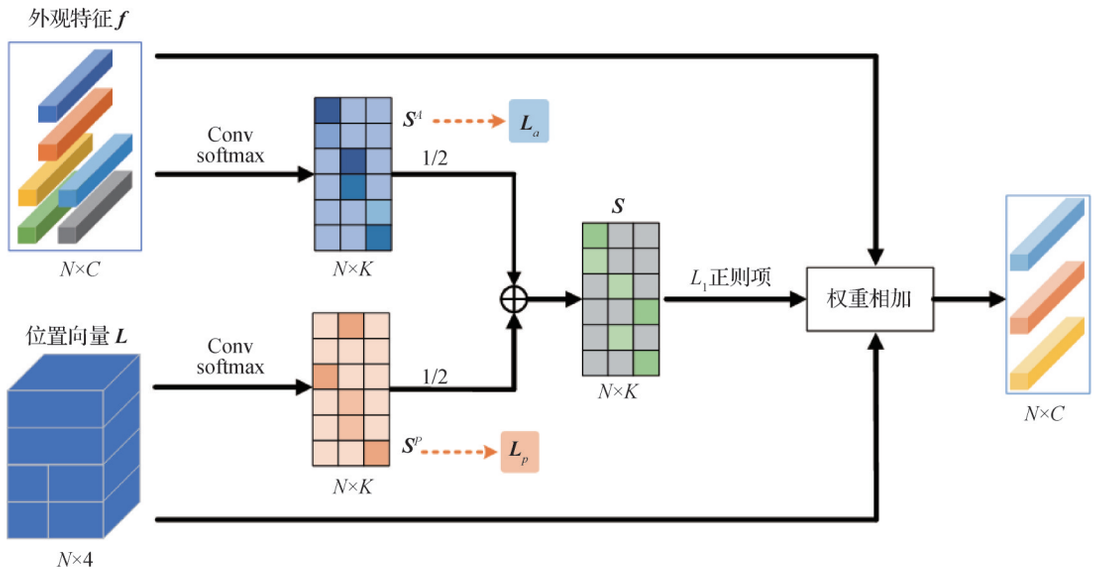

图 4　分配矩阵结构图

Fig. 4　Assignment matrix structure diagram

阵 $\boldsymbol{S}$，表示每个区域被分配给不同聚类的概率，最后使用相似度分配矩阵 $\boldsymbol{S}$ 对区域特征进行聚类。

外观相似矩阵 $\boldsymbol{S}^A$ 是通过卷积操作提取区域特征 $\boldsymbol{f}$，后经 softmax 函数得到。在聚类过程中，为了确保外观相似区域能够被精准地分配到同一聚类中，需要引入约束条件，以限定具有相似外观特征的区域。通过这种方式，可以更准确地识别和分类外观相似的区域。本文引入了外观相似度正则项 L_a，约束外观分配向量与对应区域特征的相似性一致。具体为

$$L_a = \sum_{i=1}^{N}\sum_{j=1}^{N}\left\| sim\left(S_i^A, S_j^A\right) - sim\left(\boldsymbol{f}^i, \boldsymbol{f}^j\right)\right\|_1 \tag{5}$$

$$sim\left(\boldsymbol{x}, \boldsymbol{y}\right) = \frac{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}}{\left\|\boldsymbol{x}\right\|_2\left\|\boldsymbol{y}\right\|_2} \tag{6}$$

式中，$\|\cdot\|_1$ 表示 L_1 范数；sim 表示余弦相似度；$\|\cdot\|_2$ 表示 L_2 范数；S_i^A 表示相对应于 $\boldsymbol{A}^i$ 区域特征的外观分配向量；S_j^A 表示相对应于 $\boldsymbol{A}^j$ 区域特征的外观分配向量；$\boldsymbol{f}$ 表示区域特征。

在 L_a 约束的情况下，当区域 $\boldsymbol{A}^i$ 的某些区域被部分遮挡，而该区域未被遮挡的部分与 $\boldsymbol{A}^j$ 在外观上具有相似性时，$\boldsymbol{S}^A$ 可以将 $\boldsymbol{A}^i$ 和 $\boldsymbol{A}^j$ 归并到同一个聚类中。这样，来自 $\boldsymbol{A}^j$ 的外观特征信息就能通过解码器传递到 $\boldsymbol{A}^i$，从而补全 $\boldsymbol{A}^i$ 中被遮挡的部分。然而，当人体某个区域被完全遮挡时，仅依赖相似外观信息进行补全的方式会受到限制。因此，需要根据相邻的位置信息补全被全部遮挡的区域。位置相邻矩阵 $\boldsymbol{S}^P$ 能够利用相邻区域的位置信息补全被完全遮挡的区域特征。首先，使用向量 $(\boldsymbol{x}_i, \boldsymbol{y}_i, h_i, w_i)$ 编码区域 $\boldsymbol{A}^i$ 的位置信息，$(\boldsymbol{x}_i, \boldsymbol{y}_i)$ 表示区域 $\boldsymbol{A}^i$ 的中心坐标；h_i, w_i 分别表示区域 $\boldsymbol{A}^i$ 的高度和宽度。将划分的 6 个区域的位置编码定义为 $\boldsymbol{L}$，然后通过卷积操作提取位置编码特征，经过 softmax 函数得到位置相邻矩阵 $\boldsymbol{S}^P$。

同样，为了保证将相邻位置的区域特征信息更准确地分配到一个聚类中，使用位置相邻正则项 L_p 对其进行约束。具体为

$$p_{\mathrm{sim}}\left(\boldsymbol{A}^i, \boldsymbol{A}^j\right) = 1 - 2\sqrt{\left(\frac{\boldsymbol{y}_i - \boldsymbol{y}_j}{H}\right)^2 + \left(\frac{\boldsymbol{x}_i - \boldsymbol{x}_j}{W}\right)^2} \tag{7}$$

$$L_p = \sum_{i=1}^{N}\sum_{j=1}^{N}\left\| sim(S_i^P, S_j^P) - p_{\mathrm{sim}}(\boldsymbol{A}^i, \boldsymbol{A}^j)\right\|_1 \tag{8}$$

式中，$\|\cdot\|_1$ 表示 L_1 正则化；p_{sim} 表示区域间位置相似度；H, W 分别表示特征图的高度和宽度；S_i^P 表示相对应于 $\boldsymbol{A}^i$ 区域特征的位置分配向量；S_j^P 表示相对应于 $\boldsymbol{A}^j$ 区域特征的位置分配向量。

在 L_p 约束下，区域编码器为位置相邻的区域产生一致性的位置向量，以保障相邻区域在编码过程中产生相似的表示结果。

在解码阶段，为了保证编码和解码过程信息的一致性，通过转置分配矩阵 $\boldsymbol{S}$ 得到解码矩阵 $\boldsymbol{D}$。应用该矩阵，将聚类 $\boldsymbol{c}^k$ 的解码输出映射到相应的区域从而得到特征 $\boldsymbol{d}^i$，以实现对被遮挡区域的特征重建。

$$\boldsymbol{d}^i = \sum_{k=1}^{K} \boldsymbol{D}_{ki} \times \boldsymbol{c}^k \tag{9}$$

由于聚类$\boldsymbol{c}^k$集合了相关非遮挡区域的特征，因此遮挡区域可以通过区域解码器解码相关部位的信息来恢复其特征。被遮挡图像的特征经过编码器—解码器的共同作用，最终得到被修复的图像特征$\boldsymbol{F}_S$。

1.2 姿态估计模块

姿态估计网络因其能提供行人的空间结构和姿态信息而常用于行人重识别。本文使用姿态估计模块（Sun等，2019）引导生成被修图像特征的关键点热力图，预测人体关键点，实现更精准的行人匹配。首先，姿态估计模块$\boldsymbol{H}$根据输入的图像特征$\boldsymbol{F}_S$生成K幅热力图，每幅热力图预测一个人体关键点。然后，使用归一化的方法使每个像素取值范围为（0，1），表示相应的置信度分数。由于每个初始热力图只覆盖了有限区域，因此使用全局最大池化方法扩展关键点的覆盖范围，获取更多的上下文信息。为了能更加准确地预测关键点的位置，同时降低背景噪声带来的干扰，通过设置阈值γ过滤置信度值过小的特征点。

$$L_i = \begin{cases} (x_i, y_i) & C_i \geqslant \gamma \\ 0 & \text{其他} \end{cases}, \quad i = 1, \cdots, K \tag{10}$$

式中，L_i表示第i个特征点；(x_i, y_i)表示第i个特征点的坐标；C_i表示置信度值；γ表示设置的阈值。

最后，取所有热力图中同一像素位置的最大置信度值C作为基准，合并全部的关键点热力图$\{\boldsymbol{M}_1, \cdots, \boldsymbol{M}_k\}$构建一个人体掩码$\boldsymbol{M}_f \in (0, 1)^{H_f \times W_f}$。

将人体掩码$\boldsymbol{M}_f$和特征图$\boldsymbol{F}_S$组合在一起，得到包含姿态信息的特征图$\boldsymbol{F}_P$，使网络模型学习行人的姿态特征信息。

1.3 特征增强模块

特征修复模块利用聚类方法对被遮挡的特征进行补全，但在聚类过程中不可避免地将图像背景信息包含在内。此外，姿态估计模块在生成姿态关键点时，会使用图像特征的上下文和全局信息，也会包含图像的背景信息。为消除背景信息带来的干扰，提高模型对有效信息的学习能力，本文提出了一种特征增强模块。该模块通过选择更具显著性的特征区域，以减少来自背景信息的干扰，其作用在于让特征学习集中在更有意义的区域。特征增强模块用于提取特征$\boldsymbol{F}_P$中显著性的行人特征，使网络学习行人的全局姿态信息和经过修复后的遮挡区域特征，得到特征增强分值图$\boldsymbol{a}_{i,j}$作为特征增强权重，使网络模型能够关注特征图中有效的语义信息，提高模型的性能，并使其更具自适应性，如图5所示。

首先，计算特征$\boldsymbol{F}_S$和特征图$\boldsymbol{F}_P$之间的余弦相似度，然后使用softmax函数将其归一化，得到特征增强分值图$\boldsymbol{a}_{i,j}$，使网络模型根据输入数据的不同特性，自适应地调整特征权重，从而适应不同的场景和任务。

$$\boldsymbol{F}_a = \sum_{i,j} \boldsymbol{a}_{i,j} \times \boldsymbol{F}_{i,j} \tag{11}$$

式中，$\boldsymbol{F}_{i,j}$表示特征图在(x, y)处的特征向量。

为了使模型学到更加丰富的特征信息，将修补后的特征$\boldsymbol{F}_S$和特征$\boldsymbol{F}_a$相加以保留更多的特征修复信息，最终使特征增强模块获得区域显著性特征$\boldsymbol{F}_E$。之后经过降低维度和全局平均池化得到特征$\boldsymbol{f}_e$，将其与特征$\boldsymbol{f}_s$一同用于网络模型的训练。

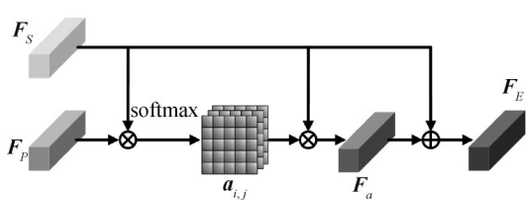

图5 特征增强模块结构图

Fig. 5 Structure diagram of feature enhancement module

1.4 损失函数

本文采用基于标签平滑的交叉熵损失（Szegedy等，2016）和难样本三元组损失（Schroff等，2015）联合监督网络。基于标签平滑的交叉熵损失函数能够解决行人重识别任务的数据集中存在噪声和错误标注等问题，比传统的交叉熵损失函数更加有效，且对噪声标签更加鲁棒，能有效提高网络模型的泛化能力。具体为

$$L_{\mathrm{TCE}} = -\sum_{i=1}^{N} q_i \log_2(P_i) \tag{12}$$

$$q_i = \begin{cases} \dfrac{\varepsilon}{N} & n \neq i \\ 1 - \dfrac{N-1}{N}\varepsilon & n = i \end{cases} \tag{13}$$

式中，N表示训练样本的行人类别数量；q_i表示经过平滑后的第i个类别的概率；n表示行人标签；τ表示平滑因子；k表示类别总数；P_i表示预测行人身份与

标签 i 一致的概率；δ_{y_i} 表示指示函数，其中，当 $i = y$ 时为 1，否则为 0，y 表示真实类别标签；ε 表示错误率。

三元组损失锚定图像和正样本图像的特征向量在特征空间中的距离尽可能小，而锚定图像和负样本图像的特征向量在特征空间中的距离尽可能大。为了使网络具有更好的泛化能力，本文使用了难样本三元组损失，它是在三元组损失函数的基础上改进的版本，具体为

$$L_{\mathrm{Th}} = \frac{\sum_{a \in batch}\left[\max d_{ap}(\eta) - \min_{n \in B} d_{an}(\eta) + \alpha\right]_+}{P \times K} \tag{14}$$

式中，a，p 和 n 分别为锚定样本、正样本和负样本，a 和 p 共享相同的行人类别，而 a 和 n 具有不同的行人类别；$d_{ap}(\eta)$ 和 $d_{an}(\eta)$ 分别表示目标样本与正、负样本之间的欧氏距离；P 和 K 分别为单个批次包含的行人数和每个行人相应的图像数；α 为间隔阈值参数，本文实验中，$\alpha = 0.5$。

动态划分单元中的定位点位置确定被转化为分类任务，使用交叉熵损失，即

$$L_k = \frac{1}{4T}\sum_{t=1}^{T}\sum_{i=1}^{4} L_{\mathrm{CE}}(p^i, \boldsymbol{l}^i) \tag{15}$$

式中，p^i 表示预测的定位点的概率；$\boldsymbol{l}^i$ 表示姿态估计其产生的 one-hot 向量；T 表示特征图帧数；L_{CE} 表示交叉熵损失。

网络模型总损失为

$$L_{\mathrm{total}} = L_{\mathrm{htri}} + L_{\mathrm{TCE}} + \lambda_1 L_k + \lambda_2(L_a + L_p) \tag{16}$$

式中，L_a 为外观相似度正则项；L_p 为位置相邻正则项；λ_1 和 λ_2 是超参数，用来平衡不同损失函数的影响。

2 实验结果与分析

2.1 数据集和评价指标

Market1501 数据集（Zheng 等，2015）由 6 个不同相机捕捉到的 1 501 位行人的图像组成，包含 1 501 个行人的 32 668 幅图像。其中训练集由 751 个不同行人身份的 12 936 幅图像组成；测试集包含 750 个行人身份，包含 19 732 幅图像。

DukeMTMC-reID（Duke multi-tracking multi-cam era re-identification）数据集（Ristani 等，2016）由 8 台高分辨率相机拍摄的 1 812 位行人的 36 411 幅图像组成。其中，训练集由属于 702 位行人的 16 522 幅图像构成，测试集包含 17 661 幅 702 个行人的图像。

Occluded-DukeMTMC（occluded Duke multi-tracking multi-camera re-identification）是一个专注于遮挡场景的数据集（Miao 等，2019）。它是通过在 Duke-MTMC-reID 数据集中留下遮挡图像作为查询，过滤掉训练集中具有相同遮挡模式的图像，遮挡数据集 Occluded-DukeMTMC 包含 15 618 幅训练图像、17 661 幅图库图像和 2 210 幅查询图像，是目前遮挡行人再识别问题中最大且最具挑战的遮挡数据集。

本实验采用平均精度均值（mean average precision，mAP）和首次命中率（rank-1 accuracy，Rank-1）作为评价指标，对本文的网络模型进行评估。

2.2 实验设置

实验操作系统为 Ubuntu18.04，CPU 参数为 Intel®Xeon（R）Silver 4112 CPU@2.6 GHz，内存为 64 GB，使用两张 NVIDIA GeForce RTX 2080Ti 显卡进行运算加速。实验环境采用 Python3.8 和 Pytorch1.7.0 深度学习框架。采用 Adam 优化器优化模型，初始学习率设置为 0.000 15，每经过 20 个 epoch 衰减 10 倍，模型总共训练 160 个 epoch，直到网络收敛。

2.3 对比实验

为了验证本文提出方法的有效性，在 Market1501、DukeMTMC-reID 和 Occluded-DukeMT MC 3 个数据集上进行对比实验，对比方法包括基于姿态估计的方法 PGFA（pose-guided feature alignment）（Miao 等，2019）、PVPM（pose-guided visible part matching）（Gao 等，2020）、Miao 等人（2022）、PGMA-Net（pose-guided multi-attention network）（Chen 等，2022）、VGTri（Yang 等，2021）；基于特征修复的方法 HOReID（high-order re-identification）（Wang 等，2020）、MFEN（multi-branch feature enhancement network）（Liu 等，2023）；基于注意力机制的方法 CASN（consistent attentive siamese networks）（Zheng 等，2019b）、DCNN（deep convolutional neural network）（Li 等，2020）、MHSA（multihead self-attention network）（Tian 等，2023）。实验结果如表 1—表 3 所示。

从实验结果可以看出，本文针对遮挡场景的行人重识别方法，无论是在无遮挡的数据集（Market1501，DukeMTMC-reID）还是遮挡数据集（Occluded-DukeMTMC）上，都表现了很好的效果。从表中可以看出，本文方法在 Market-1501 数据集上比 PGMA-Net（Chen 等，2022）在 mAP 高了 0.4%；在

表1 在Market-1501数据集上与相关方法的比较
Table 1 Comparison with mainstream related method on the Market-1501 dataset

/%

方法	mAP	Rank-1
PGFA(Miao等,2019)	76.8	91.2
DCNN(Li等,2020)	82.7	90.2
CASN(Zheng等,2019b)	82.3	94.4
PVPM(Gao等,2020)	82.3	93.1
Miao等人(2022)	81.3	92.7
HOReID(Wang等,2020)	84.9	94.2
HG(Kiran等,2023)	86.1	**95.6**
OCNet(Kim等,2022)	87.2	94.9
PGMA-Net(Chen等,2022)	88.4	95.0
MFEN(Liu等,2023)	85.7	95.3
本文	**88.8**	95.5

注:加粗字体表示各列最优结果。

表2 在DukeMTMC-reID数据集上与相关方法的比较
Table 2 Comparison with related method on the DukeMTMC-reID dataset

/%

方法	mAP	Rank-1
PGFA(Miao等,2019)	65.5	82.6
DCNN(Li等,2020)	78.0	81.0
CASN(Zheng等,2019b)	73.7	87.7
PVPM(Gao等,2020)	71.8	84.9
HOReID(Wang等,2020)	75.6	86.9
Miao等人(2022)	74.3	86.2
HG(Kiran等,2023)	72.6	86.2
OCNet(Kim等,2022)	77.2	87.8
PGMA-Net (Chen等,2022)	78.4	88.6
MFEN(Liu等,2023)	77.3	88.3
本文	**79.2**	**89.3**

注:加粗字体表示各列最优结果。

DukeMTMC-reID数据集上比PGMA-Net (Chen等,2022)在mAP和Rank-1上分别高了0.8%和0.7%;在遮挡数据集Occluded-DukeMTMC上的性能比HG(Kiran等,2023)在mAP上提高了1.2%。相比之下,Rank-1稍显弱势的原因是特征修复单元在聚类过程中存在对背景噪声和异常点敏感,会干扰到特征的提取过程,在对图像聚类后恢复其特征的过程中会损失一部分行人特征信息。但是,本文模型在3个数据集上的mAP指标都达到了最优,说明针对遮挡场景设计的网络表现出很好的泛化能力,同时也具有良好的鲁棒性,使模型能够在复杂的实际场景中更可靠地识别行人。

表3 在Occluded-DukeMTMC数据集上与相关方法的比较
Table 3 Comparison with related method on the Occluded-DukeMTMC dataset

/%

方法	mAP	Rank-1
PGFA(Miao等,2019)	37.3	51.4
HOReID(Wang等,2020)	43.8	55.1
MHSA(Tian等,2023)	44.8	59.7
VGTri(Yang等,2021)	46.3	61.2
SORN(Zhang等,2021)	46.3	57.6
Miao等人(2022)	32.0	42.3
OCNet(Kim等,2022)	49.7	59.9
HG(Kiran等,2023)	50.5	**61.4**
MFEN(Liu等,2023)	47.6	60.8
本文	**51.7**	60.3

注:加粗字体表示各列最优结果。

从以上实验结果可以看出,本文方法结合了姿态估计网络和特征缺失修补网络的优点,姿态估计网络在图像未遮挡的部分具有良好的识别能力,而针对遮挡部分的特征修补网络在图像遮挡的区域能发挥特征修补的优势,特征增强模块的应用可以使图像中显著性的特征得到充分利用,同时也能缓解背景噪声带来的干扰。

2.4 消融实验

为了进一步验证本文网络模型的有效性,在数据集Market1501、DukeMTMC-reID和Occluded DukeMTMC上进行了消融实验。其中Baseline表示只包含ResNet50骨干网的行人重识别网络;Baseline + PEM用于验证姿态估计网络预测人体关键点的有效性;Baseline + SFPM用于验证空间特征修补模块对遮挡区域修复的有效性;Baseline + PEM + FEB用于验证姿态估计模块和特征增强模块联合使用获取显著性姿态信息的有效性,Baseline + SFPM + FEB用于验证特征修复模块和特征增强模块联合

使用的有效性；Baseline + PEM + SFPM + FEB用于验证整体网络的有效性。

从表4可以看出，本文提出的SFPM、PEM和FEB模块对Baseline网络的性能都有明显的提升。在Market1501数据集上，加入PEM模块后网络的Rank-1和mAP分别提高了1.3%和7.8%，该模块通过生成人体关键点的热力图，引导模型学习人体姿态信息，显著地提高了模型的准确率。在Occluded-DukeMTMC数据集上使用SFPM模块后，网络的Rank-1和mAP分别提高了4.1%和5.2%，说明该模块能有效地补全图像被遮挡的区域特征，帮助网络模型学习更加丰富的特征信息。在此基础上添加FEB模块使网络的Rank-1和mAP都提高了0.7%，说明该模块能帮助网络提取到显著性的特征，提升网络的性能。综上可以看出，SFPM模块可以对遮挡部位特征进行有效地修复，PEM模块使用修复后的特征预测人体关键点，从而实现更加精确的匹配。此外，通过添加FEB模块，可以增强显著性的特征信息，使网络模型学习更加具有判别性的特征，提高网络模型的识别准确率。

表4　在Market1501、DukeMTMC-reID、Occluded-DukeMTMC数据集上的消融实验结果

Table 4　Results of ablation experiments on Market-1501, DukeMTMC-reID and Occluded-DukeMTMC datasets

/%

方法	Market-1501		DukeMTMC-reID		Occluded-DukeMTMC	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
Baseline	92.0	78.1	81.5	65.9	54.7	45.2
Baseline+PEM	93.9	85.9	87.3	76.4	57.0	47.0
Baseline+SFPM	94.4	87.5	88.2	77.2	58.8	50.4
Baseline+PEM+FEB	94.2	86.4	87.8	77.1	59.7	50.7
Baseline+SFPM+FEB	94.6	88.3	88.7	78.7	59.5	51.1
Baseline+PEM+SFPM+FEB	**95.5**	**88.8**	**89.3**	**79.2**	**60.3**	**51.7**

注：加粗字体表示各列最优结果。

2.5　复杂性分析

为了对网络模型进行复杂性分析，本文使用了模型参数量（parameter）、浮点运算次数（floating-point operations，FLOPs）和每秒传输帧数（frames per second，FPS）3个指标对模型进行评估。其中，parameter是指模型训练中需要训练的参数总量，用来衡量模型的大小，其单位为M。FLOPs表示模型的时间复杂度，其数值大小反映出模型过程所需要的时间，数值越大，则需要的时间越多，其单位为MFLOPs。FPS表示每秒内可以处理的图像帧数，表5中使用FPS的倒数（1/FPS）表示处理一幅图像所需时间来评估检测速度，时间越短，速度越快。测试用的数据集是Market-1501，图像尺寸为256 × 128像素，分别使用1幅图像和2幅图像测试网络模型的FLOPs值。

从表5可以看出，SFPM和FEB两个模块在参数量较少的情况可以有效提高网络识别的准确率，其所使用的计算开销比较少，同时网络模型推理时间也比较短。在增加一倍数量图像后，网络的计算量也同样增加了一倍，说明在增大训练样本数量时，相应的计算量也会同比增加。网络模型计算量最大的模块是PEM，对应的参数量也最多，原因是使用了HRNet（high-resolution network）（Wang等，2021）网络，该网络通过维持整个姿态估计过程中的高分辨率表示，使其更准确地捕捉细节和位置信息，有助于提高姿态估计的精确度，同时通过多尺度融合过程提高关键点位置的准确性和空间精度，但是这种方法会导致整个网络的参数量增加，因此需要大量的计算资源进行训练和推理。针对该问题，下一步的工作是使姿态估计模型更加轻量化，以降低网络模型的复杂度。

2.6　可视化结果

为了更好地展示本文方法的效果，在Market1501和Occluded-DukeMTMC数据集上随机选择照片进行识别，可视化结果如图6所示。

从图6可以看出，第1组在没有脸部轮廓的情况

下全部正确匹配；第2组在有其他行人遮挡的情况下，检索结果全部正确，表明模型对于未遮挡的图像同样有很好的效果；第3组行人下半身被汽车遮挡，在有汽车和背包干扰的情况下仅第9幅图像判断错误，其余全部检索正确，说明模型在多重遮挡和干扰的情况下，仍然保持了良好的性能；第4组在行人下半身被全部遮挡的情况下，模型检索结果全部正确，证明了模型在遮挡情况下的鲁棒性。从以上结果可以看出，本文提出的网络模型能有效抑制干扰，并成功地提取出具有鉴别性的行人特征，实现了准确的匹配，同时具有良好的鲁棒性，可以在实际应用中应对各种复杂的场景和挑战。

表5　在Market1501数据集上的模型复杂性评估实验结果

Table 5　The experimental results for model complexity assessment on the Market1501 dataset

方法	FLOPs/M		参数量/M	(1/FPS)/s	mAP/%
	1幅图像	2幅图像	1幅图像	1幅图像	
Baseline	4 087.98	8 175.96	25.05	0.13	78.1
Baseline+PEM	14 619.01	29 138.02	88.65	0.26	85.9
Baseline+SFPM	4 368.40	8 736.79	25.61	0.14	87.5
Baseline+PEM+FEB	14 621.14	29 240.28	88.92	0.28	86.4
Baseline+SFPM+FEB	4 371.25	8 741.31	25.83	0.23	88.3
Baseline+PEM+SFPM+FEB	19 660.93	39 590.81	92.08	0.40	88.8

图6　在Market1501和Occluded-DukeMTMC数据集上的可视化结果

Fig. 6　Visualization results on the Market1501 and Occluded-DukeMTMC datasets

3 结 论

在遮挡场景中，当图像中可见部分的外观呈现高度相似性时，进行特征提取可能会造成误识别问题。对此，本文提出了一种融合姿态引导和多尺度特征的行人重识别方法。网络模型中的特征修复模块根据不同区域之间具有的外观相似性或位置相邻性，使得区域编码器能够输出相似的分配向量，从而实现对遮挡区域特征的补全。姿态引导模块通过检测行人姿态信息作出姿态估计和预测，进而生成姿态特征向量，然后将其与行人特征向量进行融合，获得具有区分性的行人特征信息。特征增强模块使用显著性区域特征方法增强网络提取特征的能力，提高网络对有效特征学习的能力。多项公开数据集上的实验结果表明，本文方法在遮挡情况下的行人重识别任务中展现出优良的性能。然而，应该注意的是，本文使用的姿态引导网络虽然提升了网络模型的性能，但也增加了网络的参数量和训练时间，从而影响到ReID任务的时效性。在未来的工作中，将探索在减少模型参数量的同时提升网络模型的性能，进一步优化网络结构，使其能适应各种应用场景。

参考文献(References)

Chen Y, Yang Y Z, Liu W F, Huang Y W and Li J M. 2022. Pose-guided counterfactual inference for occluded person re-identification. Image and Vision Computing, 128: #104587 [DOI: 10.1016/j.imavis.2022.104587]

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. Seattle, USA: IEEE: 11741-11749 [DOI: 10.1109/CVPR42600.2020.01176]

Hou R B, Ma B P, Chang H, Gu X Q, Shan S G and Chen X L. 2019. VRSTC: occlusion-free video person re-identification//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 7176-7185 [DOI: 10.1109/CVPR.2019.00735]

Hou R B, Ma B P, Chang H, Gu X Q, Shan S G and Chen X L. 2022. Feature completion for occluded person re-identification. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(9): 4894-4912 [DOI: 10.1109/TPAMI.2021.3079910]

Kim M, Cho M, 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]

Kiran M, Praveen R G, Nguyen-Meidine L T, Belharbi S, Blais-Morin L A and Granger E. 2023. Holistic guidance for occluded person re-identification [EB/OL]. [2023-07-20]. https://arxiv.org/pdf/2104.06524.pdf

Li Q, Hu W Y, Li J Y, Liu Y and Li M X. 2022. A survey of person re-identification based on deep learning. Chinese Journal of Engineering, 44(5): 920-932 (李擎, 胡伟阳, 李江昀, 刘艳, 李梦璇. 2022. 基于深度学习的行人重识别方法综述. 工程科学学报, 44(5): 920-932) [DOI: 10.13374/j. issn2095-9389.2020.12.22.004]

Li Y, Jiang X Y and Hwang J N. 2020. Effective person re-identification by self-attention model guided feature learning. Knowledge-Based Systems, 187: #104832 [DOI: 10.1016/j.knosys.2019.07.003]

Liang X D, Gong K, Shen X H and Lin L. 2019. Look into person: joint body parsing and pose estimation network and a new benchmark. IEEE Transactions on Pattern Analysis and Machine Intelligence, 41(4): 871-885 [DOI: 10.1109/TPAMI.2018.2820063]

Liu Z G, Wang Q, Wang M and Zhao Y J. 2023. Occluded person re-identification with pose estimation correction and feature reconstruction. IEEE Access, 11: 14906-14914 [DOI: 10.1109/ACCESS.2023.3243113]

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. Seoul, Korea (South): IEEE: 542-551 [DOI: 10.1109/ICCV.2019.00063]

Miao J X, Wu Y and Yang Y. 2022. Identifying visible parts via pose estimation for occluded person re-identification. IEEE Transactions on Neural Networks and Learning Systems, 33(9): 4624-4634 [DOI: 10.1109/TNNLS.2021.3059515]

Peng Y J, Hou S H, Cao C S, Liu X, Huang Y Z and He Z Q. 2022. Deep learning-based occluded person re-identification: a survey [EB/OL]. [2023-07-20]. https://arxiv.org/pdf/2207.14452.pdf

Ristani E, Solera F, Zou R, Cucchiara R and Tomasi C. 2016. Performance measures and a data set for multi-target, multi-camera tracking//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands: Springer: 17-35 [DOI: 10.1007/978-3-319-48881-3_2]

Schroff F, Kalenichenko D and Philbin J. 2015. FaceNet: a unified embedding for face recognition and clustering//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 815-823 [DOI: 10.1109/CVPR.2015.7298682]

Sun K, Xiao B, Liu D and Wang J D. 2019. Deep high-resolution representation learning for human pose estimation//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recogni-

tion. Los Angeles, USA: IEEE: 5686-5696 [DOI: 10.1109/CVPR.2019.00584]

Szegedy C, Vanhoucke V, Ioffe S, Shlens J and Wojna Z. 2016. Rethinking the inception architecture for computer vision//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 2818-2826 [DOI: 10.1109/CVPR.2016.308]

Tian H C, Liu X P, Yin B C and Li X. 2023. MHSA-Net: multihead self-attention network for occluded person re-identification. IEEE Transactions on Neural Networks and Learning Systems, 34(11): 8210-8224 [DOI: 10.1109/TNNLS.2022.3144163]

Wang G A, Yang S, Liu H Y, Wang Z C, Yang Y, Wang S L, Yu G, Zhou E J and Sun J. 2020. High-order information matters: learning relation and topology for occluded person re-identification//Proceedings of 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 6448-6457 [DOI: 10.1109/CVPR42600.2020.00648]

Wang J D, Sun K, Cheng T H, Jiang B R, Deng C R, Zhao Y, Liu D, Mu Y D, Tan M K, Wang X G, Liu W Y and Xiao B. 2021. Deep high-resolution representation learning for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(10): 3349-3364 [DOI: 10.1109/TPAMI.2020.2983686]

Wang L B, Zhou Y, Sun Y J and Li S. 2022. Occluded person re-identification based on differential attention siamese network. Applied Intelligence, 52(7): 7407-7419 [DOI: 10.1007/s10489-021-02820-6]

Yang J, Zhang C L, Tang Y P and Li Z X. 2022. PAFM: pose-drive attention fusion mechanism for occluded person re-identification. Neural Computing and Applications, 34(10): 8241-8252 [DOI: 10.1007/s00521-022-06903-4]

Yang J R, Zhang J W, Yu F F, Jiang X Y, Zhang M D, Sun X, Chen Y C and Zheng W S. 2021. Learning to know where to see: a visibility-aware approach for occluded person re-identification//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 11865-11874 [DOI: 10.1109/ICCV48922.2021.01167]

Zahra A, Perwaiz N, Shahzad M and Fraz M M. 2023. Person re-identification: a retrospective on domain specific open challenges and future trends. Pattern Recognition, 142: #109669 [DOI: 10.1016/j.patcog.2023.109669]

Zhang X K, Yan Y, Xue J H, Hua Y and Wang H Z. 2021. Semantic-aware occlusion-robust network for occluded person re-identification. IEEE Transactions on Circuits and Systems for Video Technology, 31(7): 2764-2778 [DOI: 10.1109/TCSVT.2020.3033165]

Zhang Y F, Yang H Y, Zhang Y J, Dou Z P, Liao S C, Zheng W S, Zhang S L, Ye M, Yan Y C, Li J J and Wang S J. 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, 张史梁, 叶茫, 晏轶超, 李俊杰, 王生进. 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]

Zheng L, Huang Y J, Lu H C and Yang Y. 2019a. Pose-invariant embedding for deep person re-identification. IEEE Transactions on Image Processing, 28(9): 4500-4509 [DOI: 10.1109/TIP.2019.2910414]

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

Zheng M, Karanam S, Wu Z Y and Radke R J. 2019b. Re-identification with consistent attentive siamese networks//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 5728-5737 [DOI: 10.1109/CVPR.2019.00588]

作者简介

张红颖，女，教授，主要研究方向为机场智能信息处理、图像处理与计算机视觉。E-mail: carole_zhang0716@163.com

刘腾飞，男，硕士研究生，主要研究方向为计算机视觉和行人重识别。E-mail: tengfei240@163.com

罗谦，男，研究员，主要研究方向为民航大数据挖掘算法和民航大数据平台仿真建模。E-mail: luoqian@caacetc.com

张涛，男，高级工程师，主要研究方向为智慧机场运行技术。E-mail: zhangtao@caacetc.com