第 59 卷 第 8 期/2022 年 4 月/激光与光电子学进展 研究论文

激光与光电子学进展 先进成像

融合注意力机制的多损失联合跨模态行人重识别方法

王凤随[1,2,3*]，刘芙蓉[1,2,3]，陈金刚[1,2,3]，王启胜[1,2,3]

[1]安徽工程大学电气工程学院，安徽 芜湖 241000；

[2]检测技术与节能装置安徽省重点实验室，安徽 芜湖 241000；

[3]高端装备先进感知与智能控制教育部重点实验室，安徽 芜湖 241000

摘要 跨模态行人重识别任务的难点在于提取出更有效的模态共享特征，为解决该问题，提出基于注意力机制的多损失联合跨模态行人重识别方法。在ResNet50网络中嵌入注意力模型，保留细节信息。将特征切割成六块局部特征，以使网络关注局部深层信息，增强网络的表征能力。对提取出的局部特征列向量进行批归一化处理，并选用交叉熵损失和改进的异质中心损失进行联合监督学习，以加速模型收敛，提升模型精度。所提方法在SYSU-MM01、RegDB数据集下的平均精度(mAP)分别达到56.82%和75.44%，实验结果表明，本文方法有效地提升了跨模态行人重识别精度。

关键词 图像处理；行人重识别；跨模态；深度学习；注意力；多损失联合

中图分类号 TP391.4 **文献标志码** A **doi**：10.3788/LOP202259.0810010

Multi-Loss Joint Cross-Modality Person Re-Identification Method Integrating Attention Mechanism

Wang Fengsui[1,2,3*], Liu Furong[1,2,3], Chen Jingang[1,2,3], Wang Qisheng[1,2,3]

[1]*School of Electrical Engineering, Anhui Polytechnic University, Wuhu, Anhui* 241000, *China*;

[2]*Anhui Key Laboratory of Detection Technology and Energy Saving Devices, Wuhu, Anhui* 241000, *China*;

[3]*Key Laboratory of Advanced Perception and Intelligent Control of High-End Equipment, Ministry of Education, Wuhu, Anhui* 241000, *China*

Abstract The difficulty of cross-modality person re-identification task is to extract more effective modal shared features. To solve the problems, this paper proposes a multi-loss joint cross-modality person re-identification method based on attention mechanism. Firstly, the attention model is embedded in the ResNet50 network, preserving the details. Secondly, the feature is divided into six local features to make the network focus on local deep information and enhance the representation ability of the network. Finally, the extracted local feature column vectors were normalized by batch processing, and the cross-entropy loss and improved hetero-center loss were used for joint supervised learning to accelerate the model convergence and improve the model accuracy. The proposed method achieves an average accuracy of 56.82% and 75.44% in the SYSU-MM01 and RegDB datasets, respectively. The experimental results show that the proposed method can effectively improve the accuracy of cross-modality person re-identification.

Key words image processing; person re-identification; cross-modality; deep learning; attention; multi-loss joint

收稿日期：2021-03-30；**修回日期**：2021-04-19；**录用日期**：2021-04-29

基金项目：安徽高校省级自然科学研究重点项目(KJ2019A0162)、安徽省自然科学基金(2108085MF197，1708085MF154)、检测技术与节能装置安徽省重点实验室开放基金(DTESD2020B02)

通信作者：*fswang@ahpu.edu.cn

1 引　　言

随着人们安全防范意识的提升，监控摄像头逐渐普及。查询监控画面成为公安系统查人寻踪的一个重要技术手段。行人重识别就是对行人的再次识别，其任务就是跨监控摄像头的行人检索。行人重识别技术是一项利用计算机视觉技术判断图像或者视频中是否存在特定行人的技术，它可以弥补固定摄像头的视觉局限[1]，从不同监控摄像头捕获的图像或视频中检索出特定行人。

目前，行人重识别方法主要基于深度学习，其中，表征学习和度量学习的方法都属于深度学习范畴。表征学习法的实质是利用行人的身份属性作为标签进行分类训练；度量学习法则是计算图片间的相似度距离。跨监控摄像头在白天光照良好的情况下捕获RGB图像，在夜晚光线不佳的环境下主要由红外摄像头捕获红外(IR)图像。RGB和IR图像间的信息差异较大，例如，RGB图像中的颜色信息是行人重识别任务中的重要识别依据，IR图像却丢失了大量的颜色信息。因此，传统行人重识别中利用单一模态中的特有信息作为特征描述符的方法在处理跨模态问题上具有局限性[2-4]。此时，RGB和IR图像间更具鲁棒性的模态共享信息就成为了跨模态行人重识别任务中的关键。2017年，Wu等[5]提出了一种深层单流网络架构的跨模态行人重识别方法。2018年，Ye等[6]提出了TONE双流网络，该网络是一种包含了表征学习和度量学习的两阶段网络框架。然而，两阶段网络训练需要人工干预，不适合实际的大规模应用。于是，Ye等[7]在TONE双流网络的基础上研究出一个端到端的网络架构来学习模态不变的共享特征，该网络被称为双向双约束高阶损失(BDTR)网络[8]。2020年，Zhu等[9]从提高类内跨模态相似性的角度出发设计了双流局部特征网络(TSLFN)和异质中心损失函数；其中TSLFN关注了局部特征，但是对全局信息中的次显著信息关注不足；异质中心损失函数使用均方误差(MSE)算法计算损失值，然而，MSE算法对离群点、异常值较为敏感，梯度变化较大。

针对上述问题，本文从网络结构和损失函数的角度出发，提出了一种基于融合注意力机制的多损失联合网络：1) 采用双流网络来提取RGB、IR图像的深层特征，获取高层语义信息；2) 特征提取模块利用注意力机制并融合原有特征，提高了网络对高级别模态共享特征的捕获和表征能力；3) 设计特征约束模块来提高模型精度，使用交叉熵损失和平滑中心损失联合训练网络模型，从而减小离群点对网络的负面影响。

2 算法原理

2.1 算法流程

本文基于ResNet50[10]网络设计了一种双流跨模态行人重识别算法，如图1所示。

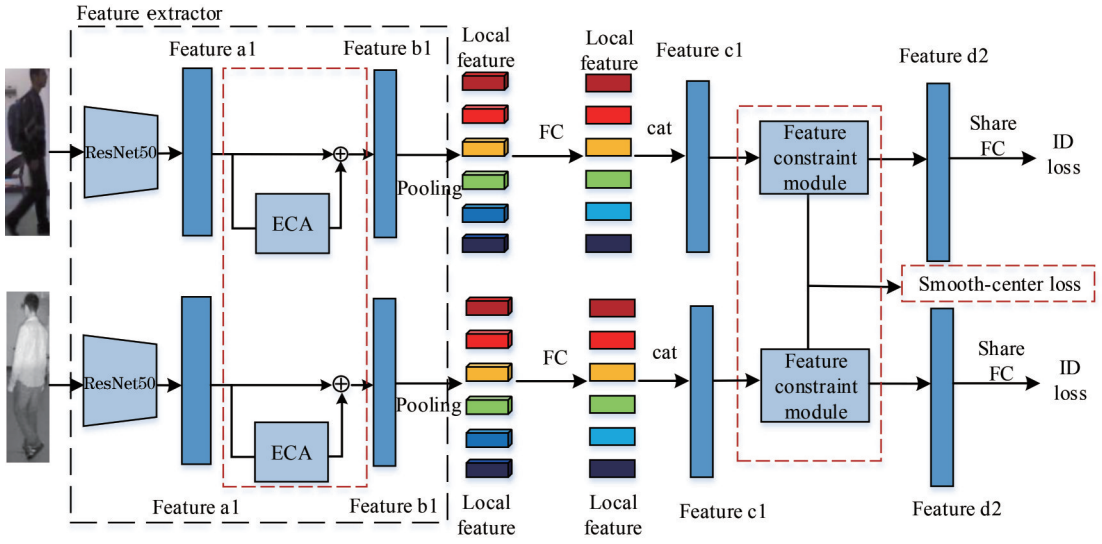

图1　融合注意力模型的多损失联合跨模态网络结构图

Fig. 1　Diagram of multi-loss joint cross-modality network structure of fused attention model

所提算法的具体步骤如下：

1）使用 ResNet50 网络模型提取基础全局特征，为扩大感受野范围、增强细粒度，去除 ResNet50 第 4 层和第 5 层之间的下采样过程，设计并使用注意力机制对特征信息进行加权并融合基础全局特征，以增强模型对高级共享语义信息的捕获能力。

2）在平均池化层后输出 6 个局部特征列向量，深层网络的感受野很大，所以每块局部特征在包含全局信息的基础上更加关注当前的局部区域。

3）各块局部特征分别通过两层全连接（FC）层，且第二层参数共享。将特征约束模块嵌入两个全连接层之间，以提高模型精度与稳定性。为了减小离群点对损失值的影响，在异质中心损失（HC）的基础上设计了平滑中心损失。最后使用平滑中心损失和交叉熵损失对各块局部特征进行联合监督学习。

2.2　深层跨模态共享特征提取网络

ResNet50 解决了网络随深度的增加而出现的退化问题，提高了深层网络的鲁棒性，其本质是一个恒等映射：

$$F(x)=H(x)-x, \tag{1}$$

式中：x 为输入；$H(x)$ 为期望输出；$F(x)$ 为残差函数，表示学习到的残差。当 $F(x)=0$ 时，残差网络满足恒等映射关系。

本文可共享的深层跨模态特征提取网络由改进的 ResNet50 网络和注意力模块组成，采用了双流网络设计，每层结构相同；其中，针对 RGB 图像的特征提取模块如图 2 所示。首先，ResNet50 网络去除了第 4 层 Conv4_x 和第 5 层 Conv5_x 之间的下采样过程，这有效扩大了感受野的范围，为后续局部特征的切割、池化操作提供基础，Conv5_x 层后输出维度为 2048×18×9 的特征图。其次，在 ResNet50 网络的第 5 层（Conv5_x）和平均池化层之间嵌入注意力模块，以加强局部跨通道交互，增强网络共享特征提取能力。受残差单元的启发，本文将基础全局特征和加权后的高层重要特征进行融合。平均池化层将特征图切割成 6 块 2048×6×1 的局部特征列向量，由于高层网络有较大的感受野，此时每一块局部特征在聚焦局部的同时也关注了全局信息。

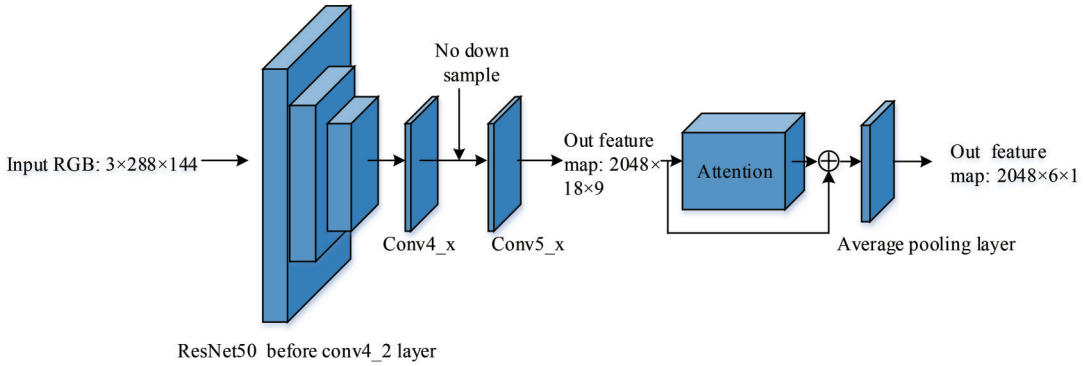

图 2　特征提取模块

Fig. 2　Feature extraction module

2.3　注意力模块

红外图像的特殊性导致颜色信息和红外信息在跨模态行人重识别任务中的鲁棒性降低，于是其他信息变得异常关键。注意力机制的本质就是通过一系列权重参数来对图像的重要信息进行加强，并抑制一些无关信息。网络越深，注意力机制捕获的特征级别也越高，信息共享能力就越强。

本文算法将全局特征水平分割并分别计算损失，由于深层网络具有较大的感受野，此时的局部特征并未完全丢失全局信息。为了防止全局特征中的重要信息丢失，首先，注意力模型 ECA[11]（efficient channel attention）在 ResNet50 的 Conv5_x 和平均池化层之间关注高层特征并对其进行加权，这些高层特征中富含的语义信息能够帮助跨模态网络辨别行人身份；其次，为了让网络更加均匀化，设计了一个平直的流行结构，即让 2048×18×9 的特征图和加权后的高层特征进行融合，以进一步保留全局特征中的重要信息和次重要信息；最后，对最终的全局特征进行水平切割，此时的局部特征中融合了更多的全局高语义共享信息。

本文的注意力模型是在 SE[12]基础上改进的，它会自适应地对特征图进行特征通道筛选，强调学习突出的信息，并抑制无关的通道。如图 3 所示（图中 W 表示宽，H 表示高，GAP 表示全局平均池化），它主要包括三项工作：

1）避免降低通道维数。

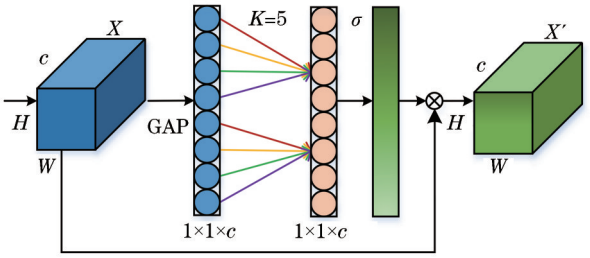

图 3　Attention结构图

Fig. 3　Attention structure

2）捕获局部的跨通道交互。本文算法共享通道参数，其中每个通道的注意力模块会涉及 $k\times c$ 个参数，参数量减少了。这种局部性约束避免了跨所有通道的交互，从而提高了模型效率。

$$\omega_i=\sigma(\alpha_i^j y_i^j),\ y_i^j\in\Omega_i^k, \tag{2}$$

式中：ω_i 为特征 y_i 的权重；α_i^j 表示特征 y_i^j 的权重；y_i^j 表示特征 y_i 的第 j 列；Ω_i^k 表示特征 y_i 的 k 个邻域的集合。

3）自适应选择卷积核大小。使用一维卷积实现局部的跨通道交互时，卷积核 k 的大小决定了局部的交互范围。卷积核 k 可以表示为

$$k=\left|\frac{\log_2 c}{\gamma}+\frac{b}{\gamma}\right|, \tag{3}$$

式中：c 表示通道数；$b=1$；$\gamma=2$。

2.4　损失函数

在跨模态行人重识别任务中，特征描述符由网络提取出的模态共享特征组成。为了提高样本的类间距离和类内相似度，本文使用交叉熵损失和改进的异质中心损失（L_{HC}）对网络进行联合监督学习。总损失（L）等于交叉熵损失（L_{softmax}）与平滑中心损失（L_{SC}）的和，L 可表示为

$$L=L_{\mathrm{softmax}}+\lambda L_{\mathrm{SC}}, \tag{4}$$

式中：λ 为平滑中心损失值占总损失值的比例；L_{softmax} 可以表示为

$$L_{\mathrm{softmax}}=-\sum_{i=1}^{K}\ln\frac{\exp(\boldsymbol{W}_{y_i}^{\mathrm{T}}\boldsymbol{x}_i+b_{y_i})}{\sum_{j=1}^{n}\exp(\boldsymbol{W}_j^{\mathrm{T}}\boldsymbol{x}_i+b_j)}, \tag{5}$$

式中：K 为批次大小；$\boldsymbol{x}_i$ 为第 y_i 类中第 i 个样本的特征；$\boldsymbol{W}_{y_i}^{\mathrm{T}}$ 为第 y_i 类中权值矩阵的转置；$\boldsymbol{W}_j$ 为 $\boldsymbol{W}$ 的第 j 行参数组成的向量；b_{y_i} 为第 j 行的偏置量；b_j 为第 y_i 类中的偏置大小。L_{HC} 约束的是样本中心距离，损失值越小说明类内相似度越高，L_{HC} 可表示为

$$L_{\mathrm{HC}}=\sum_{i=1}^{U}\|C_{i,1}-C_{i,2}\|_2^2, \tag{6}$$

式中：U 表示类别数目；$C_{i,1}$ 表示第 i 类中 RGB 图像的特征中心，$C_{i,2}$ 表示第 i 类中 IR 图像的特征中心，可分别表示为

$$\begin{cases}C_{i,1}=\dfrac{1}{M}\sum_{j=1}^{M}X_{i,1,j}\\ C_{i,2}=\dfrac{1}{N}\sum_{j=1}^{N}X_{i,2,j}\end{cases}, \tag{7}$$

式中：$X_{i,1,j}$ 和 $X_{i,2,j}$ 分别表示第 i 类中第 j 幅 RGB 图像和 IR 图像；M、N 分别表示第 i 类中 RGB 图像和 IR 图像的数量。$C_{i,1}$ 和 $C_{i,2}$ 满足

$$\frac{\partial L_{\mathrm{HC}}}{\partial X_{i,1,j}}=\frac{\partial L_{\mathrm{HC}}}{\partial C_{i,1}}\frac{\partial C_{i,1}}{\partial C_{i,1}}=\frac{2}{N}(C_{i,1}-C_{i,2})。 \tag{8}$$

异质中心损失选用 MSE 算法计算两个局部中心的差值。MSE 算法具有可导、快速收敛的优点。但是，当差值较大时，MSE 算法会给予较大的惩罚；当差值较小时，MSE 会给予较小的惩罚。因此，MSE 算法具有对偏离中心点较为敏感的特点，离群点会获得更高的权重，从而对模型造成负面影响。

平滑中心损失 L_{SC} 可表示为

$$L_{\mathrm{SC}}=\sum_{i=1}^{U}\frac{\|C_{i,1}-C_{i,2}\|_2^2}{2}, \tag{9}$$

$$L_{\mathrm{SC}}=\sum_{i=1}^{U}\left\|C_{i,1}-C_{i,2}\right\|_2-\frac{1}{2}。 \tag{10}$$

在异质中心损失函数的基础上，本文设置惩罚阈值 β：当中心距离小于 β 时，平滑中心损失函数如（9）式所示，为异质中心损失的 1/2；当中心距离大于等于阈值 β 时，平滑中心损失函数如（10）式所示。与异质中心损失不同的是，平滑中心损失采用了分段函数的设计，在阈值范围内，平滑中心损失采用 MSE 算法计算中心距离，超出阈值范围则计算中心点差值的二范数值再减去 1/2。平滑中心损失对于阈值范围外的局部中心离群样本的关注更少，这减小了离群点对网络的负面影响。平滑中心损失关于 $x_{i,1,j}$ 的偏导数为

$$\frac{\partial L_{\mathrm{SC}}}{\partial x_{i,1,j}}=\frac{\partial L_{\mathrm{SC}}}{\partial C_{i,1}}\frac{\partial C_{i,1}}{\partial C_{i,1}}=\frac{1}{N}(C_{i,1}-C_{i,2}), \tag{11}$$

$$\frac{\partial L_{\mathrm{SC}}}{\partial x_{i,1,j}}=\frac{1}{N}。 \tag{12}$$

从（8）、（11）式中可以看出，在阈值范围内，平滑中心损失函数的回传梯度是异质中心损失的一半。（12）式表示，当中心距离大于阈值时，平滑中心损失函数的偏导数为 $1/N$，即图片数量 N 越大，梯度越小。SC 损失在拉近类内中心距离的同时，受离群

点的影响更小，表现更为平滑。

2.5　特征约束模块

为了加速模型收敛、提升模型精度、缓解梯度弥散，采用本文算法，如图 4 所示。首先将提取出的局部模态不变特征直接作为平滑中心损失的优化目标；其次利用 BNNeck[13] 对特征进行批归一化处理，特征近似地分布在超球面；最后将归一化后的特征送入分类器中，使得两种损失函数同时在适合的空间里同步收敛，约束特征。

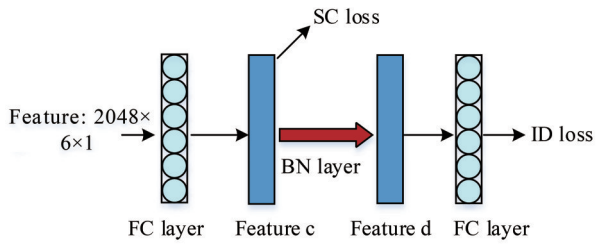

图 4　特征约束模块

Fig. 4　Feature constraint module

为了对每一批次的数据进行归一化，首先计算每一批次的均值 μ_{B} 和方差 σ_{B}^2：

$$\mu_{\mathrm{B}}=\frac{1}{m}\sum_{i=1}^{m}\left|\boldsymbol{x}_i\right|, \tag{13}$$

$$\sigma_{\mathrm{B}}^2=\frac{1}{m}\sum_{i=1}^{m}\left(\left|\boldsymbol{x}_i\right|-\mu_{\mathrm{B}}\right)^2, \tag{14}$$

式中：m 表示批次大小。

利用批均值 μ_{B} 和批方差 σ_{B}^2 对每一批次的输入数据进行归一化，可得

$$\hat{x}_i=\frac{\left|\boldsymbol{x}_i\right|-\mu_{\mathrm{B}}}{\sqrt{\sigma_{\mathrm{B}}^2+\varepsilon}}, \tag{15}$$

式中：ε 为微小正数，以避免(15)式的除数为零。最后，对 $x_{i,1,j}$ 进行尺度变换和偏移操作得到规范后的网络响应 y_i：

$$y_i=\gamma\hat{x}_i+\beta, \tag{16}$$

式中：γ、β 表示学习参数。

3　实验结果与分析

3.1　实验设置

实验使用的操作系统为 Ubuntu 16.04，GPU 选用 NVIDIA GeForce RTX 2080Ti(11 GB)，处理器为英特尔 Core i9-10900@3.7 GHz，深度学习框架为 Pytorch 1.2.0。

将行人图像预处理为 288×144 大小；数据增广策略采用对图像进行随机旋转和剪裁的方式；epoch 大小设置为 60；初始学习率设为 0.01，前 30 个 epoch 的学习率为 10^{-2}，后 30 个 epoch 的学习率为 10^{-4}；优化器采用 SGD，其中动量设置成 0.9；batch size 大小为 32。

3.2　数据集和评价标准

RegDB 数据集共采集到 4120 张可见光图像和 4120 张红外图像；其中包括 412 个行人，每个行人具有 10 张 RGB 图像、10 张 IR 图像。训练集包含 206 个行人，4120 张图；测试集包含 206 个行人身份，4120 张图。SYSU-MM01 数据集有 287628 张 RGB 图像、15792 张 IR 图像，共有 491 个不同身份的行人信息。其中训练集含有 395 人，测试集含有 96 人；训练集中有 22258 张 RGB 图像和 11909 张 IR 图像；测试集中有 3803 张可查询图像，将随机选取的 301 张 RGB 图像作为图库集。

SYSU-MM01 数据集测试模式分为两种：一种是全搜索模式，包括所有摄像头；另一种是室内模式，使用室内摄像头搭建。全搜索模式比室内模式的环境更加多样复杂，因此全搜索模式的难度更大，而室内模式能够更好地评估跨模态网络模型的检索能力，更接近理想状态。采用本文算法进行测试时选用全搜索模式并且设置了 Single-hot 和 Multi-hot。

测评时，使用 Rank-n、mAP 作为评估标准；Rank-n 表示搜索结果最靠前的 n 张图片的准确率，当 n=1，10，20 时计算测试集中前 1，10，20 张与查询集中图片经相似度排序后为同一标签的准确率。重复 10 次实验，取平均值作为最终评估结果。mAP 可表示为

$$V_{\mathrm{mAP}}=\frac{\sum_{i=1}^{n}V_{\mathrm{AP},i}}{C}, \tag{17}$$

式中：$V_{\mathrm{AP},i}$ 表示每个类别的平均精度；C 表示类别数。

3.3　实验过程

为了展现平滑中心损失对总损失值的影响，本文进行了对照组实验。在保持相同网络结构的前提下，异质中心损失和平滑中心损失分别与交叉熵损失联合监督网络。图 5 展现了训练过程中损失值与准确值随 epoch 变化的关系图。损失值与 epoch 的关系如图 5(a)所示，my_loss 曲线始终位于 base_loss 下方。准确度与 epoch 的关系如图 5(b)所示，my_acc 基本位于 base_acc 上方。实验结果说明本文算法的总损失值更小，约束力更强，平滑中心损失受

到异常值的影响更小，这弱化了离群点对总损失值的贡献。其次，基于平滑中心损失的多损失联合网络比基线网络的准确度上升得更快，且波动幅度更小。因此，平滑中心损失的有效性得到验证。

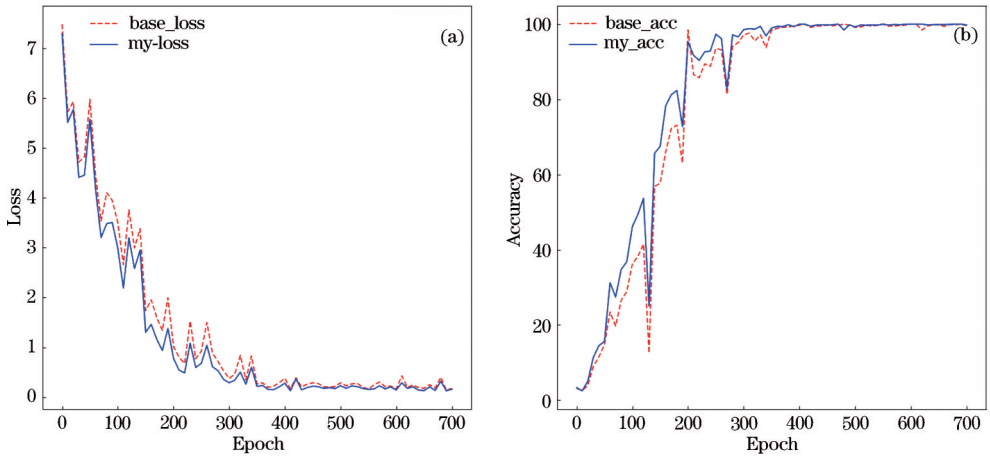

图 5　训练阶段损失值与准确度的变化趋势。(a)损失值；(b)准确度

Fig. 5　Trends of loss and accuracy during training stage. (a) Loss; (b) accuracy

为了验证本文算法关键部分的有效性，如表 1 所示，在 RegDB 数据集下分别对 ECA 模型和 BN 模块进行实验，训练过程采用平滑中心损失与交叉熵损失联合监督学习。对多损失联合基础跨模态网络进行测试，其 mAP 达到 69.89%，首位击中率为 79.17%。添加 BN 后首位击中率大幅提升 5.3%，由此可见 BN 有效缓解了梯度弥散，对网络稳定性起到促进作用。在基础网络上添加注意力模块，实验发现融合注意力模型的网络首位击中率和平均精度都有所提升，这验证了注意力机制对网络特征的提取能力有加强作用，丰富了特征描述符。在基础网络中同时引入 BN 和 ECA 模型，算法 mAP 达到 75.44%，Rank-1 达到 85.63%，与基础网络相比 mAP 和 Rank-1 分别提升了 5.55% 和 6.46%，且较单独添加 BN 或 ECA 模块的基础网络性能都得到大幅度提升。

表 1　各模块在 RegDB 数据集下的实验结果

Table 1　Experimental results of each module in RegDB dataset　%

BN	ECA	Rank-1	Rank-10	Rank-20	mAP
×	×	79.17	94.08	97.52	69.89
×	√	79.81	94.71	97.72	70.50
√	×	84.47	96.80	98.16	72.10
√	√	85.63	96.84	98.50	75.44

图 6 显示的是本文算法与 TSLFN[9] 在训练阶段的 mAP 对比图，base_mAP 表示 TSLFN 算法的训练过程，my_mAP 为本文算法的训练过程。从

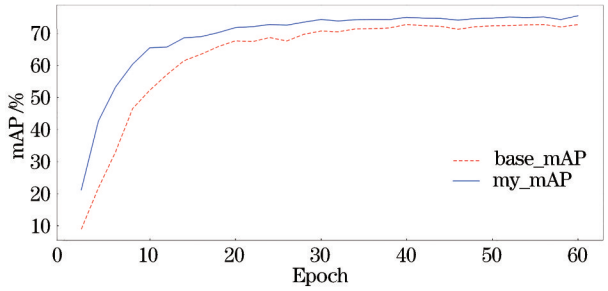

图 6　训练阶段的 mAP 变化趋势

Fig. 6　Trend of mAP in training stage

图 6 可以看出，本文算法比 TSLFN 算法的起始 mAP 值高 10%，且收敛更快。整个训练过程中本文算法的平均精度持续高于 TSLFN。

3.4　与其他方法的对比

为了验证本文算法的有效性，将本文算法与其他方法在同一数据集 SYSU-MM01 下进行对比，并且采用同一评估指标 Rank-n 和 mAP；其中，HPILN (ResNet50)[14]、AGW[15]、LZM[16] 和本文算法都是基于 ResNet50 网络框架。图 7、图 8 分别表示在 SYSU-MM01 数据集全搜索模式下本文算法与其他方法的比较结果[5-9,14-27]。从图中可以看出，本文算法的首位击中率和平均精度大幅优于传统行人重识别法。本文算法较 HPILN(ResNet50)在全搜索 single、multi 设置下的 mAP 提升了 13.87%、13.14%，Rank-1 分别增加 16.83%、16.34%。HPILN(ResNet50)采用的是 soft-max 函数作为分类损失，联合 HPI loss 和 soft-max loss 共同学习；其中 HPI loss 由全局难样本三元组和跨模态难样本三

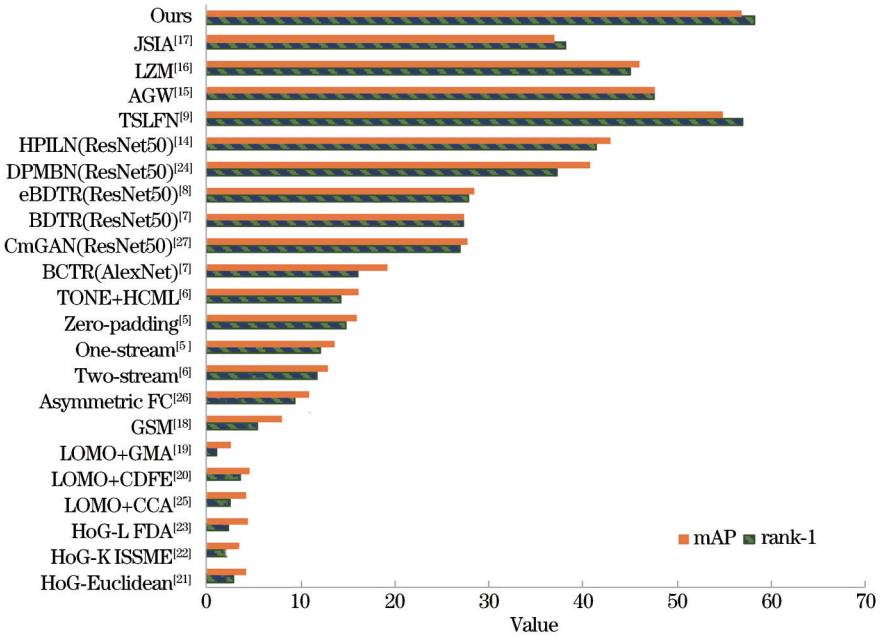

图 7　在 SYSU-MM01 全搜索 single 模式下本文算法和其他先进方法的对比

Fig. 7　Comparison of proposed method and other advanced methods for SYSU-MM01 all-search single mode

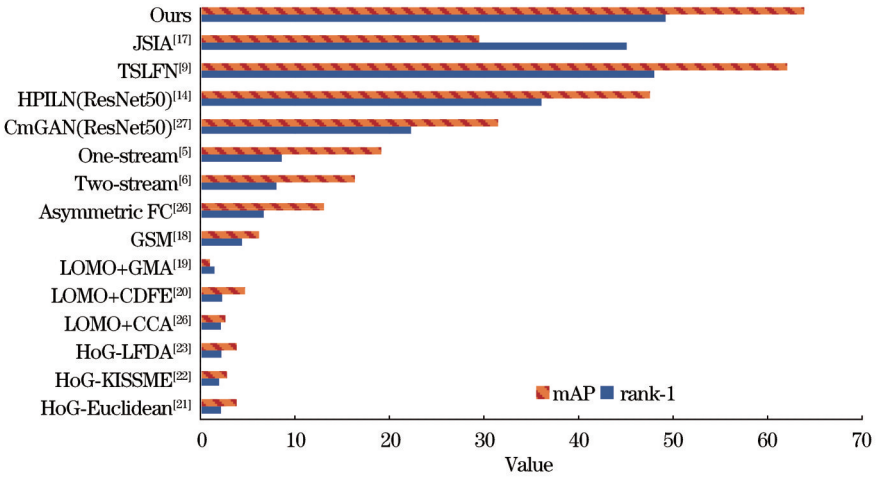

图 8　在 SYSU-MM01 全搜索 multi 模式下本文算法和其他先进方法的对比

Fig. 8　Comparison of proposed method and other advanced methods for SYSU-MM01 all-search multi mode

元组损失函数组成；三元组损失函数通过计算基准样本、正样本、负样本间的距离来实现样本间的相似性计算；但是，它计算的是相对距离，损失大小与正负样本的绝对距离无关。而本文算法采用的是平滑中心损失函数，具有很强的聚类能力，能大幅减小类内跨模态中心距离。本文算法与 AGW、LZM 算法相比，在全搜索模式 single 模式下的 mAP 分别提升了 9.17% 和 19.92%。AGW 和 LZM 算法基于图像的全局特征，本文算法设计出的双流局部网络结构在不丢失全局信息的同时更关注局部信息。对比实验结果表明本文算法对跨模态图像的深层共享特征的提取能力更强。

表 2 为在 SYSU-MM01 数据集室内模式下本文算法与其他先进方法的比较结果，从表中可以看出本文算法在室内 single-hot 和 multi-hot 设置下，mAP 分别为 68.11%、60.29%，Rank-1 分别为 60.19%，71.56%。本文算法比 HPILN 算法在室内 single-hot 和 multi-hot 设置下的 mAP 分别高出 11.59% 和 12.81%，Rank-1 分别高出 14.42% 和 18.51%；本文算法较 JSIA 算法[17]在室内 single-hot 和 multi-hot 设置下的 mAP 分别高出 15.21% 和 17.59%。

表 2 在 SYSU-MM01 indoor-search 模式下本文方法与跨模态行人再识别的对比实验结果

Table 2 Comparative experiment results between our method and others for SYSU-MM01 indoor-search mode

Method	Single-hot				Multi-hot			
	Rank-1	Rank-10	Rank-20	mAP	Rank-1	Rank-10	Rank-20	mAP
GSM[18]	9.46	48.98	72.06	15.57	11.36	51.34	73.41	9.03
LOMO+GMA[19]	1.79	17.90	36.01	5.63	1.71	18.11	36.17	2.88
LOMO+CDFE[20]	5.75	34.35	54.90	10.19	7.36	40.38	60.33	5.64
HoG-Euclidean[21]	3.22	24.68	44.52	7.52	4.75	29.06	49.38	3.51
HoG-KISSME[22]	3.11	25.47	46.57	7.43	4.10	29.32	50.59	3.61
HoG-LFDA[23]	2.44	24.13	45.50	6.87	3.42	25.27	45.11	3.19
DPMBN(ResNet50)[24]	44.47	87.12	95.24	54.51	—	—	—	—
LOMO+CCA[25]	4.11	30.60	52.54	8.83	4.86	34.40	57.30	4.47
Asymmetric FC[26]	14.59	57.94	78.68	20.33	20.09	69.37	85.80	13.04
CmGAN(ResNet50)[27]	31.36	77.23	89.18	42.19	37.00	80.94	92.11	32.76
One-stream[5]	16.94	63.55	82.10	22.95	22.62	71.74	87.82	15.04
Zero-padding[5]	20.58	68.38	85.79	26.2	24.43	75.86	91.32	18.64
Two-stream[6]	15.60	61.18	81.02	21.49	22.49	72.22	88.61	13.92
BDTR(ResNet50)[7]	31.92	77.18	89.28	41.86	—	—	—	—
eBDTR(ResNet50)[8]	32.46	77.42	89.62	42.46	—	—	—	—
TSLFN[9]	59.74	92.07	96.22	64.91	69.76	95.85	98.90	57.81
HPILN(ResNet50)[14]	45.77	91.82	98.46	56.52	53.05	93.71	98.93	47.48
AGW[15]	54.17	—	—	62.97	—	—	—	—
JSIA[17]	43.8	86.2	94.2	52.9	52.7	91.1	96.4	42.7
Ours	**60.19**	**96.74**	**99.41**	**68.11**	**71.56**	**97.23**	**99.41**	**60.29**

4 结 论

提出了融合注意力机制的多损失联合跨模态行人重识别方法。首先使用加入 ECA 模型的改进双流 ResNet50 网络提取跨模态共享特征，在确保获取全局信息的同时加强了局部深层特征的提取能力；其次，在 FC 层间嵌入特征约束模块；最后，使用平滑中心损失和交叉熵损失进行联合监督学习，以提高网络对离群点的敏感度，提高稳定性。将本文算法与 TSLFN 算法相比，可得：本文算法在 RegDB 数据集上的总损失值更小，mAP 提高了 3.44%，Rank-1 增加了 2.63%；本文算法在 SYSU-MM01 数据集全搜索 multi 设置下，mAP 提升了 1.87%，Rank-1 增加了 1.23%。本文算法有效提升了网络精度和模型收敛速度，并且提高了模型稳定性。

参 考 文 献

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

[2] Li C, Jiang M, Kong J. Multi-branch person re-identification based on multi-scale attention[J]. Laser & Optoelectronics Progress, 2020, 57(20): 201001.
李聪, 蒋敏, 孔军. 基于多尺度注意力机制的多分支行人重识别算法[J]. 激光与光电子学进展, 2020, 57(20): 201001.

[3] Zhang T, Yi Z M, Li X, et al. Improved algorithm for person re-identification based on global features [J]. Laser & Optoelectronics Progress, 2020, 57(24): 241503.
张涛, 易争明, 李璇, 等. 一种基于全局特征的行人重识别改进算法[J]. 激光与光电子学进展, 2020, 57(24): 241503.

[4] Liu K W, Fang P P, Xiong H X, et al. Person re-identification based on multi-layer feature[J]. Laser & Optoelectronics Progress, 2020, 57(8): 081503.
刘可文, 房攀攀, 熊红霞, 等. 基于多层级特征的行人重识别[J]. 激光与光电子学进展, 2020, 57(8): 081503.

[5] Wu A C, Zheng W S, Yu H X, et al. RGB-infrared cross-modality person re-identification[C]//2017 IEEE International Conference on Computer Vision (ICCV),

October 22-29, 2017, Venice, Italy. New York: IEEE Press, 2017: 5390-5399.

[6] Ye M, Lan X, Li J, et al. Hierarchical discriminative learning for visible thermal person re-identification [C]//Proceedings of the 32nd AAAI Conference on Artificial Intelligence, February 2-7, 2018, New Orleans, LA, USA. Menlo Park: AAAI, 2018: 7501-7508.

[7] Ye M, Wang Z, Lan X Y, et al. Visible thermal person re-identification via dual-constrained top-ranking [C]//Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, July 13-19, 2018, Stockholm, Sweden. Menlo Park: International Joint Conferences on Artificial Intelligence Organization, 2018: 1092-1099.

[8] Ye M, Lan X Y, Wang Z, et al. Bi-directional center-constrained top-ranking for visible thermal person re-identification[J]. IEEE Transactions on Information Forensics and Security, 2020, 15: 407-419.

[9] Zhu Y X, Yang Z, Wang L, et al. Hetero-center loss for cross-modality person re-identification[J]. Neurocomputing, 2020, 386: 97-109.

[10] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 770-778.

[11] Wang Q L, Wu B G, Zhu P F, et al. ECA-net: efficient channel attention for deep convolutional neural networks[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 11531-11539.

[12] Hu J, Shen L, Albanie S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8): 2011-2023.

[13] Luo H, Jiang W, Gu Y Z, et al. A strong baseline and batch normalization neck for deep person re-identification[J]. IEEE Transactions on Multimedia, 2020, 22(10): 2597-2609.

[14] Zhao Y B, Lin J W, Xuan Q, et al. HPILN: a feature learning framework for cross-modality person re-identification[J]. IET Image Processing, 2019, 13 (14): 2897-2904

[15] Ye M, Shen J B, Lin G J, et al. Deep learning for person re-identification: a survey and outlook[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 4775, PP(99): 1.

[16] Wang G A, Zhang T Z, Yang Y, et al. Cross-modality paired-images generation for RGB-infrared person re-identification[C]//The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, February 7-12, 2020, New York, NY, USA. Menlo Park: AAAI, 2020: 12144-12151.

[17] Basaran E, Gökmen M, Kamasak M E. An efficient framework for visible-infrared cross modality person re-identification[J]. Signal Processing: Image Communication, 2020, 87: 115933.

[18] Lin L, Wang G R, Zuo W M, et al. Cross-domain visual matching via generalized similarity measure and feature learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1089-1102.

[19] Sharma A, Kumar A, Daume H, et al. Generalized Multiview Analysis: a discriminative latent space [C]//2012 IEEE Conference on Computer Vision and Pattern Recognition, June 16-21, 2012, Providence, RI, USA. New York: IEEE Press, 2012: 2160-2167.

[20] Lin D H, Tang X O. Inter-modality face recognition [M]//Leonardis A, Bischof H, Pinz A. Computer vision-ECCV 2006. Lecture notes in computer science. Heidelberg: Springer, 2006, 3954: 13-26.

[21] Surasak T, Takahiro I, Cheng C H, et al. Histogram of oriented gradients for human detection in video[C]//2018 5th International Conference on Business and Industrial Research (ICBIR), May 17-18, 2018, Bangkok, Thailand. New York: IEEE Press, 2018: 172-176.

[22] Köstinger M, Hirzer M, Wohlhart P, et al. Large scale metric learning from equivalence constraints [C]//2012 IEEE Conference on Computer Vision and Pattern Recognition, June 16-21, 2012, Providence, RI, USA. New York: IEEE Press, 2012: 2288-2295.

[23] Pedagadi S, Orwell J, Velastin S, et al. Local fisher discriminant analysis for pedestrian re-identification [C]//2013 IEEE Conference on Computer Vision and Pattern Recognition, June 23-28, 2013, Portland, OR, USA. New York: IEEE Press, 2013: 3318-3325.

[24] Xiang X Z, Lü N, Yu Z T, et al. Cross-modality person re-identification based on dual-path multi-branch network[J]. IEEE Sensors Journal, 2019, 19 (23): 11706-11713.

[25] Rasiwasia N, Costa Pereira J, Coviello E, et al. A new approach to cross-modal multimedia retrieval

[C]//Proceedings of the international conference on Multimedia-MM'10, October 25-29, 2010, Firenze, Italy. New York: ACM Press, 2010: 251-260.

[26] Escobar-Cabrera E, Lario P, Baardsnes J, et al. Asymmetric Fc engineering for bispecific antibodies with reduced effector function[J]. Antibodies, 2017, 6(2): 7.

[27] Dai P Y, Ji R R, Wang H B, et al. Cross-modality person re-identification with generative adversarial training[C]//Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, July 13-19, 2018, Stockholm, Sweden. Menlo Park: International Joint Conferences on Artificial Intelligence Organization, 2018: 677-683.