基于阶梯型特征空间分割与局部注意力机制的行人重识别

石跃祥 周 玥*

(湘潭大学计算机学院·网络空间安全学院 湘潭 411105)

摘 要：为了让网络捕捉到更有效的内容来进行行人的判别，该文提出一种基于阶梯型特征空间分割与局部分支注意力网络(SLANet)机制的多分支网络来关注局部图像的显著信息。首先，在网络中引入阶梯型分支注意力模块，该模块以阶梯型对特征图进行水平分块，并且使用了分支注意力给每个分支分配不同的权重。其次，在网络中引入多尺度自适应注意力模块，该模块对局部特征进行处理，自适应调整感受野尺寸来适应不同尺度图像，同时融合了通道注意力和空间注意力筛选出图像重要特征。在网络的设计上，使用多粒度网络将全局特征和局部特征进行结合。最后，该方法在3个被广泛使用的行人重识别数据集Market-1501，DukeMTMC-reID和CUHK03上进行验证。其中在Market-1501数据集上的mAP和Rank-1分别达到了88.1%和95.6%。实验结果表明，该文所提出的网络模型能够提高行人重识别准确率。

关键词：行人重识别；特征空间分割；注意力机制；局部特征

中图分类号：TN911.73; TP391.41 **文献标识码**：A **文章编号**：1009-5896(2022)01-0195-08

DOI: 10.11999/JEIT201006

Person Re-identification Based on Stepped Feature Space Segmentation and Local Attention Mechanism

SHI Yuexiang ZHOU Yue

(*School of Computer Science and School of Cyberspace Security*, *Xiangtan University*, *Xiangtan* 411105, *China*)

Abstract: In order to make the network capture more effective content distinguish pedestrians, this paper proposes a multi-branch network based on Stepped feature space segmentation and Local Branch Attention Network (SLANet) mechanism to pay attention to the salient information of local images. First of all, a stepped branch attention module is introduced into the network. This module blocks the feature map horizontally in a stepped manner, and branch attention is used to assign different weights to each branch. Secondly, a multi-scale adaptive attention module is introduced into the network, which processes local features and adapts the size of the receptive field to adapt to images of different scales. Meanwhile, channel attention and spatial attention are combined to screen out the important features of the image. In the design of network, the multi-granularity network is used to combine the global feature with the local feature. Finally, the method is validated on three widely used person re-identification data sets Market-1501, DukeMTMC-reID and CUHK03. Among them, mAP and Rank-1 on market-1501 data set reach 88.1% and 95.6% respectively. The experimental results show that the proposed network model can improve the accuracy of person re-identification.

Key words: Person re-identification; Feature space segmentation; Attention mechanism; Local features

1 引言

行人重识别是在跨监控设备下检索出给定行人图像目标的技术，广泛应用于智能安防、人机交互、电子商务等领域。由于监控设备下的行人图像存在角度变化、光照、遮挡和分辨率低等问题，行人重识别仍然是一项具有挑战性的任务。早期行人重识别的研究通常是对行人图像提取颜色直方图、HOG、纹理等手工特征。但手工设计的特征描述能力有限，传统算法识别率比较低。此外，许多研究者在距离度量学习方法上进行研究，周智恒等人[1]提出一种等距离度量学习策略，陈莹等人[2]提出一种双向参考集矩阵度量学习算法。为克服度量模型的过拟合问题，He等人[3]提出环推度量学习算法。2016年后，随着深度学习在行人重识别任务中

收稿日期：2020-11-30；改回日期：2021-10-21；网络出版：2021-11-16
*通信作者： 周玥 zhyue621@163.com
基金项目：国家自然科学基金(61602397, 61502407)
Foundation Items: The National Natural Science Foundation of China (61602397, 61502407)

的应用，算法性能大大超过了早期的传统方法。

基于深度学习的行人重识别主要可以归纳为基于生成对抗网络[4]、特征空间分割、行人姿势和注意力机制的方法[5]。研究者利用生成对抗网络来扩大数据集和增加数据多样性，如Zheng等人[6]利用生成对抗网络来生成更多模拟的数据。在全局特征遇到性能瓶颈后，研究者更加注重局部特征的研究。Sun等人[7]提出的PCB模型将所提取的全局特征均匀分成6个水平块。Wang等人[8]提出了一种整合全局特征和局部特征的多粒度模型。但简单的分块可能造成块与块间有效信息的丢失，且无法实现水平块之间的对齐。Zhao等人[9]借助人体的14个姿势关键点生成7个子区域，然后将其与全局特征一起送入特征融合网络得到最终的行人特征表示。Miao等人[10]利用姿态将有用信息从遮挡中分离出来。结合人体姿势信息进行预测，能够有效避免姿态错位导致的特征对齐困难，但需要大量额外的监督和姿势预测过程。另一个比较有效的是注意力方法，它可以模仿人类的视觉信号处理机制，并且不需要使用具体的语义特征，相比结合姿势的方法，能在一定程度上减轻工作。Song等人[11]利用二值掩膜设计了一种对比注意力模型来分别学习身体和背景区域的特征。Li等人[12]提出了注意力融合卷积神经网络模型，从而优化图像未对准的情况。注意力机制方法对于行人遮挡和分辨率低等问题有一定的帮助，但在识别的过程中可能丢失一些比较重要的数据。

为了解决上述问题，本文将特征空间分割和注意力机制相结合，提出了更加有效的特征空间分割方式，针对局部分支引入了注意力机制策略。首先，改进了以往空间分割的方法，对切分为多个水平区域的特征图进行阶梯型特征提取，与之前的强制水平分割相比，阶梯型特征提取能够关注更多的边缘信息，加强局部特征之间的联系。其次，在多个局部分支中引入了分支注意力，按重要性给每一个局部分支分配相应的权重，从而在更注重包含比较多重要信息的分支同时削弱无关信息的关注程度。最后，在每一个局部分支后引入卷积核注意力、通道注意力和空间注意力，结合多类注意力机制能够获得更好的判别特征。

2 特征空间的分割及注意力机制的引入

2.1 网络结构

本文网络结构如图1所示，该网络以ResNet-50为骨干网络，且在最后一层卷积conv5_x前划分成两个独立的分支。第1个分支延续原始ResNet-50相同的体系结构，提取行人图像的全局特征。第2个分支提取多粒度的局部注意力特征。

对于第2个分支，在conv5_x中设置卷积步长为1，从而增大网络输出的特征图来获得更多的特征信息。随后又划分为3个独立的分支：分支1、分支2、分支3。将调整为384×128大小的图像输入到网络后，在该分支经过conv5_x得到大小为24×8的整体行人特征图，其维度为2048维。在分支1利用所得的整体行人特征图来获取完整图像特征信息。在分支2将整体行人特征图输入到阶梯型分支注意力模块(Stepped Branch Attention Module, SBAM)进行阶梯型特征分块，得到P个不同权重的局部特征。在这里设置P=5，表示划分得到5个局部区域，其中每一个分块区域的大小为12×8。在

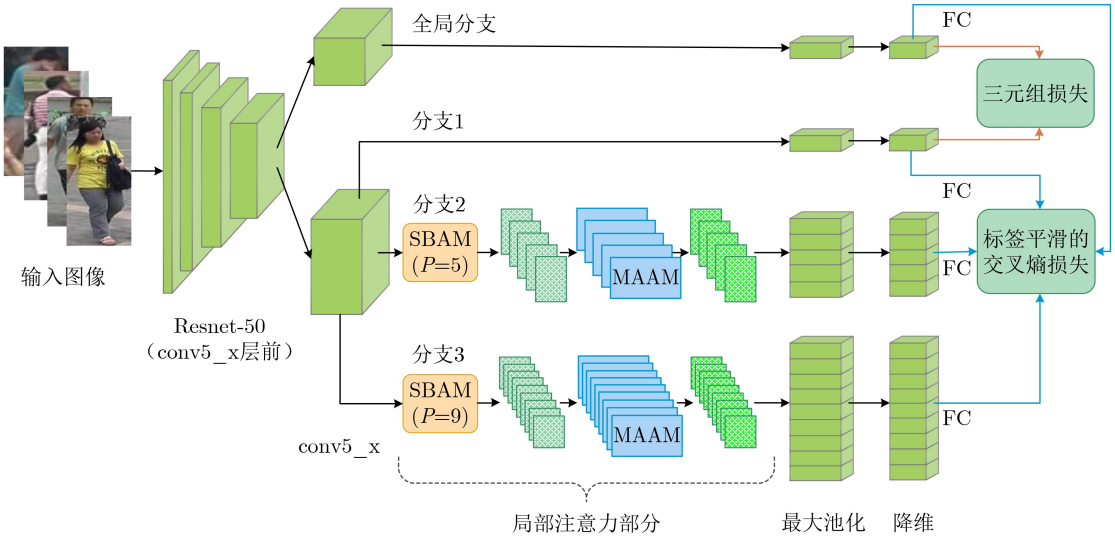


图 1 阶梯型局部分支注意力网络（Stepped Local Branch Attention Network，SLANet）结构

分支3将整体行人特征图输入到SBAM模块进行阶梯型特征分块，得到P个不同权重的局部特征，设置P=9，每一个分块区域的大小为8×8。对分支2和分支3得到的14个局部特征都单独送入到多尺度自适应注意力模块(Multi-scale Adaptive Attention Module, MAAM)，从而得到融合了卷积核注意力、通道注意力和空间注意力的局部特征图，所有经过MAAM模块的特征图尺寸不变。然后对3个分支得到的15个特征图分别进行最大池化和降维处理，得到15个265维特征向量。

在训练阶段，将所有降至256维的特征分别送入全连接层后使用标签平滑的交叉熵损失进行分类学习，对全局分支和分支1得到的256维特征使用三元组损失进行度量学习。在测试阶段，将所有降至256维的特征连接起来作为最终特征，从而充分结合全局和局部信息来获得最强大的识别能力。

2.2 阶梯型分支注意力模块

特征空间分割是一种比较有效的局部特征提取方法，但现如今许多特征空间分割的方法都是水平均匀分块，这种方式能够学习到人体不同区域的差异，但块与块之间的边缘信息容易被忽略。如图2所示，对于尺寸调整后的原始行人图像，若是把它划分成左侧图中的4个水平块，第1块和第2块会将该行人衣服上的白色标志分离，同时第2块和第3块会将红色斜跨包分成两部分。这样则会忽略以及破坏块与块边缘的重要信息，导致在对每一块进行单独分类时达不到最终期待的效果。于是本文提出了一种按特定数量以阶梯型选取图像块从而得到多个分支的方法。对于分支2首先将原始完整行人图像均匀分成8个水平块，最初以第1块为起始块，每4块为一个整体作为一个局部区域，随后以步长为1往下更改起始块进行阶梯型分块，最终得到5个局部分区。可以看到，局部分区(a)和(b)中都包含有衣服上完整的白色标志，局部分区(b)和(c)中都包含有完整的斜挎包信息。该方法更注重图像水平块间的内在联系，能够避免特征学习过程中某些重要信息的丢失。

以往基于特征空间分割的方法中每个水平分块都享有一样的权重，背包等细节信息不能有效地凸显。本文给含有较多重要信息的分支分配比较大的权重，使模型聚焦到具有强分辨力的分支上。对于网络的第2个分支，输入一张图像经过conv5_x后得到特征图$\boldsymbol{F}\in\mathbb{R}^{C\times H\times W}$，接下来在分支2和分支3将特征图$\boldsymbol{F}$输入SBAM模块。在SBAM模块中，首先进行阶梯型分块。对于分支2，将$\boldsymbol{F}$分割成8个水平部分，每4个部分为一组得到一个局部区域，局部区域的起始块从第1块开始以步长1向下移动，最终能提取5个局部区域，其中每一个特征区域为$\boldsymbol{F}_i\in\mathbb{R}^{C\times(H/2)\times W}\ (i=1,2,3,4,5)$。然后对每一个$\boldsymbol{F}_i$进行注意力聚焦，首先对其在空间维度上进行压缩，计算方法为

$$\boldsymbol{F}_i'=\mathrm{FC}_2\left(\mathrm{FC}_1\left(\mathrm{Avg}_{\mathrm{s}}\left(\boldsymbol{F}_i\right)\right)\right)+\mathrm{FC}_2\left(\mathrm{FC}_1\left(\mathrm{Max}_{\mathrm{s}}\left(\boldsymbol{F}_i\right)\right)\right) \tag{1}$$

其中，$\mathrm{Avg}_{\mathrm{s}}$和$\mathrm{Max}_{\mathrm{s}}$分别为对输入数据在空间维度上进行平均池化和最大池化，压缩后得到两个1维矢量。FC_1和FC_2作为共享部分对两个矢量在通道上进行压缩和恢复，最后相加融合后得到$\boldsymbol{F}_i'\in\mathbb{R}^{C\times1\times1}$。为了给每个局部区域赋予权重，再对$\boldsymbol{F}_i'$在通道维度进行压缩聚焦，表示为

$$\boldsymbol{s}_i=\mathrm{Sum}_{\mathrm{c}}\left(\boldsymbol{F}_i'\right) \tag{2}$$

$$\boldsymbol{m}_{\mathrm{i}}=\mathrm{Max}_{\mathrm{c}}\left(\boldsymbol{F}_i'\right) \tag{3}$$

其中，$\mathrm{Sum}_{\mathrm{c}}$和$\mathrm{Max}_{\mathrm{c}}$分别为对输入数据在通道维度上进行求取总和以及最大值，最终得到$\boldsymbol{s}_i\in\mathbb{R}^{1\times1\times1}$和$\boldsymbol{m}_i\in\mathbb{R}^{1\times1\times1}$。根据计算各个局部区域的$\boldsymbol{s}_i$和$\boldsymbol{m}_i$能求解出某一分支在所有分支所占的比例，计算公式为

$$\begin{aligned}V_i&=\lambda\left(\frac{\boldsymbol{s}_i}{\sum_{i=0}^{P}\boldsymbol{s}_i}+\frac{\boldsymbol{m}_i}{\sum_{i=0}^{P}\boldsymbol{m}_i}\right)\\&=\lambda\left(\mathrm{F}_{\mathrm{sum}}\left(\boldsymbol{s}_i\right)+\mathrm{F}_{\mathrm{max}}\left(\boldsymbol{m}_i\right)\right)\end{aligned} \tag{4}$$

其中，$\sum_{i=0}^{P}\boldsymbol{s}_i$和$\sum_{\mathrm{i}=0}^{P}\boldsymbol{m}_i$分别为$P$个分支的$\boldsymbol{s}_i$和$\boldsymbol{m}_i$总值，$\lambda$为已设定好的参数，根据实验设定$\lambda=6$。

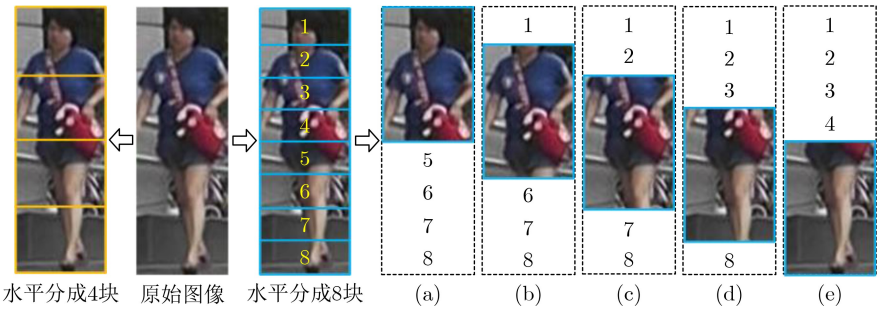


图 2 阶梯型分块方式

最后通过sigmoid函数将占比值调整为0到1之间，将原始局部区域$\boldsymbol{F}_i$与调整好的占比值相乘得到该局部区域的更新结果$\boldsymbol{S}_i \in \mathbb{R}^{C\times(H/2)\times W}$，即

$$\boldsymbol{S}_i = \boldsymbol{F}_i \times \text{sigmoid}\left(V_i\right) \tag{5}$$

2.3 多尺度自适应注意力模块

注意力机制类似于人类观察物体时将视线聚焦于重要区域，其目标是增强模型对图像重要区域的关注从而获得更有区分力的特征信息。本文结合了SKNet[13]自适应感受野大小和CBAM[14]多注意力融合的思想，在网络最后一个卷积层和池化层间引入多尺度自适应注意力模块(MAAM)。

对于输入MAAM模块的局部特征图$\boldsymbol{S}\in\mathbb{R}^{C\times(H/2)\times W}$，我们将其复制为$\boldsymbol{S}_1$和$\boldsymbol{S}_2$来分别完成两个任务。第1部分任务为在自适应调整感受野大小的同时将通道注意力融入其中。首先进行式(6)计算

$$\boldsymbol{S}_{13} = f^{3\times3}\left(\boldsymbol{S}_1\right) + f^{5\times5}\left(\boldsymbol{S}_1\right) = \boldsymbol{S}_{11} + \boldsymbol{S}_{12} \tag{6}$$

其中，$f^{3\times3}$和$f^{5\times5}$分别为对$\boldsymbol{S}_1$进行3×3和5×5大小卷积核的卷积操作，为了不增加计算量，5×5的卷积核是使用膨胀系数为2的3×3卷积核来代替的。对得到的两组不同尺度的特征相加融合后再进行特征聚焦，计算方式为

$$\begin{aligned}\boldsymbol{V} =& \text{FC}_3\left(\text{FC}_1\left(\text{Avg}_\text{s}\left(\boldsymbol{S}_{13}\right)\right)\right)\\ &+ \text{FC}_3\left(\text{FC}_1\left(\text{Max}_\text{s}\left(\boldsymbol{S}_{13}\right)\right)\right)\end{aligned} \tag{7}$$

其中，Avg_s和Max_s分别为对输入数据在空间维度上进行平均池化和最大池化，压缩后得到两个1维矢量。FC_1和FC_3为先降维再升维的两层全连接层，其中在FC_3分成两个矢量以便之后自适应地选择不同空间尺度的信息，最后相加融合得到$\boldsymbol{V}\in\mathbb{R}^{2\times C\times1\times1}$。$\boldsymbol{V}$经过softmax后输出两个权重矩阵$\boldsymbol{V}_a$和$\boldsymbol{V}_b$，其中$\boldsymbol{V}_b$为冗余矩阵，在这里$\boldsymbol{V}_a+\boldsymbol{V}_b=1$。使用$\boldsymbol{V}_a$和$\boldsymbol{V}_b$两个权重矩阵对$\boldsymbol{S}_{11}$和$\boldsymbol{S}_{12}$进行加权操作，从而利用各种卷积核的注意力权重得到最终的输出特征$\boldsymbol{Y}_{13}\in\mathbb{R}^{C\times(H/2)\times W}$，即

$$\boldsymbol{Y}_{13} = \boldsymbol{V}_a \times \boldsymbol{S}_{11} + \boldsymbol{V}_b \times \boldsymbol{S}_{12} \tag{8}$$

由于$\boldsymbol{V}$中已经含有行人图像的通道注意力信息，可利用该矩阵对$\boldsymbol{Y}_{13}$融入通道注意力，即

$$\boldsymbol{Y}_1 = \boldsymbol{Y}_{13} \times \text{sigmoid}\left(\text{Sum}_\text{v}\left(\boldsymbol{V}\right)\right) \tag{9}$$

其中，Sum_v为对两个矢量进行元素相加，将每个通道的权重压缩到一个特征向量中。

第2部分任务为对输入特征融入空间注意力，首先提取出位置信息特征，表示为

$$\boldsymbol{S}_{23} = \left[\text{Avg}_\text{c}\left(\boldsymbol{S}_2\right);\text{Max}_\text{c}\left(\boldsymbol{S}_2\right)\right] \tag{10}$$

其中，Avg_c和Max_c分别为对输入数据在通道维度上进行平均池化和最大池化，得到两个不同的特征描述$\boldsymbol{S}_{21}\in\mathbb{R}^{1\times(H/2)\times W}$，$\boldsymbol{S}_{22}\in\mathbb{R}^{1\times(H/2)\times W}$，将其合并后得到$\boldsymbol{S}_{23}\in\mathbb{R}^{2\times(H/2)\times W}$。之后通过卷积将维度重新恢复成1，且卷积操作也使用自适应卷积核大小的方式，即

$$\boldsymbol{Z}_i = \text{Avg}_\text{s}\left(f^{i\times i}\left(\boldsymbol{S}_{23}\right)\right) + \text{Max}_\text{s}\left(f^{i\times i}\left(\boldsymbol{S}_{23}\right)\right), i = 3,5 \tag{11}$$

其中，$f^{i\times i}$表示进行卷积核大小为$i\times i$的卷积操作，其中i为3或5。通过不同的卷积操作可计算出

$$\boldsymbol{S}_{24} = f^{3\times3}\left(\boldsymbol{S}_{23}\right) \tag{12}$$

$$\boldsymbol{S}_{25} = f^{5\times5}\left(\boldsymbol{S}_{23}\right) \tag{13}$$

求得的$\boldsymbol{Z}_3\in\mathbb{R}^{1\times1\times1}$和$\boldsymbol{Z}_5\in\mathbb{R}^{1\times1\times1}$经过softmax后输出两个权重矩阵$\boldsymbol{Z}_\text{a}$和$\boldsymbol{Z}_\text{b}$，然后分别与$\boldsymbol{S}_{24}$和$\boldsymbol{S}_{25}$相乘。求和后经过sigmoid激活函数再与原始特征$\boldsymbol{S}_2$相乘，表示为

$$\boldsymbol{Y}_2 = \boldsymbol{S}_2 \times \text{sigmoid}\left(\boldsymbol{Z}_a \times \boldsymbol{S}_{24} + \boldsymbol{Z}_b \times \boldsymbol{S}_{25}\right) \tag{14}$$

最后将两个部分得到的结果相加融合，即

$$\boldsymbol{Y} = \boldsymbol{Y}_1 + \boldsymbol{Y}_2 \tag{15}$$

2.4 损失函数

行人重识别网络使用的损失函数一般分为分类损失和度量损失两类。本文使用标签平滑[15]的交叉熵损失对全局和局部特征进行分类，同时使用难样本采样三元组损失[16]对全局特征实现度量学习，选择更难的样本去训练网络能提高网络的泛化能力，根据全局特征和局部特征的差异性联合使用两种损失能够获得更优的性能[8]。将行人图像输入到网络后，负交叉熵损失公式可表示为

$$L_{id} = -\sum_{i=1}^{N} q_i \lg p_i \tag{16}$$

其中，N为类别总数，p_i为网络预测类别为i的概率，q_i为标签。由于训练集中每张图像只有1个标签，真实标签类别对应的q_i为1，其余都是0，但这样网络会太倾向真实类别而造成过拟合。为了减小过拟合而增加泛化性能，本文引入标签平滑的思想，对所有标签数据进行了平滑处理，将q_i设置为

$$q_i = \begin{cases} 1-\dfrac{N-1}{N}\varepsilon, i = y \\ \varepsilon/N, \qquad i \neq y \end{cases} \tag{17}$$

其中，ε为一个较小的超参数，y代表图像的真实标签，N为类别总数。

三元组损失优化的目标是拉近相同行人类别的图像距离，同时拉远不同行人类别的图像距离。本文使用更有效的难样本采样三元组损失，对于每一个训练批次，随机挑选P个行人身份，每个行人随机挑选K张图像。之后对于训练批次中的每一张图

像a，选择与之距离最远的正样本和距离最近的负样本同a组成一个三元组。公式为

$$L_{\text{triplet}} = -\sum_{i=1}^{P}\sum_{a=1}^{K}\left[\alpha + \max_{p=1,\cdots,K}\left\|f_a^{(i)} - f_p^{(i)}\right\|_2 - \min_{\substack{n=1,\cdots,K\\ j=1,\cdots,P\\ j\neq i}}\left\|f_a^{(i)} - f_n^{(j)}\right\|_2\right]_+ \tag{18}$$

其中，$f_a^{(i)}$, $f_p^{(i)}$和$f_n^{(j)}$分别是从锚样本、正样本和负样本中提取的特征，α是用于控制样本内和样本间距离差异的阈值参数，P为每批次中行人的个数，K为每批次中每个行人的图像数量。

3　实验结果与分析

3.1　数据集和评价指标

为了验证本文方法的有效性，在Market-1501[17]，DukeMTMC-reID[18]和CUHK03[19]3个数据集上进行了验证。本文使用平均精度均值(mean Average Precision, mAP)和累计匹配特征曲线(Cumulative Match Characteristic, CMC)两种评价指标来评价算法的性能。

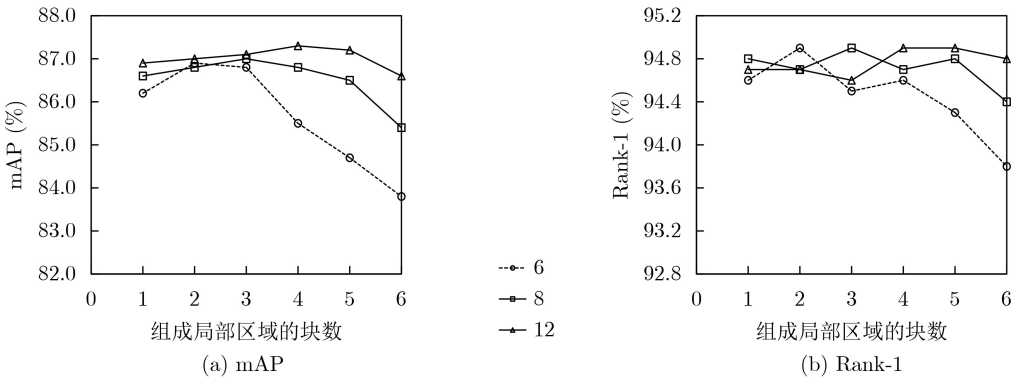
3.2　实验环境及参数设置

本文实验在配有硬件Intel(R) Xeon(R) W-2155 CPU@3.30GHz和NVIDIA RTX 2080Ti GPU，操作系统为Ubuntu 19.10的设备上进行，并基于python 3.7.5编程语言和Pytorch 1.3.0深度学习框架来完成算法实现。实验数据集调整图像大小为384×128，采用随机翻转、随机裁剪和随机擦除的数据增强方法。训练批次大小为N=P×K，其中P为每批次中行人的个数，K为每批次中每个行人的图像数量。设置P为8，K为4，批次大小为32。在训练过程中使用Adam优化器来优化网络参数，选择Warmup预热学习率策略，设置初始学习率为0.0004，权重衰减系数为0.0005，网络迭代次数为400次，且分别在第220次迭代和第320次迭代将学习率降为之前的0.1。

3.3　分块策略的有效性

在SBAM模块中，我们将行人图像均匀分成M个水平块，最初以第1块为起始块，每N块为一个整体作为一个局部区域，起始块位置以步长为1向下移动进行阶梯型分块。为了确定M和N的取值，从SLANet网络中删去分支3并移除MAAM模块，在分支2单分支中以控制变量法进行分块实验。由于输入SBAM模块的特征图大小为24×8，为了方便实验，分别以4, 3, 2的高度划分水平块，则对应的总块数分别为6块、8块和12块。图3给出了在3种分块数量下组成局部区域的块数取不同值时的实验结果，该实验在Market-1501数据集上进行。从实验数据可以看出，当分块数分别为6块、8块和12块时，组成局部区域的每组块数分别为2，3，4时性能较优。且这3种块数分组方式都比每组块数为1时效果好，验证了该方案下的阶梯型分块相比以往普通的分块方式的有效性。

文献[8]提出了一种多粒度的特征学习策略。在这里，我们将完整图像作为1级粒度，将尺寸水平裁剪成原来的1/2和1/3的图像分别作为2级粒度和3级粒度。对于输入SBAM模块的大小为24×8的特征图，粒度等级为2的情况有：(1)将其划分成6个水平块，每3个水平块为一组；(2)将其划分成8个水平块，每4个水平块为一组；(3)将其划分成12个水平块，每6个水平块为一组。粒度等级为3的情况有：(1)将其划分成6个水平块，每2个水平块为一组；(2)将其划分成12个水平块，每4个水平块为一组。同时将划分成8个水平块，每3个水平块为一组得到的9×8的局部区域近似为粒度等级3。最后将全局特征和这两种粒度作为多分支的方式进行组合实验，从表1展示的结果可以看出，8_4 + 12_4分块方式的性能效果最优，mAP和Rank-1分别达到了87.7%和95.3%。表2展示了单独送入SBAM模块的局部分支个数对行人重识别任务性能的影响。2个局部分支的效果要比只有全局分支和1个局部分支的效果都要好，表示在一定程度上多粒度局部特



(a) mAP　　(b) Rank-1

图 3 不同分块方式的对比结果

征方法能够提升行人重识别任务性能。但当局部分支个数增加为3个时，不仅计算复杂度增大，效果也没有明显增强。这是因为前两个分支间的相互作用已经对行人图像达到了一个比较好的局部特征判别能力，并且在分支数过多时，会存在较多的不交叉重叠区域，这些区域单独计算损失，完成这些额外任务在降低速度的同时对性能不会有显著提升。

3.4 联合训练的有效性

表3展示了是否使用两个模块联合训练的性能效果对比，该实验在Market-1501数据集上进行。其中baseline是只使用一个全局分支的情况，在此基础上增加SBAM模块可以明显提高判别性能，mAP和Rank-1分别提升了10.4%和4.8%。将每一个局部区域单独送入到MAAM模块后，mAP和Rank-1分别继续提升了0.4%和0.3%，MAAM模块只是对每个局部特征进行处理，通过对其融合多类注意力机制来辅助完成局部区域的识别，因此相对提升幅度不大。实验证明本文提出的模块在行人重识别任务上能获得优异的性能。

表 1 多粒度分块方法比较(%)

分块方式	mAP	Rank-1	Rank-5	Rank-10
6_3 + 6_2	87.6	95.0	98.4	99.1
6_3 + 8_3	87.1	94.9	98.3	98.9
6_3 + 12_4	87.3	94.9	98.4	99.0
8_4 + 6_2	87.4	95.1	98.4	99.0
8_4 + 8_3	87.1	94.9	98.3	99.0
8_4 + 12_4	**87.7**	**95.3**	**98.5**	**99.2**
12_6 + 6_2	87.5	95.0	98.4	99.1
12_6 + 8_3	**87.7**	95.2	98.4	99.1
12_6 + 12_4	87.4	95.2	98.4	**99.2**

表 2 阶梯型多分支的有效性(%)

分块方式	mAP	Rank-1	Rank-5	Rank-10
全局特征	77.3	90.5	96.4	97.8
8_4	86.8	94.7	98.4	99.0
12_4	87.3	94.9	98.4	99.1
8_4 + 12_4	87.7	**95.3**	**98.5**	**99.2**
8_4 + 12_4 + 8_2	**87.8**	95.1	98.4	**99.2**
8_4 + 12_4 + 12_3	87.3	94.9	98.4	99.1

表 3 联合训练的有效性(%)

方法	mAP	Rank-1	Rank-5	Rank-10
baseline	77.3	90.5	96.4	97.8
+ SBAM	87.7	95.3	98.5	99.2
+ SBAM + MAAM	88.1	95.6	98.6	99.2

3.5 与现有方法的比较

为了进一步说明本文模型的有效性，将其与多种主流行人重识别算法相比，实验数据如表4、表5和表6所示。从表中数据可以看出，本文提出的方法在Market-1501数据集上的mAP和Rank-1分别为88.1%和95.6%，在DukeMTMC-reID数据集上的mAP和Rank-1分别为80.0%和88.6%。在DukeMTMC-reID上的相对效果较低，主要因为该数据集中不同行人之间的相似性很高且同一行人会存在遮挡、行人重叠等情况，挑战难度较大。同时在CUHK03-Labeled数据集上的mAP和Rank-1分别为78.4%和80.7%，在CUHK03-Detected数据集上

表 4 在Market-1501数据集上的性能比较(%)

方法	mAP	Rank-1	Rank-5	Rank-10
SVDNet[20]	62.1	82.3	92.3	95.2
HA-CNN[12]	75.7	91.2	–	–
PCB[7]	77.4	92.3	97.2	98.2
PCB+RPP[7]	81.6	93.8	97.5	98.5
HPM[21]	82.7	94.2	97.5	98.5
MHN[22]	85.0	95.1	98.1	98.9
SLANet(本文)	88.1	95.6	98.6	99.2
SLANet(+RK)	94.6	96.5	98.1	98.8

表 5 在DukeMTMC-reID数据集上的性能比较(%)

方法	mAP	Rank-1	Rank-5	Rank-10
SVDNet[20]	56.8	76.7	86.4	89.9
HA-CNN[12]	63.8	80.5	–	–
PCB[7]	66.1	81.7	89.7	91.9
PCB+RPP[7]	69.2	83.3	90.5	92.5
HPM[21]	74.3	86.6	–	–
MHN[22]	77.2	89.1	94.6	96.2
SLANet(本文)	80.0	88.6	95.0	96.7
SLANet(+RK)	90.0	91.7	95.3	96.5

表 6 在CUHK03数据集上的性能比较(%)

方法	Labeled		Detected	
	mAP	Rank-1	mAP	Rank-1
SVDNet[20]	37.8	40.9	37.3	41.5
HA-CNN[12]	41.0	44.4	38.6	41.7
MLFN[23]	49.2	54.7	47.8	52.8
PCB+RPP[7]	–	–	57.5	63.7
MGN[8]	67.4	68.0	66.0	66.8
MHN[22]	72.4	77.2	65.4	71.7
SLANet(本文)	78.4	80.7	75.2	78.6
SLANet(+RK)	88.6	87.1	85.5	83.8

的mAP和Rank-1分别为75.2%和78.6%，后者使用DPM自动检测标注方法，存在更多遮挡、错位和人体部分缺失等问题从而更接近实际情况。另外在Market-1501数据集上使用重排序[24]后，本文方法的mAP可以达到94.6%，Rank-1可以达到96.5%。本文方法在3个数据集上的mAP和Rank-1相比于其他方法均有明显提升，表明本文方法对行人重识别任务的性能提升是有效的。

图4展示了本文方法SLANet在Market-1501数据集上得到的MAAM模块后输出特征的可视化结果，该可视化结果展示了模型对相关区域的关注程度。从图中可以看出相比CBAM注意力模块，SBAM和MAAM模块的联合更加强调了局部区域的局部注意力，能够带来优秀的行人重识别能力。

4　结束语

对于行人重识别网络模型，本文提出了有效的阶梯型局部分支注意力模块和多尺度自适应注意力模块。阶梯型分支注意力模块以阶梯型划分图像区域从而加强局部特征之间的联系，同时以分支注意力的方式使网络关注更重要的分支从而能提高行人重识别的准确度。多尺度自适应注意力模块对输入特征进行自适应感受野大小的处理，并对特征融入通道注意力和空间注意力，使网络提取的局部特征更具有鲁棒性。同时设计了合理的多分支网络结构，使用多粒度的方法将全局特征和局部特征进行融合，得到更具判别能力的融合特征表示。实验结果表明本文算法在3大主流数据集上均取得了较好的效果。如何高效地提取更有效的局部特征表示并寻求准确率和时间开销的平衡仍然是今后需要重点研究探讨的。

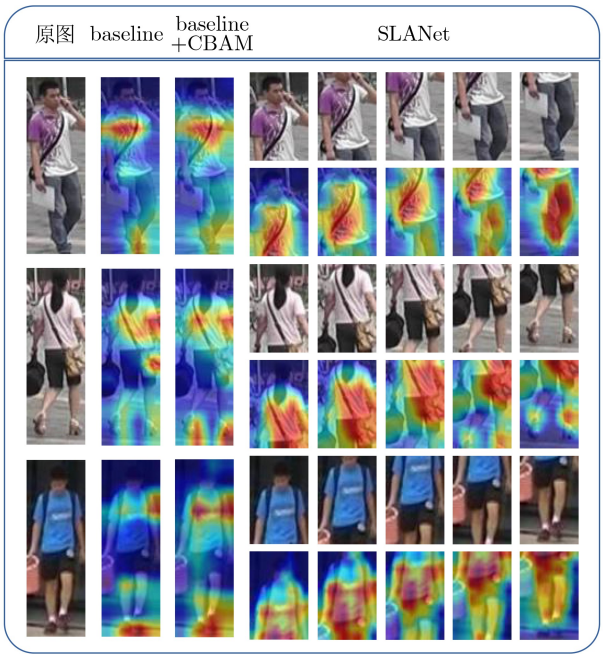


图 4 可视化结果

参 考 文 献

[1] 周智恒, 刘楷怡, 黄俊楚, 等. 一种基于等距度量学习策略的行人重识别改进算法[J]. 电子与信息学报, 2019, 41(2): 477–483. doi: 10.11999/JEIT180336.
ZHOU Zhiheng, LIU Kaiyi, HUANG Junchu, *et al*. Improved metric learning algorithm for person re-identification based on equidistance[J]. *Journal of Electronics & Information Technology*, 2019, 41(2): 477–483. doi: 10.11999/JEIT180336.

[2] 陈莹, 许潇月. 基于双向参考集矩阵度量学习的行人再识别[J]. 电子与信息学报, 2020, 42(2): 394–402. doi: 10.11999/JEIT190159.
CHEN Ying and XU Xiaoyue. Matrix metric learning for person re-identification based on bidirectional reference set[J]. *Journal of Electronics & Information Technology*, 2020, 42(2): 394–402. doi: 10.11999/JEIT190159.

[3] HE Botao and YU Shaohua. Ring-push metric learning for person reidentification[J]. *Journal of Electronic Imaging*, 2017, 26(3): 033005. doi: 10.1117/1.JEI.26.3.033005.

[4] GOODFELLOW I J, POUGET-ABADIE J, MIRZA M, *et al*. Generative adversarial nets[C]. Proceedings of the 27th International Conference on Neural Information Processing Systems, Montreal, Canada, 2014: 2672–2680.

[5] 杨锋, 许玉, 尹梦晓, 等. 基于深度学习的行人重识别综述[J]. 计算机应用, 2020, 40(5): 1243–1252. doi: 10.11772/j.issn.1001-9081.2019091703.
YANG Feng, XU Yu, YIN Mengxiao, *et al*. Review on deep learning-based pedestrian re-identification[J]. *Journal of Computer Applications*, 2020, 40(5): 1243–1252. doi: 10.11772/j.issn.1001-9081.2019091703.

[6] ZHENG Zhedong, ZHENG Liang, and YANG Yi. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro[C]. Proceedings of 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 3774–3782. doi: 10.1109/ICCV.2017.405.

[7] SUN Yifan, ZHENG Liang, YANG Yi, *et al*. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline)[C]. Proceedings of the 15th European Conference on Computer Vision, Munich, Germany, 2018: 480–496. doi: 10.1007/978-3-030-01225-0_30.

[8] WANG Guanshuo, YUAN Yufeng, CHEN Xiong, *et al*. Learning discriminative features with multiple granularities for person Re-identification[C]. Proceedings of the 26th ACM international conference on Multimedia, Seoul, Republic of Korea, 2018: 274–282. doi: 10.1145/3240508.

3240552.

[9] ZHAO Haiyu, TIAN Maoqing, SUN Shuyang, *et al*. Spindle Net: Person re-identification with human body region guided feature decomposition and fusion[C]. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 907–915. doi: 10.1109/CVPR.2017.103.

[10] MIAO Jiaxu, WU Yu, LIU Ping, *et al*. Pose-guided feature alignment for occluded person re-identification[C]. Proceedings of 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Korea (South), 2019: 542–551. doi: 10.1109/ICCV.2019.00063.

[11] SONG Chunfeng, HUANG Yan, OUYANG Wanli, *et al*. Mask-guided contrastive attention model for person re-identification[C]. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 1179–1188. doi: 10.1109/CVPR.2018.00129.

[12] LI Wei, ZHU Xiatian, and GONG Shaogang. Harmonious attention network for person re-identification[C]. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 2285–2294. doi: 10.1109/CVPR.2018.00243.

[13] LI Xiang, WANG Wenhai, HU Xiaolin, *et al*. Selective kernel networks[C]. Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 510–519. doi: 10.1109/CVPR.2019.00060.

[14] WOO S, PARK J, LEE J Y, *et al*. CBAM: Convolutional block attention module[C]. Proceedings of the 15th European Conference on Computer Vision, Munich, Germany, 2018: 3–19. doi: 10.1007/978-3-030-01234-2_1.

[15] LUO Hao, GU Youzhi, LIAO Xingyu, *et al*. Bag of tricks and a strong baseline for deep person re-identification[C]. Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, USA, 2019: 1487–1495. doi: 10.1109/CVPRW.2019.00190.

[16] HERMANS A, BEYER L, and LEIBE B. In defense of the triplet loss for person re-identification[EB/OL]. https://arxiv.org/abs/1703.07737, 2017.

[17] ZHENG Liang, SHEN Liyue, TIAN Lu, *et al*. Scalable person re-identification: A benchmark[C]. Proceedings of 2015 IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 1116–1124. doi: 10.1109/ICCV.2015.133.

[18] RISTANI E, SOLERA F, ZOU R, *et al*. Performance measures and a data set for multi-target, multi-camera tracking[C]. Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 2016: 17–35. doi: 10.1007/978-3-319-48881-3_2.

[19] LI Wei, ZHAO Rui, XIAO Tong, *et al*. DeepReID: Deep filter pairing neural network for person re-identification[C]. Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA, 2014: 152–159. doi: 10.1109/CVPR.2014.27.

[20] SUN Yifan, ZHENG Liang, DENG Weijian, *et al*. SVDNet for pedestrian retrieval[C]. Proceedings of 2017 IEEE International Conference on Computer Vision, Venice, Italy, 2017: 3820–3828. doi: 10.1109/ICCV.2017.410.

[21] FU Yang, WEI Yunchao, ZHOU Yuqian, *et al*. Horizontal pyramid matching for person re-identification[C]. Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, USA, 2019: 8295–8302. doi: 10.1609/aaai.v33i01.33018295.

[22] CHEN Binghui, DENG Weihong, and HU Jiani. Mixed high-order attention network for person re-identification[C]. Proceedings of 2019 IEEE/CVF International Conference on Computer Vision, Seoul, Korea (South), 2019: 371–381. doi: 10.1109/ICCV.2019.00046. doi: 6.

[23] CHANG Xiaobin, HOSPEDALES T M, and XIANG Tao. Multi-level factorisation net for person re-identification[C]. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 2109–2118. doi: 10.1109/CVPR.2018.00225.

[24] ZHONG Zhun, ZHENG Liang, CAO Donglin, *et al*. Re-ranking person re-identification with k-reciprocal encoding[C]. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 3652–3661. doi: 10.1109/CVPR.2017.389.

石跃祥：男，1964年生，教授，硕士生导师，研究方向为图像处理与智能系统.

周 玥：女，1996年生，硕士生，研究方向为图形图像处理与行人重识别.

责任编辑：陈 倩