

DOI:10.13364/j.issn.1672-6510.20220107

## 基于 HSA 注意力模块和关联规则的行人属性识别算法

孙志伟<sup>1</sup>, 师亚涛<sup>1</sup>, 马永军<sup>1</sup>, 闫潇宁<sup>2</sup>, 许能华<sup>2</sup>

(1. 天津科技大学人工智能学院, 天津 300457; 2. 深圳市安软科技股份有限公司, 深圳 518131)

**摘要:** 针对先前研究工作对行人属性相关性分析不足和行人图像中细粒度属性特征难以捕获的问题, 首先基于关联规则寻找属性之间的相关性, 然后根据相关性改变网络结构, 提升有较强相关性属性的准确率并嵌入改进的 HSA (hierarchy split attention) 注意力模块, 将特征图中的潜在信息进行深度挖掘. HSA 注意力模块将特征图进行分组和在子特征图之间增加通道交互操作, 整合后的特征图输入挤压激励 (squeeze and excitation, SE) 模块中, 提取图像在通道上的信息. 在 PA100K、Market-1501 和 PETA 数据集上的实验结果表明, 本算法的精确率、召回率、 $F_1$  值与其他算法相当, 但是准确率有较大提升.

**关键词:** 图像处理; 行人属性识别; 注意力模块; 协同训练; 残差网络

中图分类号: TP301.6 文献标志码: A 文章编号: 1672-6510(2023)01-0061-09

### Pedestrian Attribute Recognition Algorithm Based on HSA Attention Module and Association Rules

SUN Zhiwei<sup>1</sup>, SHI Yatao<sup>1</sup>, MA Yongjun<sup>1</sup>, YAN Xiaoning<sup>2</sup>, XU Nenghua<sup>2</sup>

(1. College of Artificial Intelligence, Tianjin University of Science & Technology, Tianjin 300457, China;  
2. Shenzhen Softsz Co., Ltd., Shenzhen 518131, China)

**Abstract:** In view of insufficient correlation analysis of pedestrian attributes in previous research and the difficulty in capturing fine-grained attribute features in pedestrian images, in our present study we first find the correlation between attributes based on association rules, then change the network structure according to the correlation, improve the accuracy of highly correlated attributes, and embed the improved hierarchy split attention (HSA) module to deeply mine the latent information in the feature map. HSA module groups feature maps and adds channel interaction between sub feature maps, inputs the integrated feature map into the squeeze and excitation (SE) module, and extracts the information of the image on the channel. The experimental results on PA100k, Market-1501 and PETA data sets show that this algorithm has little difference with other algorithms in precision, recall and  $F_1$ , but has a great improvement in accuracy indicators.

**Key words:** image processing; pedestrian attribute recognition; attention module; collaborative training; residual network

行人属性识别是计算机视觉领域的一项重要任务, 它在视频监控中扮演着重要的角色<sup>[1]</sup>, 目的是识别图片中行人的属性, 比如年龄、性别、上衣穿搭、上衣颜色等; 在其他视觉任务中也有较强的辅助作用, 如行人重识别<sup>[2]</sup>、行人跟踪<sup>[3]</sup>、行人检测<sup>[4]</sup>.

早期的行人属性识别算法主要用的是手工设计

的低级特征, 如 HOG<sup>[5]</sup>、SIFT<sup>[6]</sup>、LibSVM<sup>[7]</sup>、CRF<sup>[8]</sup>等. 由于获取的是低级特征, 这些算法在实际应用中的效果不理想. 近年来, 深度学习通过多层非线性变换可以提取更加丰富的特征, 也出现了很多基于深度学习的行人属性识别算法, 如基于全局特征<sup>[9-10]</sup>、基于局部特征<sup>[11-12]</sup>、基于注意力机制<sup>[13-14]</sup>等的算法.

收稿日期: 2022-04-23; 修回日期: 2022-07-14

基金项目: 国家自然科学基金资助项目(61603273); 天津市自然科学基金资助项目(18JQNJC69500)

作者简介: 孙志伟(1973—), 男, 河北保定人, 副教授, zhwsun@tust.edu.cn

Li 等<sup>[9]</sup>提出的 DeepMAR 网络将整幅行人图像作为输入,对所有属性进行预测. Sudowe 等<sup>[10]</sup>提出的基于卷积网络的多分支分类分层方法 ACN(a jointly-trained holistic CNN model)考虑全局图像,联合学习行人所有属性进行预测. 但是这类算法缺乏对图像的细粒度识别. Li 等<sup>[11]</sup>提出了 PGDM(pose guided deep model)模型,利用行人姿态估计模型获取图像关键点,提取局部图像,获得局部特征. Liu 等<sup>[12]</sup>提出了一种定位引导网络 LGNet(localization guided network),将行人局部特征与全局特征融合,可以对不同属性对应的区域进行定位. 但是这类算法依赖于辅助模型,获取到的局部特征和全局特征可能有效性不足. Liu 等<sup>[13]</sup>提出的 HydraPlus-Net 将注意力机制引入行人属性识别算法,采用多分支结构,在每条分支上嵌入注意力机制获取多层注意力图. Tang 等<sup>[14]</sup>提出的 ALM(attribute localization module)模块中包含通道注意力子网络和 STN(spatial transformer networks)网络,用于特征提取和空间信息变换. 但是这类算法性能取决于注意力模块的性能,如何建立更高效的注意力模块成为此类算法的关键. Hu 等<sup>[15]</sup>提出的挤压激励(squeeze and excitation, SE)模块能够对各个通道自适应地进行权重分配,提升了网络性能. Saini 等<sup>[16]</sup>提出的 ULSAM(ultra-light weight subspace attention)注意力模块,从每个特征子空间获得各自的注意力图,可以获得多尺度和多频率的特征表示. ULSAM 注意力模块通过分组技术从每个特征子空间获得各自的注意力图,得到了互为补充的子特征图,对细粒度图像分类更为理想,但是 ULSAM 注意力模块没有考虑相邻通道信息之间的联系. Woo 等<sup>[17]</sup>提出的 CBAM(convolutional block attention module)模块可以沿着通道和空间维度有序地获得注意力向量. 石方炎<sup>[18]</sup>通过嵌入 CBAM 注意力模块获得丰富的行人特征信息,使网络性能得到提高;但是 CBAM 中的空间注意力对不同通道上的空间信息以同样的注意力图进行提取,获得的特征图信息之间的耦合性较高. 因此,目前算法对于挖掘行人属性之间相关性均没有得到很好的解决.

基于以上问题,本文提出了一种嵌入 HSA(hierarchy split attention)注意力模块和基于关联规则的网络模型 AA-Net(Apriori and attention network). 模型采用 ResNet50 作为基础网络,将 HSA 注意力模块嵌入 ResNet 的残差模块中,对特征图中的潜在注意力信息进行深度挖掘,通过 Apriori 算法得出属性之间

的关联规则,通过改变全连接层结构提升强相关性属性的准确率.

## 1 相关工作

### 1.1 注意力机制

#### 1.1.1 CBAM

CBAM 的创新之处在于沿着两个独立的维度(通道和空间)有序地进行信息提取,可以获得更全面、更可靠的注意力向量,见图 1 下半部分图形. 图中  $C, H, W$  分别表示特征图矩阵的通道数、高度、宽度,  $\otimes$  表示矩阵点乘. CBAM 中的空间注意力 SAM(spatial attention module)将通道注意力模块的输出作为输入,通过最大池化和平均池化操作得到两个  $1 \times H \times W$  的特征图,接着进行通道连接、卷积、sigmoid 操作获得空间注意力向量,最后与输入特征图相乘得到最终特征,见图 1 上部方框部分. 图中 MaxPool、AvgPool 指通道维度的最大池化、平均池化, sigmoid 为激活函数, Conv 表示卷积层.

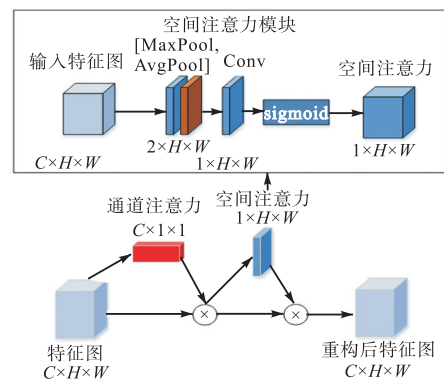


图 1 CBAM 及子模块

Fig. 1 CBAM and sub modules

#### 1.1.2 ULSAM

Saini 等<sup>[16]</sup>在 2020 年提出了注意力模块 ULSAM, 其与一般空间注意力模块的区别在于其应用了分组技术. 分组技术的优势在于经过不同卷积路径得到的特征图间的耦合性较低,各子路径关注的主要特征不同,通道连接后可以得到更加全面的特征图.

ULSAM 模块如图 2 所示. ULSAM 将输入特征图  $F(m \times h \times w)$  划分成  $g$  个通道数大小均为  $G$  的子特征图  $[F_1, \dots, F_n, \dots, F_g]$ , 从每个特征子空间获得各自的注意力图,可以获得多尺度和多频率的特征表示. 这对于细粒度图像分类更为理想,并且会减少网络的参数量. 在获得子特征图的注意力图时采用了

深度可分离卷积<sup>[19]</sup>, 由 DW (deep wise) 和 PW (point wise) 两部分卷积层组成. 与普通卷积的区别在于可以进一步减少参数量, 一般用于轻量级网络, 类似于

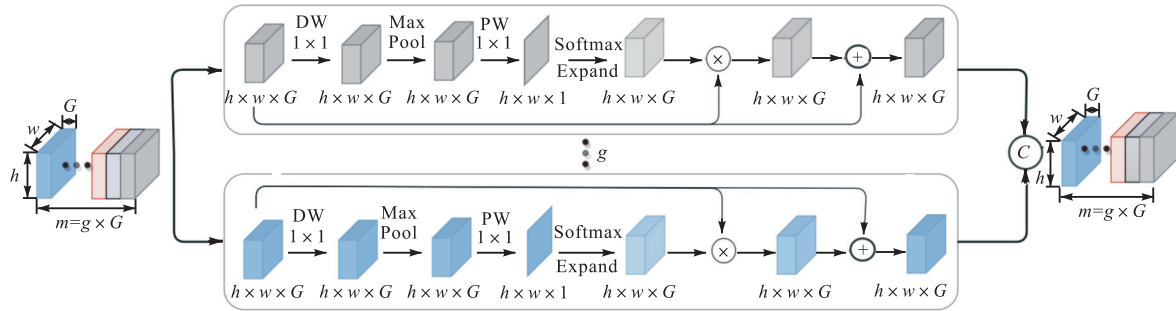


图2 ULSAM 模块

Fig. 2 ULSAM module

## 1.2 关联规则

Lin 等<sup>[20]</sup>通过嵌入行人属性识别网络, 辅助行人重识别任务. 该算法中通过条件概率挖掘属性的相关性, 进行属性权重赋值, 提高网络性能. 但是该算法仅考虑两个属性之间的相关性, 忽略了多属性之间的相关性. 与两属性关联相比, 多属性关联更能确保获得的关联规则的正确性.

本算法基于 Apriori 算法<sup>[21]</sup>挖掘行人数据中属性之间的相关性, 设置最小支持度 (min\_sup), 然后获取项集和频繁项集. 项集是由若干个属性组成的集合; 频繁项集为支持度大于最小支持度的项集. 设置最小置信度 (min\_conf) 后, 与提升度进行过滤, 从而获取有强相关性的频繁项集, 置信度和提升度按照式 (1) 和式 (2) 进行计算.

$$\text{Conf}(y \rightarrow x) = P(x|y) = P(xy)/P(y) \quad (1)$$

$$\text{Lift}(y \rightarrow x) = P(x|y) = \text{Conf}(y \rightarrow x)/P(x) \quad (2)$$

其中:  $y$  表示前件,  $x$  表示后件, Conf 表示置信度, Lift 表示提升度. 当置信度大于 min\_conf 且提升度大于 1 时, 表示此规则是有效的强相关规则.

但是, Apriori 算法存在以下问题: 当支持度设置过高时, 可能会丢掉频率低却有用的规则; 当支持度设置过低时, 可能会获得较多无意义的规则.

## 1.3 ResNet 模型

He 等<sup>[23]</sup>提出的深度残差网络 ResNet (residual neural network), 通过残差操作在一定程度上解决了传统卷积网络的信息丢失、梯度消失、梯度爆炸等问题. 残差模块的思想是通过直接将输入信息绕道传到输出, 保护信息的完整性, 整个网络只需要学习输

入与输出的差别部分, 降低了学习难度. 残差网络结构与  $F_n$  相加得到子空间的特征图. 最后对所有子特征图进行通道连接操作, 获得提取后的特征图. 图中  $\oplus$  表示矩阵相加,  $\otimes$  表示矩阵点乘.

入与输出的差别部分, 降低了学习难度.

残差模块结构如图 3 (a) 所示, 输入特征图直接经过 2 个  $1 \times 1$  卷积层和 1 个  $3 \times 3$  卷积层就可以得到输出特征图. 本算法将不同的注意力模块嵌入第 2 个  $1 \times 1$  卷积层之后, 用来提取特征图信息, 如图 3 (b) 所示.

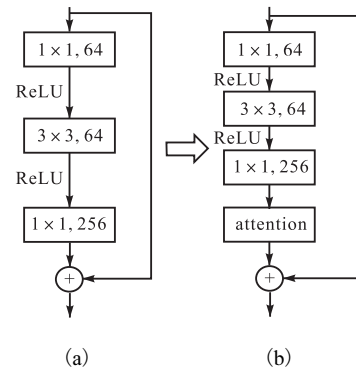


图3 ResNet 残差模块

Fig. 3 ResNet residual module

## 2 本文方法

本算法主要从两方面进行改进: (1) 提出 HSA 注意力模块, 较之 SE、ULSAM 等注意力模块, 能够更好地提升网络性能; (2) 通过 Apriori 算法挖掘属性之间的关系, 获取合适的关联规则, 改变原网络的全连接层结构.

### 2.1 改进的注意力模型

HSA 注意力模块首先将 CBAM 中的空间注意力模块替换成应用了分组技术的 ULSAM 注意力模块, 使模块可以捕获跨通道信息. 接着在 ULSAM 注意力模块中添加了通道交互操作, 使不同组的特征图

享受到不同尺度的感受野,丰富了特征图所包含的信息. HSA 模块的结构如图 4 所示.

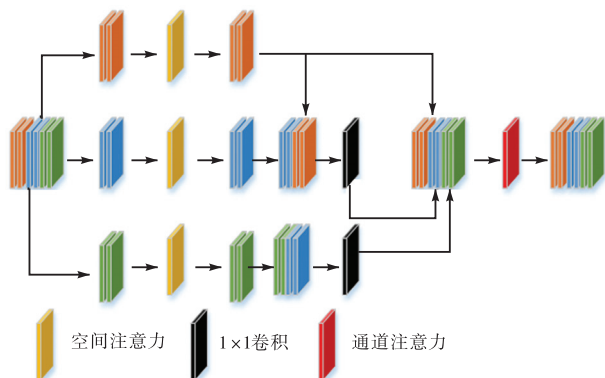


图 4 HSA 模块  
Fig. 4 HSA module

假设残差模块第 2 个  $1 \times 1$  卷积后输出特征图  $F(C \times H \times W)$ , 首先, 将输入特征图  $F(C \times H \times W)$  划分成  $g$  个通道大小均为  $G$  的子特征图  $[F_1, \dots, F_n, \dots, F_g]$ , 取子特征图  $F_n$ , 大小为  $[(C/g) \times H \times W]$ ; 接着, 将  $F_n$  输入空间注意力模块中, 输出特征图  $F'_n$ , 见公式 3 (MaxPool 指通道维度的最大池化操作, sigmoid 为激活函数); 然后, 将相邻的通道做通道连接操作和卷积操作 (Concat 指通道连接操作, Conv 指卷积核大小  $1 \times 1$  的卷积操作), 输出特征图  $F''_n$ ; 最后, 对子特征图  $[F'_1, \dots, F'_n, \dots, F'_g]$  进行通道连接操作后输入 SE 注意力模块中, 输出修正后的特征图  $O$ . FC 为 SE 模块中的全连接层, ReLU 为 SE 模块中的激活函数.

$$F'_n = \text{sigmoid}(\text{MaxPool}(F_n)) * F_n \quad (3)$$

$$F''_n = \begin{cases} F'_n & n = 1 \\ \text{Conv}(\text{Concat}(F'_n \oplus F'_{n-1})) & n > 1 \end{cases} \quad (4)$$

$$O = \text{FC}_2(\text{ReLU}(\text{FC}_1(\text{MaxPool}(\text{Concat}[F'_1, \dots, F'_n, \dots, F'_g]))) \quad (5)$$

## 2.2 基于属性相关性的改进

针对 Apriori 参数设置问题, 柴岩等<sup>[22]</sup>通过引入统计分析等方法, 将项目的权重设为区间值, 将项目的权重和项目的支持数应用于区间估计方法计算其分布, 获得最小支持度区间.

本文提出的基于关联规则的协同训练算法步骤如下:

(1) 获得各数据集上的最小支持度区间, 将区间左端点、中点、右端点设置为最小支持度, 设置最小置信度, 通过 Apriori 算法筛选具有强相关性的属性

对, 获取大于阈值的关联规则.

(2) 通过关联规则找出与后件 (Consequents) 属性存在强相关性的前件 (Antecedents) 高准确率属性集.

(3) 将不同支持度和置信度获取到的关联规则应用于网络中, 改变全连接层结构, 使网络具有更好的判别效果. 根据结果选择对网络性能提升较大的支持度、置信度作为阈值.

协同训练的本质是通过某种方法增加二分类中正确类别的权重, 因此本文将与之相关的且准确率高的属性的全连接层相连接. 以 Market-1501 关联规则中的  $\{\text{'down'}, \text{'cloth'}\} \Rightarrow \{\text{'teenager'}\}$  为例, 后件属性 teenager 的前件属性集为 down 与 cloth, 接着在预训练结果中判断 down 与 cloth 的准确率是否大于设定阈值. 将满足条件的前件与后件的全连接层相连, 见式 (6), 全连接层结构如图 5 所示.

$$\begin{cases} \text{FC}_{\text{teenager}} = \text{FC}_{\text{teenager}} \oplus \text{FC}_{\text{cloth}} \oplus \text{FC}_{\text{down}} \\ \text{FC}_{\text{upsleeve}} = \text{FC}_{\text{upsleeve}} \oplus \text{FC}_{\text{teenager}} \oplus \text{FC}_{\text{down}} \oplus \text{FC}_{\text{cloth}} \end{cases} \quad (6)$$

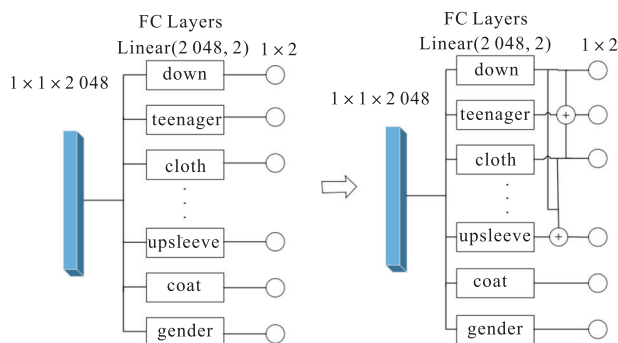


图 5 全连接层结构

Fig. 5 Structure of the whole connection layer

以往的行人算法网络一般通过一个全连接层 FC (channel, numClass) 对所有属性进行预测, 输出向量  $O(1 \times \text{numClass})$  (其中 channel 为最后一个卷积层输出特征图的通道数, numClass 为属性个数). 模型对每一个属性进行预测时, 概率大于 0.5 的预测值为 1, 反之则为 0.

如图 5 所示, 本算法将属性识别作为多任务学习算法, 共享浅层网络参数的同时为每一个属性创建独立的全连接层进行预测, 输出向量  $O(\text{numClass} \times 1 \times 2)$ . 对于每个属性的全连接层输出  $O_i(i = 0, 1, \dots, \text{numClass} - 1)$ , 大小为  $1 \times 2$ , 下标的值是对应预测属性值的概率, 通过比较概率, 选择概率较大值对应的下标作为预测值.

### 3 实验和分析

#### 3.1 数据集

为了验证 HSA 注意力模块和基于关联规则的协同训练策略, 在 PA100K 和 Market-1501 数据集上进行消融实验, 在 PETA 和 PA100K 数据集上进行对比实验。

PA100K 数据集: 共 10 万张行人图像, 其中 8 万张为训练集, 1 万张为验证集, 1 万张为测试集, 每张图像都有 26 个常用二值属性。

Market-1501 数据集: 共 32 668 张图像, 其中训练集包含 12 936 张图像, 测试集包含 19 732 张图像。每张图像含有 9 个二值属性和 3 个多值属性, 将其转换成二值属性共 32 个。

PETA 数据集: 共 19 000 张图像, 11 400 张为训练集, 7 600 张为测试集, 并选取其中正负比例均衡的 35 个属性进行实验。

#### 3.2 实验环境及参数设置

以 PyCharm 为实验平台, 使用 PyTorch 框架进行实验, GPU 型号为 3090, 内存 24 GB。训练次数 epoch 参数设置为 40; 学习率参数初始值设置为 0.000 1, 使用 Warm Up 策略自动调整学习率; batch size 参数设置为 64; 利用反向传播算法和 SGDM 优化器对模型进行训练。多任务联合学习可以共享数据集, 改善模型效果。网络采用多任务学习中的硬连接网络结构, 如图 6 所示。

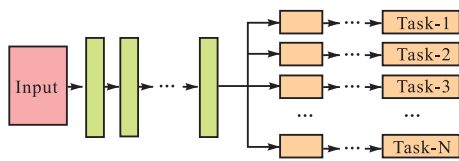


图 6 硬连接网络结构

Fig. 6 Structure of hard connection network

本算法基于实例的评价方式, 根据每个样本的对属性和分错属性关系得到 4 个评价指标, 分别为准确率 (Accuracy)、精确率 (Precision)、召回率 (Recall) 和  $F_1$  值。

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (7)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (8)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (9)$$

$$F_1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

其中: TP 为将正类预测为正类, TN 为将负类预测为负类, FP 为将负类预测为正类, FN 为将正类预测为负类。

#### 3.3 实验结果

以采用硬连接网络结构的 ResNet-50 为基础网络, 即为每个二值属性创建全连接层, 使用 Apriori 协同算法策略 CT (co-train)。通过区间估计方法<sup>[22]</sup>计算最小支持度区间, 各数据集的最小支持度区间见表 1, 将区间左端点、中点、右端点设置为最小支持度。最小置信度设置为 0.5。

表 1 不同数据集的最小支持度区间

Tab. 1 Minimum support interval of different datasets

数据集	左端点	中点	右端点
PA100K	0.245	0.298	0.352
Market-1501	0.353	0.40	0.447
PETA	0.212	0.258	0.304

为了寻找合适的最小置信度, 将 0.5、0.6、0.7、0.8、0.9 分别作为最小置信度, 若相邻的置信度获取的关联规则相同, 则只显示较高的置信度作为最小置信度。从数据集 PA100K 和 Market-1501 获取的关联规则见表 2。其中 PA100K 数据集在最小支持度为 0.245、0.298 且最小置信度为 0.9 时获取的关联规则相同; Market-1501 数据集在最小支持度为 0.353、0.400 且最小置信度为 0.9 时获取的关联规则相同。

##### 3.3.1 消融实验

嵌入注意力机制的网络能够将特征图中潜在的注意力信息进行进一步挖掘, 从而提升网络的性能。为了验证 HSA 注意力模块的有效性, 将不同的注意力模块嵌入残差模块中, 在 PA100K 和 Market-1501 数据集上进行行人属性识别任务, 结果表明 HSA 注意力模块的性能较 SE、ULSAM 和 CBAM 注意力模块的性能有所提高 (表 3)。

SE、ULSAM 注意力模块分别从通道和空间对特征图中的信息进行提取。本实验中 ULSAM 注意力模块的准确率高于 SE 注意力模块, 可能与所选数据集有关。相较于 CBAM 注意力模块, HSA 注意力模块在从多个维度进行特征提取的基础上, 添加了分组和通道交互操作, 可以捕获跨通道信息, 使不同子组的特征图获得不同尺度的感受野。从表 3 中可以看出, 相较于 SE、ULSAM 和 CBAM 注意力模块, HSA 注意力模块的准确率有所提高, 表明了该模块的有效性。

表 2 数据集关联规则

Tab. 2 Rules of dataset association

数据集	最小支持度	最小置信度	前件	后件	支持度	置信度	提升度	
Market-1501	0.353	0.7	{'down', 'cloth'}	{'teenager'}	0.442 5	0.904 4	1.158 2	
			{'teenager', 'down', 'cloth'}	{'upLongSleeve'}	0.424 3	0.958 9	1.009 1	
			{'teenager', 'upLongSleeve'}	{'down'}	0.372 0	0.710 2	1.150 2	
	0.400	0.9	{'down', 'cloth'}	{'teenager'}	0.442 5	0.904 4	1.158 2	
			{'teenager', 'down', 'cloth'}	{'upLongSleeve'}	0.424 3	0.958 9	1.009 1	
	0.447	0.9	{'down', 'cloth'}	{'teenager'}	0.442 5	0.904 4	1.158 2	
			{'down', 'cloth'}	{'upLongSleeve'}	0.471 1	0.962 8	1.013 2	
	PA100K	0.245	0.5	{'LongCoat', 'Skirt/dress'}	{'Female'}	0.330 4	0.913 7	1.030 4
				{'LongSleeve', 'Age18-60'}	{'Trousers'}	0.376 2	0.976 3	1.045 3
				{'Trousers', 'LongSleeve'}	{'Age18-60'}	0.376 2	0.976 3	1.045 2
{'Trousers', 'Age18-60'}				{'LongSleeve'}	0.276 2	0.624 2	1.328 1	
{'Female', 'Age18-60'}				{'ShortSleeve'}	0.268 7	0.549 9	1.065 2	
0.298		0.6	{'LongCoat', 'Skirt/dress'}	{'Female'}	0.330 4	0.913 7	1.030 4	
			{'LongSleeve', 'Age18-60'}	{'Trousers'}	0.376 2	0.976 3	1.045 3	
			{'Trousers', 'LongSleeve'}	{'Age18-60'}	0.376 2	0.976 3	1.045 2	
			{'Trousers', 'Age18-60'}	{'LongSleeve'}	0.276 2	0.624 2	1.328 1	
0.352		0.9	{'LongCoat', 'Skirt/dress'}	{'Female'}	0.330 4	0.913 7	1.030 4	
			{'LongSleeve', 'Age18-60'}	{'Trousers'}	0.376 2	0.976 3	1.045 3	
			{'Trousers', 'LongSleeve'}	{'Age18-60'}	0.376 2	0.976 3	1.045 2	

表 3 在 Market-1501 和 PA100K 上不同注意力的对比

Tab. 3 Comparison of different attentions on Market-1501 and PA100K

数据集	模型	准确率	召回率	精确率	$F_1$
PA100K	Base	0.905 4	0.719 0	0.818 4	0.765 5
	+ SE	0.914 1	0.782 4	0.810 9	0.796 4
	+ ULSAM	0.920 5	0.767 3	0.848 2	0.805 7
	+ CBAM	0.924 3	0.794 5	0.843 9	0.818 4
	+ HSA	0.927 7	0.802 7	0.852 3	0.826 7
Market-1501	Base	0.889 8	0.717 4	0.815 3	0.763 2
	+ SE	0.896 4	0.745 4	0.819 7	0.780 8
	+ ULSAM	0.903 8	0.774 5	0.826 0	0.799 4
	+ CBAM	0.906 7	0.776 9	0.834 8	0.804 8
	+ HSA	0.911 1	0.793 2	0.838 7	0.815 3

通过对比嵌入不同关联规则的 Base 模型的性能, 选择合适的最小支持度和置信度. 通过区间量化方法计算出最小支持度满足的区间, 然后分别选择区间的左端点、右端点和中点作为最小支持度; 将不同置信度作为最小置信度进行消融实验, 结果见表 4. 在 PA100K、Market-1501 数据集上进行实验, 最小支持度、最小置信度分别取 0.245、0.6 和 0.353、0.7 时,

添加 CT 操作后, 准确率的提升最大.

表 4 支持度和置信度的选择

Tab. 4 Selection of support and confidence

数据集	支持度	置信度	模型	准确率	召回率	精确率	$F_1$
PA100K	0.245	0.5	Base	0.905 4	0.719 0	0.818 4	0.765 5
			+ CT	0.912 4	0.714 6	0.850 2	0.776 5
			+ CT	0.912 9	0.716 3	0.851 0	0.777 8
	0.245	0.6	+ CT	0.912 6	0.714 9	0.850 0	0.776 7
			+ CT	0.912 6	0.714 9	0.850 0	0.776 7
	0.352	0.9	+ CT	0.911 6	0.759 1	0.816 4	0.786 7
			Base	0.889 8	0.717 4	0.815 3	0.763 2
Market-1501	0.353	0.7	+ CT	0.902 6	0.748 1	0.818 2	0.781 5
			+ CT	0.899 1	0.732 8	0.827 7	0.777 3
	0.400	0.9	+ CT	0.899 1	0.732 8	0.827 7	0.777 3
			+ CT	0.902 4	0.747 3	0.817 2	0.780 7

在 Base 模型中添加 HSA 注意力模块和 CT 策略验证其有效性, 结果见表 5.

协同训练通过寻找属性之间的关联性, 改变网络结构, 增加正确预测值的权重. 在数据集 PA100K 和 Market-1501 上的实验结果表明: Base 模型添加 CT 操作后准确率分别提高了 0.75%、1.28%; 嵌入 HSA 注意力模块的 Base 模型, 进一步对特征图中的潜在

信息进行挖掘, 准确率分别提高了 2.23%、2.13%; 将两者嵌入 Base 模型中, 准确率进一步提升. 其中 PA100K 数据集虽然在嵌入 HSA 注意力模块的 ResNet 中添加协同策略后, 精确率略微降低, 但是综合指标  $F_1$  值是有提升的. 添加协同训练策略和 HSA 注意力模块后, 关联属性的提升结果见表 6.

表 5 协同策略和 HSA 注意力模块的消融实验

Tab. 5 Ablation experiment of cooperative strategy and HSA attention module

数据集	模型	准确率	召回率	精确率	$F_1$
PA100K	Base	0.905 4	0.719 0	0.818 4	0.765 5
	+ CT	0.912 9	0.716 3	0.851 0	0.777 8
	+ HSA	0.927 7	0.802 7	0.852 3	0.826 7
	本文	0.929 7	0.811 3	0.850 2	0.830 3
Market-1501	Base	0.889 8	0.717 4	0.815 3	0.763 2
	+ CT	0.902 6	0.748 1	0.818 2	0.781 5
	+ HSA	0.911 1	0.793 2	0.838 7	0.815 3
	本文	0.913 2	0.795 3	0.844 7	0.819 3

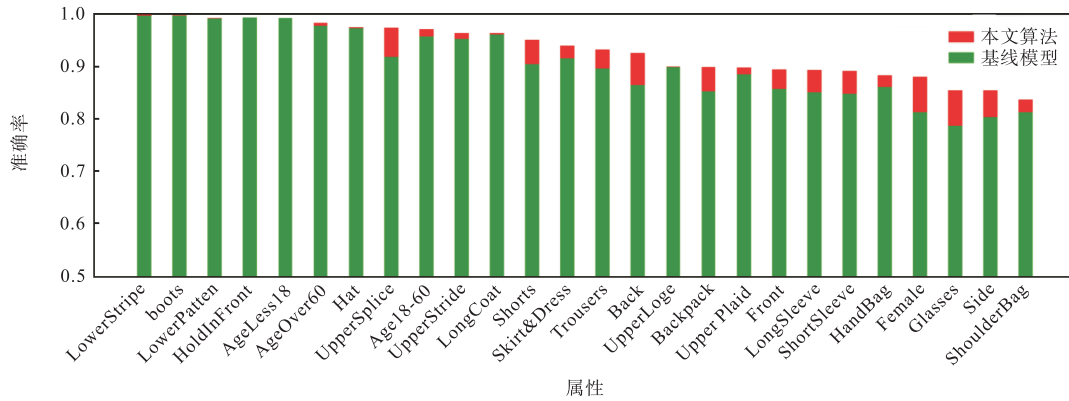


图 7 PA100K 数据集中各属性的提升

Fig. 7 Promotion of attributes in PA100K dataset

### 3.3.2 对比实验

为了验证本文算法的有效性, 在 PA100K 数据集和 PETA 数据集上对行人属性识别任务的结果与近年来较先进的算法进行比较, 包括基于全局特征的 DeepMAR 网络、基于注意力机制的 HP-Net 模型、基于局部特征的 PGDM 网络、基于联合姿态学习的 VeSPA<sup>[24]</sup>等, 结果见表 7 和表 8.

表 7 不同算法在 PA100K 数据集上的识别结果

Tab. 7 Recognition results of different algorithms on PA100K dataset

模型	准确率	召回率	精确率	$F_1$
DeepMAR <sup>[9]</sup>	0.703 9	0.804 2	0.822 4	0.813 2
HP-Net <sup>[13]</sup>	0.721 9	0.820 9	0.829 7	0.825 3
PGDM <sup>[11]</sup>	0.730 8	0.822 4	0.843 6	0.832 9
VeSPA <sup>[24]</sup>	0.730 0	0.814 9	0.849 9	0.832 0
IA2-Net <sup>[25]</sup>	0.747 3	0.833 4	0.850 1	0.845 2
MAN <sup>[26]</sup>	0.781 7	0.828 3	0.849 8	0.857 1
本文算法	0.929 7	0.811 3	0.850 2	0.830 3

表 6 在 Market-1501 和 PA100K 上关联属性的提升

Tab. 6 Promotion of associated attributes on Market-1501 and PA100K

数据集	模型/关联属性	准确率	召回率	精确率	$F_1$
PA100K	Female	0.812 3	0.836 1	0.870 9	0.878 9
	Age18-60	0.954 3	0.959 1	0.961 2	0.967 9
	Trousers	0.893 7	0.917 3	0.924 6	0.929 2
	LongSleeve	0.849 4	0.854 7	0.884 4	0.891 6
Market-1501	teenager	0.826 6	0.835 8	0.857 7	0.866 4
	upLongSleeve	0.918 2	0.924 0	0.928 0	0.934 3
	down	0.887 9	0.910 3	0.918 7	0.920 9

在 PA100K 数据集上进行实验, 将本文算法和基线模型每个属性的准确率进行比较, 根据本文算法的结果进行降序排列, 如图 7 所示. 结果显示: 本文算法在许多属性上取得了较大改进, 包括一些细粒度属性, 例如 Glasses 属性提升了 6.7%、UpperSplice 属性提高了 5.5% 和 HandBag 属性提升了 2.18%. 这些属性识别准确率的提升表明了本文算法的有效性.

表 8 不同算法在 PETA 数据集上的识别结果

Tab. 8 Recognition results of different algorithms on PETA dataset

模型	准确率	召回率	精确率	$F_1$
DeepMAR <sup>[9]</sup>	0.750 7	0.836 8	0.831 4	0.834 1
HP-Net <sup>[13]</sup>	0.791 3	0.849 2	0.832 4	0.840 7
PGDM <sup>[11]</sup>	0.780 8	0.868 6	0.846 8	0.857 6
IA2-Net <sup>[24]</sup>	0.786 2	0.857 3	0.860 7	0.858 8
ALM <sup>[14]</sup>	0.795 2	0.871 1	0.861 8	0.863 6
DTM <sup>[27]</sup>	0.797 5	0.868 1	0.866 7	0.864 8
MLASC <sup>[28]</sup>	0.791 0	0.879 0	0.853 9	0.866 3
本文算法	0.911 9	0.851 3	0.866 8	0.858 9

本文算法与其他算法相比, 召回率、精确率、 $F_1$  值相差不大, 但准确率远高于其他算法. 在行人数据集中, 值为 0 的属性占比较大, 由式 (7) 可知, 当分母一定时, 分子 TN、TP 越大, 准确率越高. 本文算法的准确率远高于其他算法的原因在于 TN 值比较高, 即

预测属性值为 0 时更准确. 召回率、精确率、 $F_1$  值计算时不考虑 TN 值, 所以与其他算法相差不大.

#### 4 结 语

针对行人图像中细粒度属性特征难以捕获的问题, 将 HSA 注意力模块嵌入残差模块进行特征信息提取. HSA 注意力模块首先从空间和通道多个方向获取特征信息, 并且增加了分组和通道交互操作, 使之可以获得具有丰富信息的特征图; 其次针对先前研究工作对行人属性相关性分析不足的问题, 提出了基于关联规则的协同训练算法. 协同训练算法通过 Apriori 算法充分挖掘属性之间的相关性信息, 改变原有的全连接层结构, 增加正确预测值的权重. 结果表明, 本文算法在召回率、精确率、 $F_1$  值上与其他算法相当, 在准确率指标上有较大提升. 未来行人属性识别仍有很多工作要做, 包括优化模型和数据集本身的问题, 对于实际采集的数据集存在各属性值分布不平衡的问题, 可以通过算法进行优化.

#### 参考文献:

- [ 1 ] WANG X, ZHENG S, YANG R, et al. Pedestrian attribute recognition: a survey[J]. Pattern recognition, 2022, 121: 108220.
- [ 2 ] SCHUMANN A, STIEFELHAGEN R. Person re-identification by deep learning attribute-complementary information[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. New York: IEEE, 2017: 8014920.
- [ 3 ] 曹自强, 赛斌, 吕欣. 行人跟踪算法及应用综述[J]. 物理学报, 2020, 69(8): 41-58.
- [ 4 ] 苏松志, 李绍滋, 陈淑媛, 等. 行人检测技术综述[J]. 电子学报, 2012, 40(4): 814-820.
- [ 5 ] DALAL N, TRIGGS B. Histograms of oriented gradients for human detection[C]//IEEE. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2005: 886-893.
- [ 6 ] LOWE D G. Distinctive image features from scale-invariant keypoints[J]. International journal of computer vision, 2004, 60(2): 91-110.
- [ 7 ] CHANG C C, LIN C J. LIBSVM: a library for support vector machines[J]. ACM Transactions on intelligent systems and technology, 2011, 2(3): 1-27.
- [ 8 ] LAFFERTY J, MCCALLUM A, PEREIRA F C N. Conditional random fields: probabilistic models for segmenting and labeling sequence data[C]// ICML. Proceedings of the Eighteenth International Conference on Machine Learning. San Francisco: Morgan Kaufmann Publishers Inc., 2001: 282-289.
- [ 9 ] LI D, CHEN X, HUANG K. Multi-attribute learning for pedestrian attribute recognition in surveillance scenarios[C]//IEEE. 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR). New York: IEEE, 2015: 111-115.
- [ 10 ] SUDOWE P, SPITZER H, LEIBE B. Person attribute recognition with a jointly-trained holistic CNN model[C]//Proceedings of the IEEE International Conference on Computer Vision Workshops. New York: IEEE, 2015: 87-95.
- [ 11 ] LI D, CHEN X, ZHANG Z, et al. Pose guided deep model for pedestrian attribute recognition in surveillance scenarios[C]//IEEE. 2018 IEEE International Conference on Multimedia and Expo (ICME). New York: IEEE, 2018: 1-6.
- [ 12 ] LIU P, LIU X, YAN J, et al. Localization guided learning for pedestrian attribute recognition[EB/OL]. [2022-04-01]. <http://arxiv.org/abs/1808.09102>.
- [ 13 ] LIU X, ZHAO H, TIAN M, et al. HydraPlus-Net: attentive deep features for pedestrian analysis[C]//IEEE. Proceedings of the IEEE International Conference on Computer Vision. New York: IEEE, 2017: 350-359.
- [ 14 ] TANG C, SHENG L, ZHANG Z, et al. Improving pedestrian attribute recognition with weakly-supervised multi-scale attribute-specific localization[C]//IEEE. Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE, 2019: 4997-5006.
- [ 15 ] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2018: 7132-7141.
- [ 16 ] SAINI R, JHA N K, DAS B, et al. ULSAM: ultra-lightweight subspace attention module for compact convolutional neural networks[C]//IEEE. Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. New York: IEEE, 2020: 1627-1636.
- [ 17 ] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). Berlin: Springer, 2018: 3-19.
- [ 18 ] 石方炎. 人体检测与外观属性识别一体化算法研究[D]. 成都: 电子科技大学, 2020.
- [ 19 ] CHOLLET F. Xception: deep learning with depth wise separable convolutions[C]//IEEE. Proceedings of the



- IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2017: 1251–1258.
- [20] LIN Y, ZHENG L, ZHENG Z, et al. Improving person re-identification by attribute and identity learning[J]. Pattern recognition, 2019, 95: 151–161.
- [21] AGRAWAL R, SRIKANT R. Fast algorithms for mining association rules[C]//Proceedings of the 20th International Conference on Very Large Data Bases. San Francisco: Morgan Kaufmann Publishers Inc., 1994, 1215: 487–499.
- [22] 柴岩, 张京辉, 鲁新新. 最小支持度为区间值的加权 Apriori 算法[J]. 辽宁工程技术大学学报(自然科学版), 2016, 35(12): 1512–1516.
- [23] HE K, ZHANG X, REN S, et al. Deep residual learning for image recognition[C]//IEEE. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE, 2016: 770–778.
- [24] SARFRAZ M S, SCHUMANN A, WANG Y, et al. Deep view-sensitive pedestrian attribute inference in an end-to-end model[EB/OL]. [2022-04-01]. <http://arxiv.org/abs/1707.06089v1>.
- [25] JI Z, HE E, WANG H, et al. Image-attribute reciprocally guided attention network for pedestrian attribute recognition[J]. Pattern recognition letters, 2019, 120: 89–95.
- [26] 李娜, 武阳阳, 刘颖, 等. 基于多尺度注意力网络的行人属性识别算法[J]. 激光与光电子学进展, 2021, 58(4): 290–296.
- [27] 王林, 李聪会. 基于多级注意力跳跃连接网络的行人属性识别[J]. 计算机工程, 2021, 47(2): 314–320.
- [28] 张再腾, 张荣芬, 刘宇红. 基于多尺度残差注意网络的轻量级行人属性识别算法[J/OL]. 控制与决策: 1–10[2022-04-01]. DOI:10.13195/j.kzyjc.2021.0411.

责任编辑: 郎婧

(上接第 60 页)

- [13] 王朋鹤. 基于双向门控循环神经网络的风电机组状态检测[D]. 保定: 华北电力大学, 2021.
- [14] 杨云, 丁磊, 张昊宇. 改进一维卷积神经网络与双向门控循环单元的轴承故障诊断研究[J/OL]. 机械科学与技术: 1–8[2022-04-11]. DOI:10.13433/j.cnki.1003-8728.20200638.
- [15] 沈艳霞, 周文晶, 纪志成, 等. 基于小波包与 SVM 的风电变流器故障诊断[J]. 太阳能学报, 2015, 36(4): 785–791.
- [16] 朱永超, 朱才朝, 宋朝省, 等. PCA-PSO/GS-SVM 组合方法在风电齿轮箱故障预测中的应用研究[J]. 太阳能学报, 2021, 42(3): 35–42.
- [17] IOFFE S, SZEGEDY C. Batch normalization: accelerating deep network training by reducing internal covariate shift[J]. Journal of machine learning research, 2015, 37: 448–456.
- [18] 刘秀丽, 徐小力. 基于深度信念网络的风电机组齿轮箱故障诊断方法[J]. 可再生能源, 2017, 35(12): 1862–1868.
- [19] SUN W, ZHAO R, YAN R, et al. Convolutional discriminative feature learning for induction motor fault diagnosis[J]. IEEE Transactions on industrial informatics, 2017, 13(3): 1350–1359.
- [20] 靳志杰, 霍志红, 许昌, 等. 基于特征选择和 XGBoost 的风电机组故障诊断[J]. 可再生能源, 2021, 39(3): 353–358.

责任编辑: 郎婧