

# 局部特征重叠的行人属性识别方法<sup>①</sup>



陈伟航, 刘志刚, 黄朝, 谢东军

(东北石油大学 计算机与信息技术学院, 大庆 163318)

通信作者: 刘志刚, E-mail: 2803930906@qq.com

**摘要:** 针对目前的行人属性识别方法存在行人属性数据不均衡、行人特征表达能力不足、鲁棒性差的问题, 本文提出局部特征重叠与行人属性识别相结合的方法. 网络使用全局和局部两个分支来提升网络整体特征表达能力, 在局部分支中将得到的特征图切分为几块大小相同的几个部分并使用 Focal loss 计算每个属性的损失解决行人属性不均衡问题. 最后将投票选出的各属性最优损失与全局特征计算出来的 ID 损失共同作为模型总损失. 在 Market-1501\_attribute 和 DukeMTMC-attribute 两个行人属性数据集上进行测试, 实验结果表明所提方法能够有效提高行人重识别的准确性.

**关键词:** 行人重识别; 全局特征; 局部特征; 属性信息; Focal loss

引用格式: 陈伟航, 刘志刚, 黄朝, 谢东军. 局部特征重叠的行人属性识别方法. 计算机系统应用, 2022, 31(4): 381-385. <http://www.c-s-a.org.cn/1003-3254/8444.html>

## Pedestrian Attribute Recognition Method Based on Local Feature Overlap

CHEN Wei-Hang, LIU Zhi-Gang, HUANG Zhao, XIE Dong-Jun

(School of Computer and Information Technology, Northeast Petroleum University, Daqing 163318, China)

**Abstract:** Given the unbalanced pedestrian attribute data, insufficient expression ability of pedestrian features, and poor robustness of current pedestrian attribute recognition methods, this study proposes a method based on local feature overlap and pedestrian attribute recognition. The network uses global and local branches to improve the overall feature expression ability of the network. In the local branch, the feature graph obtained is divided into several parts with the same size, and the loss of each attribute is calculated with the Focal loss function to solve the problem of pedestrian attribute imbalance. Finally, the optimal loss of each attribute selected by voting and the ID loss calculated through global features are taken as the total loss of the model. The proposed method is tested on Market-1501\_attribute and DukeMTMC-attribute pedestrian attribute datasets, and the experimental results show that this method can effectively improve the accuracy of pedestrian re-recognition.

**Key words:** pedestrian recognition; global features; local features; attribute information; Focal loss

行人重识别是指在多个不重叠摄像头中检索出目标行人的技术, 可以广泛应用于智慧城市、图像管理等领域. 但是, 由于使用不同摄像头所拍摄的行人图像存在光照、视角、遮挡以及摄像头分辨率等问题, 使得行人重识别技术目前仍面临着很大的困难. 随着深度学习的发展, 深度学习方法在行人重识别上也得到

了很好的应用. 实验结果也表明使用深度学习的方法相较于传统方法能够学习到更深层次的图像特征, 从而更好地实现行人重识别. 但现有的使用深度学习的行人重识别方法大多采用加深神经网络层数的方法, 会导致神经网络参数急剧增多, 从而使神经网络训练的难度增大, 耗费大量的计算资源也难以提取到比较

<sup>①</sup> 基金项目: 黑龙江省自然科学基金 (LH2020F003)

收稿时间: 2021-07-02; 修改时间: 2021-08-10; 采用时间: 2021-08-17; csa 在线出版时间: 2022-03-22

高层的行人语义信息. 因此一些学者提出通过利用标注好的行人属性信息来提高对行人的特征提取能力从而提高行人重识别效果. 如: Zhu 等<sup>[1]</sup>提出使用多标签的卷积神经网络 (multi-label CNN, MLCNN) 模型, 将行人图像划分为多个可重叠的部分, 并对每个部分都进行相应的属性预测. Fabbri 等<sup>[2]</sup>同样是将行人分为头、肩、上身、下身几个部分来分别进行属性预测. Liu 等<sup>[3]</sup>提出多方向注意力网络 (hydraplus-net: attentive deep features for pedestrian analysis, HP-Net), 将多层注意力特征图多向映射到不同的特征层中, 从而能够挖掘到不同尺度的注意力特征, 充实行人的特征表示. Li 等<sup>[4,5]</sup>提出深度多属性识别 (deep learning based multiple attributes recognition, DeepMAR) 和深度单属性识别 (deep learning based single attribute recognition, DeepSAR) 两个网络模型, Deep SAR 能独立识别每个属性, 而 Deep MAR 则是利用属性之间的关系进行属性识别, 最后将两个模型联合共同进行识别. 但这些网络都没有考虑局部特征对整体网络识别的影响, 也没有关注到行人属性不均衡所导致的网络训练效果不佳的情况.

针对上述问题, 本文提出了一种结合行人属性标签的方法, 通过对多个有重叠的行人局部特征进行属性识别作为行人全局特征的补充信息来提高识别的准确率. 其次, 使用 Focal loss<sup>[6]</sup> 损失函数来解决属性识别任务中的正负样本不平衡和难易样本不平衡问题.

## 1 相关理论

### 1.1 Focal loss

Focal loss 损失函数是在标准交叉熵损失<sup>[7]</sup>的基础上修改得到的, 它通过引入参数  $\gamma$  来降低易分类样本的权重从而使模型更专注于难分类样本的训练, 同时还加入了平衡因子  $\alpha$  用来平衡正负样本比例不均, 从而也有效解决属性样本不平衡的问题. 二分类交叉熵损失函数为:

$$L_{fl} = \begin{cases} -\alpha(1-y')^\gamma \log y', & y = 1 \\ -(1-\alpha)y'^\gamma \log(1-y'), & y = 0 \end{cases} \quad (1)$$

$$\begin{aligned} L &= -y \log y' - (1-y) \log(1-y') \\ &= \begin{cases} -\log y', & y = 1 \\ -\log(1-y'), & y = 0 \end{cases} \end{aligned} \quad (2)$$

其中,  $y$  为真实值,  $y'$  为预测值. 当  $y=1$  时, 预测值越接近真实标签 1, 损失函数越小, 预测值越接近真实标签 0, 损失越大.  $y=0$  时, 预测值越接近真实值 0, 损失越小, 预测值越接近真实值 1, 损失越大. 此时的损失函数就可能在大量简单样本的迭代中训练缓慢且很难达到最优的训练效果. Focal loss 的引入主要是为解决训练过程中难易样本和正负样本不平衡的问题, 它从交叉熵函数改进而来, 其函数为:

首先在原有的交叉熵函数基础上加入了两个参数  $\gamma$  和  $\alpha$ . 加入参数  $\gamma>0$  是用来降低易分类样本的损失, 使得网络更关注困难的、容易被错分的样本. 此外, 加入了平衡因子  $\alpha$ , 用来平衡正负样本本身的比例不均.  $\gamma$  调节简单样本权重降低的速率, 当  $\gamma$  为 0 时就变成了传统的交叉熵损失函数, 当  $\gamma$  增加时调制系数也在增加, 对网络的影响也越大.

### 1.2 行人属性识别网络

行人属性识别网络 (attribute-person recognition network, APR)<sup>[8]</sup> 是将行人属性识别当作一个多任务学习的方式, 其结构如图 1 所示. 在损失计算前有  $m+1$  个全连接层 (fully connected layer, FC), 其中有一个身份分类损失和  $m$  个属性分类损失.  $m+1$  个 FC 层分别表示为 FC, FC<sub>1</sub>, ..., FC<sub>m</sub>, 其中 FC 用于 ID 分类, FC<sub>1</sub>, ..., FC<sub>m</sub> 用于属性识别, 预训练模型使用 ResNet50<sup>[9]</sup>. 对于属性预测, 使用  $M$  个 Softmax loss 来进行计算. 通过用一个多属性分类损失函数和一个 id 分类损失函数, 训练 APR 网络来预测属性和 id 标签, 最后的损失函数定义如下:

$$L = \lambda L_{ID} + \frac{1}{M} \sum_{i=1}^M L_{att} \quad (3)$$

其中,  $L_{ID}$  和  $L_{att}$  分别表示行人的 ID 分类和属性分类的交叉熵损失, 参数  $\lambda$  用于衡量这两个损失函数在 Loss 的权重.

## 2 局部特征重叠的行人属性识别方法

### 2.1 ALFO 网络

本文提出的局部特征重叠的行人属性网络 (attribute-local feature overlap network, ALFO) 主要由 3 部分组成, 其结构如图 2 所示. 首先是特征提取部分, 利用 ResNet50 作为基础网络, 其次是进行属性识别任务的局部分支部分, 最后是进行行人 ID 预测的全局分支部分.

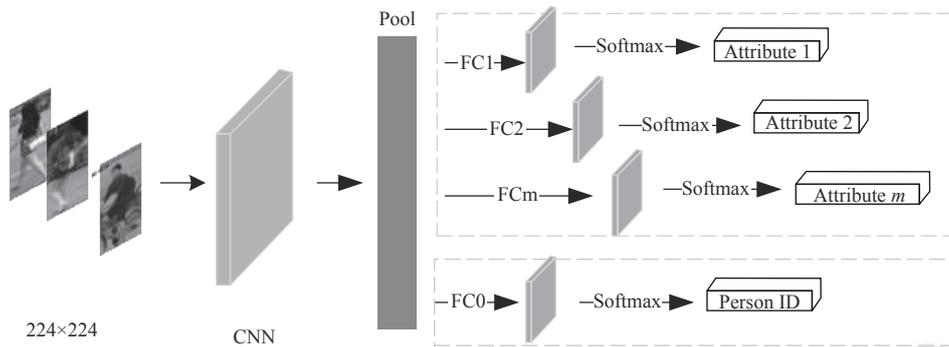


图1 行人属性识别 APR 网络模型图

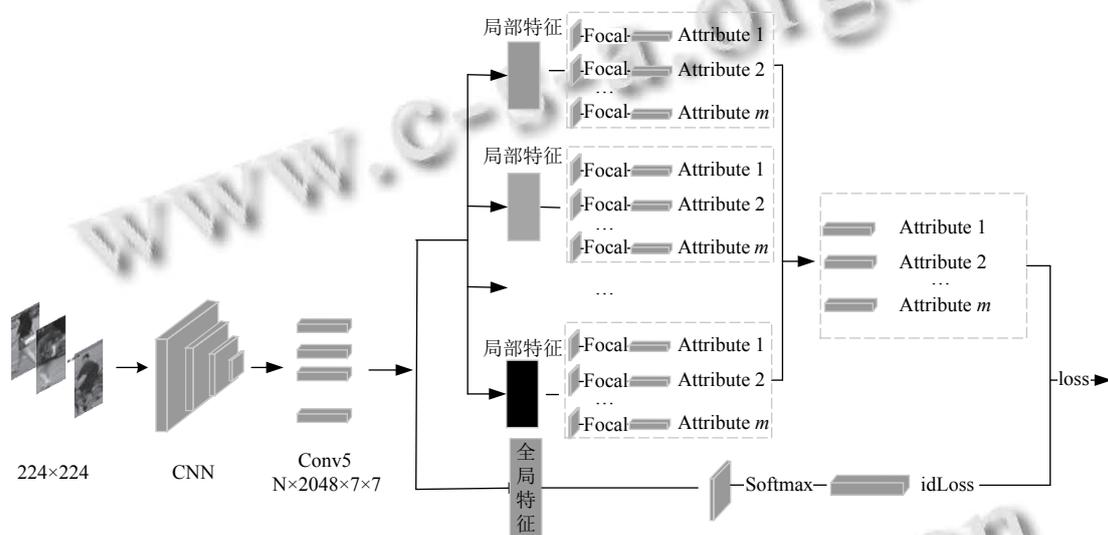


图2 局部特征重叠的行人属性识别模型图

本文基准网络采用的是 Resnet50 网络, 并去掉网络最后两层使其输出大小为  $7 \times 7$  的 2048 维图像. 在局部分支中首先将通过基准网络得到的特征图像切分为多个大小相同且有重叠的行人局部特征图, 实验发现将特征图切分为 6 块得到的最终训练效果最好. 再将每块局部特征都分别进行多任务属性学习, 并利用  $m$  个全连接层来同时预测行人的属性, 其中  $m$  是标注的属性种类数量. 然后对每块局部特征计算出的属性损失通过投票得出每个属性所对应的最优损失. 在全局分支中使用全连接层来预测行人的身份标签并计算 ID 损失, 最终网络的总损失为行人的 ID 损失与投票选出的属性损失之和.

### 2.2 损失函数

该网络的总损失包括使用行人身份标签预测的 ID 损失和使用行人属性标签预测的损失. 在行人属性

数据集中, 由于行人包含有多种不同种类的属性会导致出现属性样本不平衡问题, 包括类内样本不平衡和类间样本不平衡. 其中类内样本不平衡是指在某一类别属性下, 如在下身衣服颜色的属性下, 不同颜色下身衣服的样本数量不平衡. 类间样本不平衡表示不同属性类别间的样本数量不平衡: 如帽子和上衣属性样本数量不平衡. 属性样本不平衡会导致网络训练效果差或者训练效率低下等问题, 一般可以使用三元组损失 (triplet loss)<sup>[10]</sup> 或者四元组损失 (quadruplet loss)<sup>[11]</sup> 等度量学习的方式或者使用数据预处理的方式对解决样本不平衡问题有一定效果. 本文通过改进损失函数, 在局部分支进行属性识别时使用 Focal loss 计算损失, 能够解决在训练中属性样本不平衡的问题.

### 2.3 行人局部特征属性识别

前期基于深度学习的行人重识别研究中, 虽然是

基于神经网络自动提取特征,但是较少有方法能够提取到行人高层语义信息.随着行人重识别研究的不断深入以及标注行人属性的大规模数据集的提出,研究者们越来越关注通过行人属性特征来提取行人的高层语义信息从而提高行人重识别的准确率.另外,考虑到行人重识别问题经常面临视角变化、遮挡等情况,而之前的行人重识别方法大多只学习了行人的全局特征没有考虑人体的整体空间结构导致学习困难,因此在全局特征上强调局部差异是非常重要的.本文将行人特征图像切分为大小相同且有重叠的行人局部特征图,然后对每块局部特征分别计算属性损失,最后再通过投票得到识别效果最好的属性损失,这样网络模型同时学习到了行人的局部特征和全局特征从而提高行人重识别的准确率.

### 3 实验

#### 3.1 数据集

本文实验使用了两个行人属性数据集: Market-1501\_attribute<sup>[8]</sup>, DukeMTMC-attribute<sup>[8]</sup>. 这两个行人属性数据集,分别在 Market-1501 和 DukeMTMC 这两个行人重识别数据集的基础上进行行人属性信息标注得到的. Market-1501 数据集在清华大学校园中采集,拍摄于夏季.它是由 6 个摄像头在不同区域所拍摄到的 1501 个行人组成,且每个行人至少由 2 个摄像头所捕获到.共包含训练集 751 人,由 12 936 张图像组成,测试集由 750 个行人的共 19 732 张图像组成. Market-1501\_attribute 为每一个行人标注包括性别、年龄、上衣颜色、下身颜色等 27 个属性. DukeMTMC 数据集拍摄于冬天的杜克大学,它由 8 个高清摄像头拍摄,包含来自 702 个人的 16, 522 张训练图片和来自另外 702 个人的 2 228 张查询图片. DukeMTMC-attribute 数据集为每一个行人标注了包括性别、上衣长度、上衣颜色、下身衣服颜色等 23 个属性. 通过使用标注了行人属性信息的数据集,使得网络能通过对相应部位的属性进行学习来更好地关注到不同部位的细节,从而提取更有代表性的特征.

#### 3.2 实验结果及分析

本文基线方法采用属性识别网络,因为 Market-1501\_attribute 和 DukeMTMC-attribute 两个数据集是 Lin 等人<sup>[8]</sup>标注并公开的,且 APR 网络与本文方法相似,都是同时学习了行人身份和属性信息.同时本文也

加入了局部分支的方法,与行人重识别 PCB (part-based convolutional baseline)<sup>[12]</sup>方法有一定相似之处,因此本文与 APR 和 PCB 网络分别比较,可以验证本文方法的有效性.

本文采用了两种对照实验,分别是通过消融实验证实使用局部分支以及使用 Focal loss 的有效性以及与主流行人重识别算法效果的对比,具体如下.

实验 1. 分别只使用全局分支、只使用局部分支以及不使用 Focal loss 方法与本文算法进行对比来验证使用全局和局部特征相结合的方法比只使用局部或者全局行人特征的效果更好,同时也验证了 Focal loss 函数的有效性.

实验 2. 将本文算法与 PCB, APR, DeepMar 等主流行人重识别算法在 Market-1501 和 DukeMTMC 数据集上进行对比实验以验证本文算法相比以往一些主流的行人重识别算法效果有所提升.

实验结果表明:由表 1 可知,在 Market-1501 和 DukeMTMC 数据集上分别进行测试,结果表明,相比于只使用全局分支的网络,只用局部分支的网络其 Rank-1 和 mAP 都更高,说明只有局部信息没有行人 ID 信息不能很好的识别行人身份,而本文算法相比于只用全局分支和只用局部分支以及不使用 Focal loss 函数的模型, Rank-1 和 mAP 都有较大提升.说明使用提取全局分支和局部分支相结合的方法有一定效果.

表 1 使用单独分支实验对比表(%)

| 方法          | Market-1501 |       | DukeMTMC |       |
|-------------|-------------|-------|----------|-------|
|             | Rank-1      | mAP   | Rank-1   | mAP   |
| 局部分支        | 70.44       | 58.91 | 76.31    | 52.43 |
| 全局分支        | 74.46       | 66.67 | 73.23    | 62.44 |
| 无Focal loss | 86.52       | 77.42 | 74.69    | 70.53 |
| ALFO        | 91.25       | 77.83 | 78.86    | 74.12 |

实验结果表明:由表 2 可以看出,使用本文方法与有代表性的 APR、PCB、DeepMar 以及 TriHard 和 PAR 等基准网络相比,在 Market-1501 数据集上 mAP 分别提升了 11.24%、7.43%、4.11%. 在 DukeMTMC 数据集上 mAP 分别提升了 15.38%、8.02%、1.42%. 说明使用局部特征重叠的行人属性识别方法后相比 APR、PCB 等主流行人重识别算法在性能上得到了进一步提升,证明本文所提方法有效.

## 4 结论

行人重识别领域一直存在着许多挑战,现有的方

法是直接使用卷积神经网络来提取行人特征,但没有充分利用行人的属性信息.在实际场景中,行人不对齐、行人属性数据集样本不平衡等问题一直很难得到有效解决.本文提出了一种将局部特征重叠与行人属性识别相结合的网络结构,能够同时学习图像的全局特征和局部特征,从而能一定程度上解决行人不对齐和特征表达能力不足的问题,使性能指标得到提升.同时,对于行人属性样本不平衡问题,使用 Focal loss 代替原来的交叉熵损失,能够一定程度上解决属性样本不平衡问题所带来的训练效果差或训练效率低下问题,从而提高算法精度.

表2 与其他主流行人重识别算法实验对比表(%)

| 方法                     | Market-1501 |       | DukeMTMC |       |
|------------------------|-------------|-------|----------|-------|
|                        | Rank-1      | mAP   | Rank-1   | mAP   |
| 本文算法                   | 91.25       | 77.83 | 78.86    | 74.12 |
| TriHard <sup>[9]</sup> | 83.4        | 71.23 | 78.55    | 69.74 |
| PAR <sup>[13]</sup>    | 80.1        | 65.32 | 81.46    | 58.15 |
| APR <sup>[8]</sup>     | 81.03       | 66.59 | 70.69    | 61.74 |
| PCB <sup>[12]</sup>    | 90.3        | 70.40 | 80.6     | 66.1  |
| DeepMar <sup>[4]</sup> | 88.75       | 73.72 | 82.6     | 72.7  |

### 参考文献

- Zhu JQ, Liao SC, Yi D, *et al.* Multi-label CNN based pedestrian attribute learning for soft biometrics. Proceedings of 2015 International Conference on Biometrics. Phuket: IEEE, 2015. 535–540.
- Fabrizi M, Calderara S, Cucchiara R. Generative adversarial models for people attribute recognition in surveillance. Proceedings of the 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance. Lecce: IEEE, 2017. 1–6.
- Liu XH, Zhao HY, Tian MQ, *et al.* HydraPlus-Net: Attentive deep features for pedestrian analysis. Proceedings of 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 350–359.
- Li DW, Chen XT, Huang KQ. Multi-attribute learning for pedestrian attribute recognition in surveillance scenarios. Proceedings of the 2015 3rd IAPR Asian Conference on Pattern Recognition. Kuala Lumpur: IEEE, 2015. 111–115.
- Sudowe P, Spitzer H, Leibe B. Person attribute recognition with a jointly-trained holistic CNN model. Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago: IEEE, 2015. 329–337.
- Lin TY, Goyal P, Girshick R, *et al.* Focal loss for dense object detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318–327. [doi: [10.1109/TPAMI.2018.2858826](https://doi.org/10.1109/TPAMI.2018.2858826)]
- Kobayashi T. Group softmax loss with discriminative feature grouping. Proceedings of 2021 IEEE Winter Conference on Applications of Computer Vision. Waikoloa: IEEE, 2021. 2614–2623.
- Lin YT, Zheng L, Zheng ZD, *et al.* Improving person re-identification by attribute and identity learning. Pattern Recognition, 2019, 95: 151–161. [doi: [10.1016/j.patcog.2019.06.006](https://doi.org/10.1016/j.patcog.2019.06.006)]
- He KM, Zhang XY, Ren SQ, *et al.* Deep residual learning for image recognition. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 770–778. [doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90)]
- Schroff F, Kalenichenko D, Philbin J. FaceNet: A unified embedding for face recognition and clustering. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston: IEEE, 2015. 815–823.
- Chen WH, Chen XT, Zhang JG, *et al.* Beyond triplet loss: A deep quadruplet network for person re-identification. Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: IEEE, 2017. 1320–1329.
- Sun YF, Zheng L, Yang Y, *et al.* Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). Proceedings of the 15th European Conference on Computer Vision. Munich: Springer, 2018. 501–518.
- Zhao LM, Li X, Zhuang YT, *et al.* Deeply-learned part-aligned representations for person re-identification. Proceedings of 2017 IEEE International Conference on Computer Vision. Venice: IEEE, 2017. 3239–3248.