

基于改进损失函数的多阶段行人属性识别方法

郑少飞¹ 汤进¹ 罗斌¹ 王逍¹ 王文中¹

摘要 大量研究工作通过挖掘属性间的正相关性提高视频监控场景下的行人属性识别性能,但对属性间负相关性的探索仍存在不足.为此,文中基于深度学习理论提出多阶段行人属性识别方法,同时探索属性间的正、负相关性.第一阶段计算每个属性在训练过程中的损失值和正确率.第二阶段为平均损失较大且正确率较小的属性单独建立一个网络分支,其它属性仍保留在原分支上,然后两个分支联合预测所有属性.第三阶段新建两个网络分支,结构与第二阶段的分支相同,优化新分支的参数,使其属性识别性能优于第二阶段.最终使用第三阶段的模型进行属性预测.此外,构建增大正负样本差异的改进损失函数,应用于三个阶段的训练,进一步提升模型性能.在两个行人属性识别数据集 RAP 和 PETA 上的实验表明,文中方法性能较优.

关键词 视频监控, 行人属性, 深度学习, 多阶段, 损失函数

引用格式 郑少飞,汤进,罗斌,王逍,王文中.基于改进损失函数的多阶段行人属性识别方法.模式识别与人工智能,2018,31(12):1085-1095.

DOI 10.16451/j.cnki.issn1003-6059.201812003

中图法分类号 TP 391.4

Multistage Pedestrian Attribute Recognition Method Based on Improved Loss Function

ZHENG Shaofei¹, TANG Jin¹, LUO Bin¹, WANG Xiao¹, WANG Wenzhong¹

ABSTRACT There are plenty of studies on improving the performance of pedestrian attribute recognition in video surveillance scenarios by mining the positive correlations between attributes. However, the research on negative correlations is far from enough. In this paper, a deep learning based multi-stage pedestrian attribute recognition method is proposed to simultaneously explore the positive and negative correlations between attributes. In the first stage, the loss value and the accuracy of each attribute are calculated during training. In the second stage, a new network branch is designed for the attributes with larger average loss and lower average accuracy, while other attributes remain on the original branch. All attributes are predicted by these two branches jointly. In the third stage, two new network branches with same structure as the second stage are designed to optimize the parameters and improve the performance during attribute recognition. Moreover, an improved loss function increasing the distance between positive and negative samples is proposed, and it is applied in all training stages to further improve the performance. Experiments on datasets RAP and PETA validate the promising performance of the proposed method.

Key Words Video Surveillance, Pedestrian Attribute, Deep Learning, Multistage, Loss Function

Citation ZHENG S F, TANG J, LUO B, WANG X, WANG W Z. Multistage Pedestrian Attribute Recognition Method Based on Improved Loss Function. Pattern Recognition and Artificial Intelligence, 2018, 31(12): 1085-1095.

收稿日期:2018-09-15;录用日期:2018-11-21

Manuscript received September 15, 2018;

accepted November 21, 2018

国家自然科学基金项目(No. 61671018, 61472002, 61502006)
资助

Supported by National Natural Science Foundation of China(No.

61671018, 61472002, 61502006)

本文责任编辑 陈恩红

Recommended by Associate Editor CHEN Enhong

1. 安徽大学 计算机科学与技术学院 合肥 230601

1. School of Computer Science and Technology, Anhui University, Hefei 230601

近年来,安防产业发展迅猛,视频监控设备大量增长,利用计算机技术实现监控系统的智能化成为一个热门的研究领域.行人作为监控场景下的重要目标之一,视觉属性(性别、年龄和衣着类型等)识别成为计算机视觉中的一个重要课题,该课题的任务是:输入监控场景下的行人图像,经算法处理后,输出一系列行人属性的预测.

行人视觉属性作为中级语义特征,对视角变化及观察条件具有鲁棒性,可以建立人的低层特征和高层认知的联系,辅助应用到人的面部识别^[1-3]、行人再识别^[4-5]和身份确认^[6]等多种视觉任务.真实监控场景下的行人属性识别极具挑战性,主要难点有:1)图像质量较差,分辨率较低,遮挡及运动模糊等;2)大多数属性的识别需要局部的细粒度信息,如“眼镜”等;3)属性外观和空间位置容易发生变化,如不同类型的“包”等;4)缺乏大规模数据集并且样本分布严重不均衡.早期的属性识别方法依赖手工特征^[7-9],并单独为每个属性设计分类器,但传统的手工特征难以应对多变的监控场景及行人外观,忽略属性间的关系.属性间的相关性可作为一种约束推理以弥补图像中因噪声而缺失的信息.学者们提出基于图模型的方法^[10-12],利用马尔科夫随机场挖掘属性间的相关性,但需要计算不同属性对之间的关系,由此产生大量模型参数,当属性过多时计算开销过大.

近年来,深度学习的兴起使计算机视觉的研究取得长足进步.基于深度学习的行人属性识别方法^[13-16]获得研究者们的关注.主流方法一般基于3种方式进行模型构建.1)探索属性间的相关性^[10,13,17-19],如“裙子”和“女性”两个属性具有强相关性,这种关系可以为视觉属性提供约束推理;2)使用注意力机制^[20-21],在预测不同属性时,加强对该属性对应位置或特征的关注度.如预测属性“眼镜”时,加强对头部或面部的关注度;3)探索视觉上下文信息^[22-23],作为辅助属性识别的额外信息源,因为在同一场景下,不同的人有较大概率出现同一种属性,如婚礼中的人大多数正装出席.

围绕上述方法,在具体实现中,使用较多的是基于图像部件的方式.Gkioxari等^[23]和Zhang等^[24]将基于部件的模型和卷积神经网络(Convolutional Neural Networks, CNN)结合使用,训练姿态正则化的CNN进行属性分类.Zhu等^[5]提出多标签卷积神经网络(Multi-label CNN, MLCNN)模型,将行人图像划分为多个可重叠部分,为每一部分训练一个模型,选择相应部分进行属性预测.Li等^[25]提出结合

整个图像和图像部件的深度多属性识别(Deep Learning Based Multiple Attributes Recognition*, DeepMAR*),将行人图像及三种图像部件(头肩和上下身部分)共同送入模型进行训练.类似于DeepMAR*,Fabbri等^[26]将行人图像、肩部、上下身作为输入,利用不同部分的得分预测属性.Wang等^[27]提出基于CNN和循环神经网络(Recurrent Neural Network, RNN)序列化编解码的联合循环学习(Joint Recurrent Learning, JRL),将行人图像转换为区域序列,属性集转换为顺序的列表,利用长短期记忆网络(Long Short-Term Memory, LSTM)对属性进行有序预测.另一类行人属性识别方法仅基于整个行人图像进行建模.Li等^[13]和Sudowe等^[28]分别提出属性卷积网络(Attributes Convolutional Net, ACN)和深度多属性识别(Deep Learning Based Multiple Attributes Recognition, DeepMAR),两者都将整个行人图像作为模型输入,联合学习所有属性的预测.Liu等^[20]提出多方向注意力网络(Multi-directional and Attention-Based Net, HP-net),将深度网络的多层注意力特征图传送到不同的特征层,融合低层局部特征和高层信息,提升模型对较小尺度属性的预测能力.

虽然上述研究均取得不错进展,但这些方法将所有的属性放在同个网络分支中进行学习,忽略两个事实:1)不同属性间的学习难度不一样,2)属性所处的语义层次不一样,在一个分支中联合学习忽略属性间可能存在的抑制关系,导致模型对不同属性的预测能力彼此压制.为了解决这个问题,需要构建一种自适应模型,消减负相关性对属性识别的影响.本文提出基于深度学习的多阶段行人属性识别(Multi-stage Pedestrian Attribute Recognition, MPAR),同时探索属性间的正、负相关性,联合学习属性的识别.MPAR不依赖于行人的姿态对齐^[24]、图像块^[26,29]和上下文信息^[22-23],但能更好地探索属性之间的相关性,提高预测的准确率.此方法的动机来源于:在多属性学习中,不同属性间的学习存在相互促进和彼此抑制现象,具有促进作用的属性的学习会提高模型的预测性能,但某些属性的学习也可能会使模型对其它属性的预测能力变弱^[30].由此可知,更好地探索行人多个属性之间的正、负相关性是监控场景下行人属性识别的关键.

MPAR探索属性间的关系,基于深度网络模型VGG16,把除全连接层之外的网络层作为共享层,分三个阶段进行学习.第一阶段在共享层之后建立一个网络分支,联合学习所有属性的预测,并记录每个

属性的学习状态. 第二阶段根据第一阶段的学习状态,为学得不好的属性在共享层后单独建立一个网络分支进行学习,而其它属性仍然保留在原分支上,然后结合两个分支进行联合学习. 第三阶段在共享层后建立两个新的网络分支,结构和第二阶段的两个分支相同,然后迫使新网络分支对正、负样本预测的概率值的差距大于第二阶段. 三个阶段的学习仅利用图像及属性标签,无需额外操作.

另外,本文提出改进的损失函数,用于增大正负

样本预测概率值的差距,提高模型的判别力. 此损失函数可直接嵌入到三阶段的训练中,提高模型每一阶段的预测性能. 在两个公共数据集 RAP^[25] 和 PETA^[7] 上的实验结果表明本文方法性能较优.

1 多阶段行人属性识别

MPAR 框架建立在深度卷积神经网络 VGG16^[31] 上,如图 1 所示,共有三个阶段.

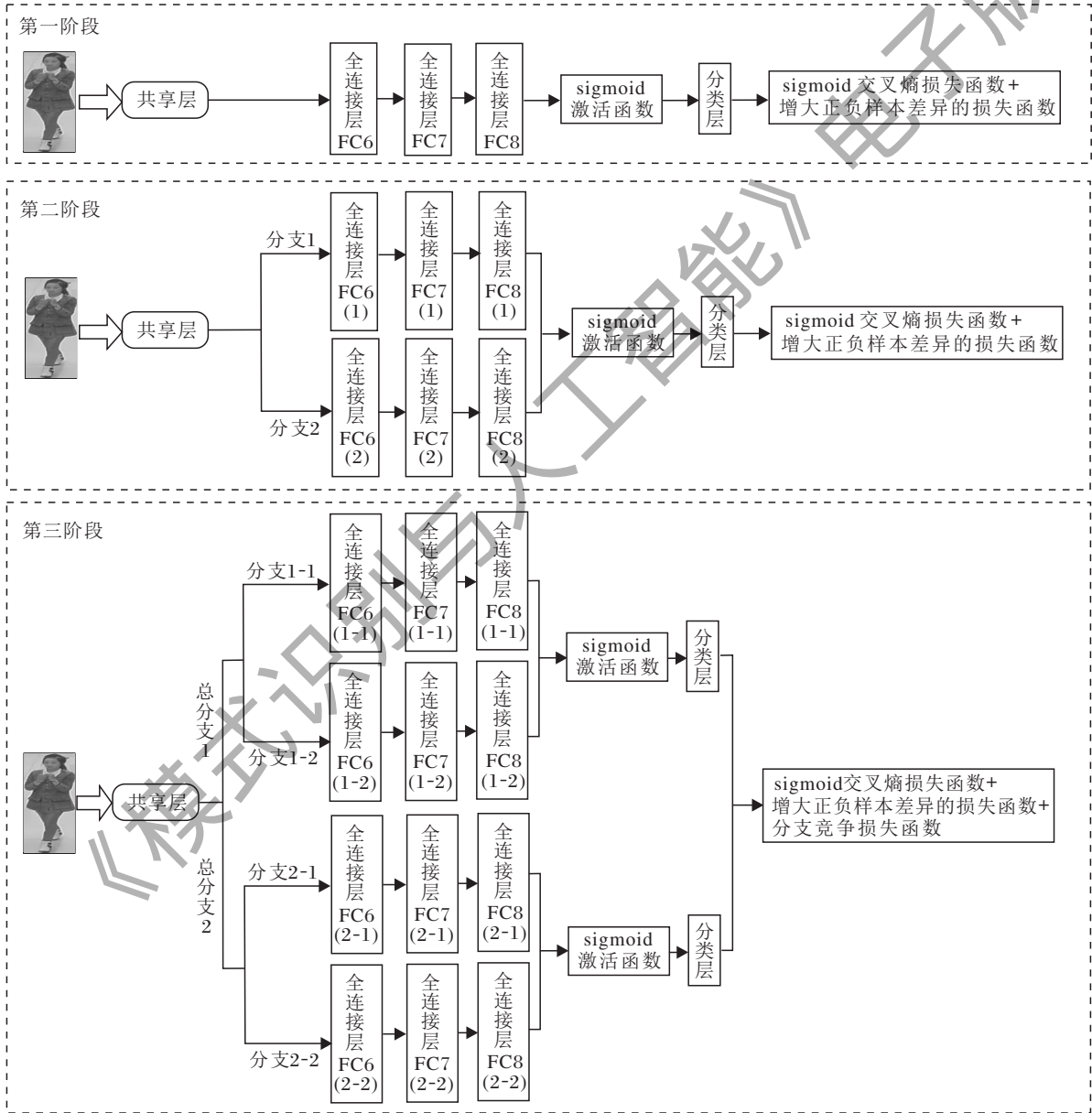


图 1 基于改进损失函数的多阶段行人属性识别网络结构图

Fig. 1 Architecture of multistage pedestrian attribute recognition based on improved loss function

1.1 第一阶段:属性正、负相关性的探索

第一阶段将网络最后一个全连接层的输出神经

元个数修改为行人属性的数目,网络的其它层仍然保持原有结构. 将最后一层全连接层的输出作为

sigmoid 激活函数的输入,经此激活函数输出每个属性的预测值.对于 RAP 和 PETA 数据集,最后的全连接层 FC8 的输出神经元个数分别为 51 和 35.

第一阶段联合学习所有的行人属性.受迁移学习提高模型泛化能力的启发^[32],使用 VGG16 在 ImageNet 数据集上训练得到的模型对网络进行微调.对于 FC8 层,将其参数初始化为满足均值为 0、标准差为 0.001 的截断正态分布的随机值,而其它层均采用预训练模型对参数进行初始化.

本文采用小批量 (Mini-Batch) 梯度下降法训练模型参数.第一阶段联合学习所有属性的预测,在训练过程中,记录每个小批次中的每个属性产生的平均损失值

$$L = \{l_{ij} | i = 1, 2, \dots, B; j = 1, 2, \dots, N\},$$

和每个属性的分类精度

$$D = \{d_{ij} | i = 1, 2, \dots, B; j = 1, 2, \dots, N\},$$

其中, B 表示模型训练的小批次数量, N 表示属性的数目.训练完成后,根据记录的损失值和精度值,得出每个属性在所有小批次训练集上的平均损失值

$$C_j = \frac{1}{B} \sum_{i=1}^B l_{ij},$$

和平均精度值

$$A_j = \frac{1}{B} \sum_{i=1}^B d_{ij},$$

作为第一阶段每个属性学习状态的两个指标,并为 2 个指标设置对应的阈值 t_c 和 t_a .然后把属性分成两组,平均损失值小于 t_c 且平均精度大于 t_a 的属性为一组,记为

$$G_1 = \{g_j | C_j < t_c, A_j > t_a\},$$

其余的属性共同组成另一组

$$G_2 = \{g_j | g_j \notin G_1\},$$

符号 g 表示属性集合.集合 G_1 和 G_2 内部元素具有正相关性的属性,在分类学习中能够相互促进,而 G_1 和 G_2 之间的属性具有抑制关系的属性,抑制对方分类性能的提升.至此,第一阶段的学习任务完成.

1.2 第二阶段:分组训练

第二阶段在共享层(全连接之前的网络层)后新建一个网络分支,如分支 2 所示.两分支的结构如下:

1) 分支 1 根据 G_1 建立,其全连接层 FC8(1) 的输出神经元数目等于 G_1 中的属性个数,用于学习 G_1 中属性的预测.

2) 分支 2 根据 G_2 建立,其全连接层 FC8(2) 输出神经元数目等于 G_2 中的属性个数,用于学习 G_2 中属性的预测.

3) 分支 2 的两个全连接层 FC6(2) 和 FC7(2) 的结构与分支 1 中的 FC6(1) 和 FC7(1) 相同.

将两个分支最后的全连接层进行连接操作,然后进行多属性的联合学习.使用第一阶段得到的模型微调第二阶段的共享层和两分支的前两个全连接层.全连接层 FC8(1) 和 FC8(2) 的参数初始化为满足均值为 0、标准差为 0.001 的截断正态分布的随机值.第二阶段的模型具有如下特点:

1) 两分支协同训练,削弱多属性联合学习中属性的抑制性.

2) 同一分支探索属性间的正相关性,不同分支抑制负相关性的影响.

3) 全连接层之外的浅层网络作为共享层,用于捕获两组属性间的共享信息,减少过拟合的风险.

1.3 第三阶段:竞争学习和强化训练

第三阶段在共享层后连接两个大分支:总分支 1 和总分支 2.两个总分支结构如下:

1) 总分支 1 连接两个小分支——总分支 1-1、总分支 1-2,总分支 2 连接总分支 2-1、总分支 2-2;

2) 总分支 1-1 和总分支 2-1 的结构与第二阶段的分支 1 相同,总分支 1-2 和总分支 2-2 的结构与第二阶段的分支 2 相同.

使用第二阶段训练得到的模型微调第三阶段的所有网络参数,保证总分支 1 和总分支 2 最初的预测能力一致.

此阶段的目的在于微调第二阶段训练得到的全连接层参数,强化网络的预测能力.采用的竞争训练方式如下:在保证共享层和总分支 1 的网络参数不变的基础上,训练总分支 2 的网络参数,使用分支竞争损失函数使总分支 2 的预测结果优于总分支 1,使总分支 2 的预测性能强于第二阶段.

模型的训练随着第三阶段的结束而完成.最终需要保留的参数为第三阶段的共享层参数及总分支 2 的网络层参数.综上所述,模型的三阶段算法步骤如下.

算法 MPAR 训练过程

输入 监控场景下行人图像,

对应的一系列属性标签

输出 对应的属性出现与否的预测概率值

step 1 第一阶段.

step 1.1 在训练过程中,记录每个小批次中每个属性的平均精度值 D 和平均损失值 L ,保留训练模型.

step 1.2 根据 D 和 L 计算每个属性的平均损失值 C 及平均精度值 A .

step 1.3 设置阈值 t_c 和 t_a , 得出两个属性分组 G_1 和 G_2 .

step 2 第二阶段.

step 2.1 根据第一阶段的两个属性组 G_1 和 G_2 , 添加网络分支并且修改每个分支最后的全连接层参数.

step 2.2 使用第一阶段的预训练模型对网络进行微调, 训练结束后保留模型参数.

step 3 第三阶段.

step 3.1 为网络另建两个和第二阶段同样结构的分支, 作为总分支 2.

step 3.2 通过分支间的竞争学习优化总分支 2 的预测性能.

step 3.3 保留共享层和总分支 2 的参数.

2 改进的损失函数

监控场景下的行人属性数据为二值属性, 相应属性出现为正样本, 否则为负样本, 如本文使用的公共数据集 RAP 和 PETA, 在此类数据集上的研究工作^[7, 20-21, 25-27]均围绕多个属性的二分类问题展开. 考虑到 logistic 回归函数在二分类问题上的有效性和便捷性, 本文模型宜采用基于 sigmoid 激活函数的交叉熵损失函数进行多属性的训练学习. 分类层的输出值表示相应属性出现的概率, 损失函数定义如下:

$$Loss = -\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (y_{ij} \ln(p_{ij}) + (1 - y_{ij}) \ln(1 - p_{ij})). \quad (1)$$

其中: M, N 分别表示每个小批次中的样本数目、属性数量; y_{ij} 为真值, 表示第 i 个样本的第 j 个属性是否出现, 出现为 1, 否则为 0; p_{ij} 表示每个小批次中的第 i 个样本的第 j 个属性的输出预测值,

$$p_{ij} = \frac{1}{1 + \exp(-x_{ij})},$$

x_{ij} 表示第 i 个样本的第 j 个属性在全连接层 FC8 上的输出值.

式(1)的损失函数统一考虑所有属性, 却忽略属性的正、负样本的分布不均衡问题. 在行人属性数据集(如 PETA 和 RAP)中, 大多数属性的正负样本分布不均匀, 式(1)作为损失函数会导致类别数目较多的样本贡献大部分的损失值, 使神经网络的梯度下降方向持续倾向于类别数目较多的样本能够被正确预测的方向, 导致类别数目较少的样本精度过低. 本文使用加权损失函数以解决此问题, 修改式(1), 得到如下损失函数:

$$Loss = -\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N (w_1 y_{ij} \ln(p_{ij}) + w_2 (1 - y_{ij}) \ln(1 - p_{ij})),$$

其中, $w_1 = \exp(1 - r_j)$, $w_2 = \exp(r_j)$ 表示权重参数, r_j 表示属性 j 的正样本比例^[13].

为了增加模型对正负样本的判别力度, 本文提出改进的损失函数:

$$Loss_{imp} = 1 - \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \alpha_{ij}, \quad (2)$$

$$\alpha = \begin{bmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1N} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2N} \\ \vdots & \vdots & & \vdots \\ \alpha_{M1} & \alpha_{M2} & \cdots & \alpha_{MN} \end{bmatrix}, \quad (3)$$

$$\alpha_{ij} = y_{ij} p_{ij} + (1 - y_{ij}) (1 - p_{ij}). \quad (4)$$

若第 i 个样本第 j 个属性为正例, 式(4)表示该属性的预测值与 0 的距离, 否则, 式(4)表示该属性的预测值与 1 的距离. 使用该距离构造式(3)所示的矩阵, 称为正负样本差异值矩阵. 该矩阵均值越大(小), 则模型对属性的预测值越接近(远离)真值. 将式(2)作为损失成分, 称为增大正负样本差异的损失函数, 迫使模型对属性的预测值更接近于真值. 第一、二阶段的损失函数包括两个损失成分:

$$Loss_{1-2} = Loss + Loss_{imp}.$$

以第 i 个样本的第 j 个属性为例分析 $Loss_{1-2}$ 特性. 假设 $Loss$ 中的权重 w_1 为 1, 真值 y_{ij} 为 1, 此时

$$Loss = -\ln p_{ij}, \quad Loss_{1-2} = -\ln p_{ij} + 1 - p_{ij}.$$

从损失值及梯度两方面分析损失函数 $Loss_{1-2}$ 的优势.

1) 图 2 为 $Loss$ 和 $Loss_{1-2}$ 两者损失曲线, 除 0 和 1 两端点之外, 均有 $Loss_{1-2} > Loss$, $Loss_{imp}$ 成分增大损失. 当 $p_{ij} = x_2$ 时, $y_1 > y_2$, $Loss_{1-2}$ 比 $Loss$ 产生更强烈的惩罚, 迫使参数的更新倾向于使 p_{ij} 更靠近于真值 1 的方向. 对于同一损失值 $y_1, x_2 > x_1$, $Loss_{1-2}$ 使预测值 p_{ij} 比 $Loss$ 更接近于 1.

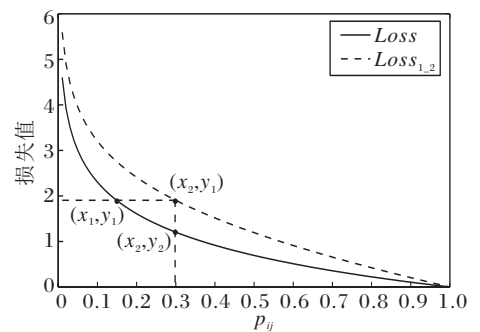


图 2 交叉熵和改进损失函数的损失值曲线图

Fig. 2 Curves of cross entropy and improved loss function

2) 两损失函数 $Loss$ 和 $Loss_{1,2}$ 对于 p_{ij} 的导数为

$$\frac{\partial Loss}{\partial p_{ij}} = -\frac{1}{p_{ij}}, \quad \frac{\partial Loss_{1,2}}{\partial p_{ij}} = -\left(\frac{1}{p_{ij}} + 1\right).$$

在两导数相等的情况下, $Loss_{1,2}$ 会使预测值 p_{ij} 比 $Loss$ 更接近于 1, 利于模型的快速收敛。

模型第三阶段的训练目的在于强化网络的预测能力, 使总分支 2 的预测性能优于总分支 1. 设计分支之间竞争学习的损失函数:

$$Loss_{com} = \frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N \beta_{ij}, \quad (5)$$

$$\beta = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1N} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2N} \\ \vdots & \vdots & & \vdots \\ \beta_{M1} & \beta_{M2} & \cdots & \beta_{MN} \end{bmatrix},$$

$$\beta_{ij} = -\min\{(y_{ij}p_{ij}^2 + (1 - y_{ij})(1 - p_{ij}^2)) - (y_{ij}p_{ij}^1 + (1 - y_{ij})(1 - p_{ij}^1)), 0\},$$

其中, p_{ij}^1 和 p_{ij}^2 表示总分支 1 和总分支 2 对第 i 个样本的第 j 个属性的预测值. 若 $\beta_{ij} > 0$, 表示总分支 2 对该属性的判别力弱于总分支 1, 若 $\beta_{ij} = 0$, 表示总分支 2 对该属性的判别力强于总分支 1. 将 β 的均值作为损失函数值, 称式(5) 为分支竞争损失函数.

第三阶段使用的损失函数具有三个成分:

$$Loss_{1,2,3} = Loss + Loss_{imp} + Loss_{com}.$$

以第 i 个样本的第 j 个属性为例说明 $Loss_{1,2,3}$ 的特性. 假设 $Loss$ 中的权重 w_1 为 1, 真值 y_{ij} 为 1, p_{ij}^1 为 0.9, 该损失函数的分析如下.

1) 若总分支 2 的预测优于总分支 1, 则 $p_{ij}^2 > p_{ij}^1$,

$$Loss_{com} = -\min\{p_{ij}^2 - p_{ij}^1, 0\} = 0.$$

总分支 2 损失函数为

$$Loss_{1,2,3} = -\ln(p_{ij}^2) + 1 - p_{ij}^2,$$

此时仅使用该损失函数进行参数更新.

2) 若总分支 2 的预测弱于总分支 1, 则 $p_{ij}^2 < p_{ij}^1$,

$$Loss_{com} = p_{ij}^1 - p_{ij}^2 > 0,$$

此时总分支 2 的损失函数为

$$L_1 = -\ln(p_{ij}^2) + 1 - 2p_{ij}^2 + p_{ij}^1.$$

若无 $Loss_{com}$ 成分, 则总分支 2 损失函数为

$$L_2 = -\ln(p_{ij}^2) + 1 - p_{ij}^2.$$

图 3 为 L_1 和 L_2 的曲线图, 除端点 0 和 1 外, 均有 $L_1 > L_2$, $Loss_{com}$ 成分增大损失. 当 $p_{ij}^2 = x_2$ 时, $y_1 > y_2$, L_1 比 L_2 产生更强烈的惩罚, 迫使参数的更新倾向于使 p_{ij}^2 更靠近于真值 1 的方向. 对于同一损失值 y_1 , $x_2 > x_1$, $Loss_{1,2,3}$ 使总分支 2 对第 i 个样本第 j 个属性的预测值更接近于 1. 另外, L_1 和 L_2 对 p_{ij}^2 的导数分别为

$$\frac{\partial L_1}{\partial p_{ij}^2} = -\left(\frac{1}{p_{ij}^2} + 2\right), \quad \frac{\partial L_2}{\partial p_{ij}^2} = -\left(\frac{1}{p_{ij}^2} + 1\right).$$

如果两者相等, 前者会使预测值 p_{ij}^2 更接近于 1, 有利于提升模型对样本的预测力度.

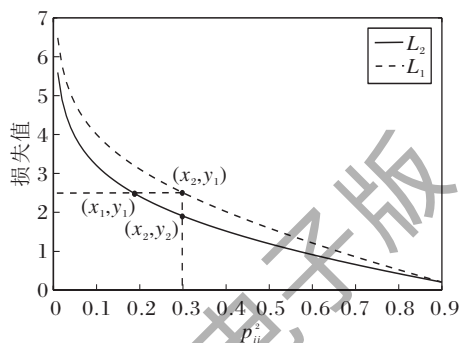


图 3 有无 $Loss_{com}$ 成分损失值曲线图

Fig. 3 Loss value curves with and without $Loss_{com}$ component

3 实验及结果分析

3.1 实验数据集

本文方法在 2 个大规模行人属性数据集 (RAP^[25] 和 PETA^[7]) 上进行验证.

RAP 数据集^[25] 来源于 26 个摄像头收集的室内的监控场景, 共有 41 585 幅图像, 标注信息十分丰富, 每幅图像有 72 个标签, 其中有 69 个二值标签和 3 个多值标签. 训练集和测试集采用 Li 等^[25] 提出的划分. 33 268 幅图像用于训练, 8 317 幅图像用于测试. 采用 Li 等^[25] 提出的 51 个二值属性进行实验评估.

PETA 数据集^[7] 来源于 10 个小规模行人再识别的数据集. 包含 19 000 幅图像, 每幅图像有 65 个属性标注, 其中有 61 个二值属性和 4 个多值属性. 根据 Deng 等^[7] 提出的划分, 随机选择 7 600 幅图像用于测试, 其它图像用于训练, 使用 35 个属性进行评估.

3.2 参数设置

使用谷歌开源深度学习框架 Tensorflow 建立模型, 网络模型使用 VGG16, 输入为 RGB 三通道图像, 尺寸缩放至 224×224 . 对于 RAP 数据集, 仅对图像进行缩放处理, 采用双线性插值法. 对于 PETA 数据集, 由于训练图像过少, 对数据进行增广操作, 采用左右反转及图像色彩抖动, 包括饱和度调整、亮度调整、对比度调整等处理方式.

对于 RAP 和 PETA 数据集, 第一阶段至第三阶

段的训练迭代次数分别设置为 20 000、10 000、6 000 和 10 000、6 000、2 000。在每个阶段,学习率随着训练次数而变化,图 4 为两个数据集上各阶段的学习率变化情况。

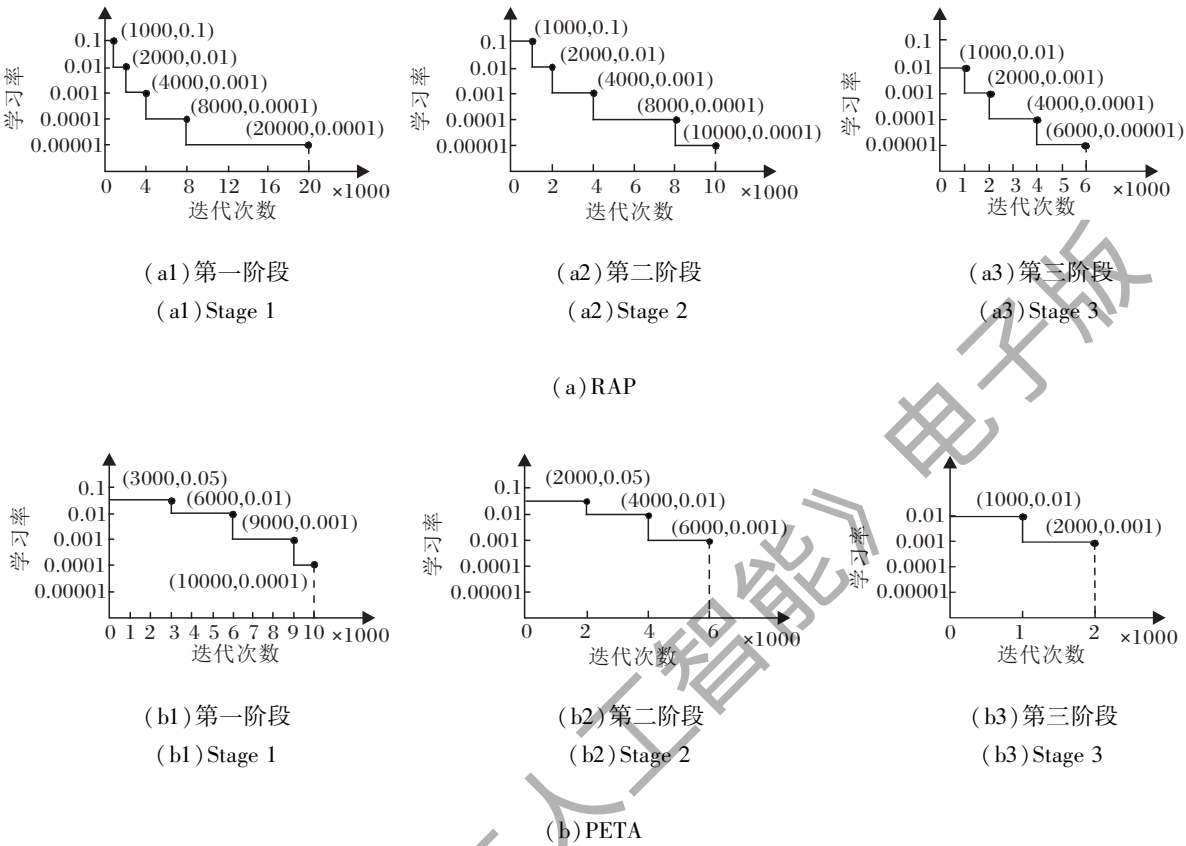


图 4 2 个数据集上各训练阶段的学习率变化
Fig. 4 Different learning rates in each training stage on 2 datasets

在不同的训练阶段,每次迭代读取的图像数目(batch_size)不同.对于 2 个数据集,同一阶段的迭代读取的图像数目相同,第一、第二和第三阶段的 batch_size 分别为 83、80 和 110。

对于 RAP 和 PETA 数据集,第一阶段设置的两个阈值 T_c 和 T_a 分别为 0.35、0.928 和 0.1、0.975。在三阶段的训练中,每阶段使用的优化方法相同,均采用随机梯度下降法(Stochastic Gradient Descent, SGD)。VGG16 网络结构采用丢弃(dropout)层随机选择前一层的部分神经元进行后续训练,防止模型过拟合。在三阶段的训练中,对于 dropout 层,均随机选择前一层的 50% 的神经元进行后续层的训练。

3.3 评价指标

采用如下 2 种评价方式对模型性能进行评估。
1) Deng 等^[7]提出基于标签的评价方式,使用平均正确率作为评价指标:

$$mA = \frac{1}{2N} \sum_{i=1}^N \left(\frac{TP_i}{P_i} + \frac{TN_i}{N_i} \right).$$

其中: N 表示属性的数目; $|TP_i|$ 和 $|TN_i|$ 分别表示第 i 个属性被正确预测的正、负样本的数目; $|P_i|$ 和 $|N_i|$ 分别表示对于第 i 个属性,正、负样本的数目。

2) Li 等^[25]提出基于样本的评价方式,共有 4 个评价指标,分别为正确率、精度、召回率和 $F1$ 值:

$$\begin{aligned} Acc_{\text{exam}} &= \frac{1}{N} \sum_{i=1}^N \frac{|Y_i \cap f(x_i)|}{|Y_i \cup f(x_i)|}, \\ Prec_{\text{exam}} &= \frac{1}{N} \sum_{i=1}^N \frac{|Y_i \cap f(x_i)|}{|f(x_i)|}, \\ Rec_{\text{exam}} &= \frac{1}{N} \sum_{i=1}^N \frac{|Y_i \cap f(x_i)|}{|Y_i|}, \\ F1 &= \frac{2 \times Prec_{\text{exam}} \times Rec_{\text{exam}}}{Prec_{\text{exam}} + Rec_{\text{exam}}}, \end{aligned}$$

其中, N 表示样本数目, Y_i 表示第 i 个样本真值中的正例标签, $f(x_i)$ 表示对第 i 个样本预测标签中的正例标签, $|\cdot|$ 表示集合基数。

3.4 对比方法介绍

对比方法中包括 2 种传统判别式模型. 1) 使用局部特征集合 (Ensemble of Localized Features, ELF) 和深度网络模型 AlexNet 提取全连接层 FC7 的特征, 使用 SVM 判别器进行训练, 形成 2 种对比方法: ELF+SVM 和 CNN+SVM. 2) Deng 等^[11] 提出带有随机森林的马尔科夫随机场模型 (Markov Random Forest, MRFr2), 为基于图的属性识别方法, 经马尔科夫随机场探索相邻行人图像的上下文信息, 挖掘不同图像间的视觉外观相似性进行属性推理.

剩余 10 种对比方法均为基于深度学习的方法. 1) Sudowe 等^[28] 提出 ACN, 所有属性联合训练, 为每个属性单独设置一个损失函数, 所有损失函数的加权平均作为总的损失值. 2) Li 等^[13] 提出深度单属性识别 (Deep Learning Based Single Attribute Recognition, DeepSAR) 和 DeepMAR, 前者单独训练每个属性, 后者将所有属性进行联合训练, 捕获属性间的相关性. 3) Li 等^[13] 在评估行人属性数据集时, 提出基于部件的模型 DeepMAR*, 将行人分为头肩、上、下身三部分, 结合整个行人图像, 使用 DeepMAR 进行训练. 4) Yu 等^[21] 提出弱监督行人属性定位 (Weakly-Supervised Pedestrian Attribute Localization, WPAL), 为弱监督属性定位框架. 基于 GoogleNet 网络, 设计检测层提取中级属性特征, 并使用基于最大池化的弱监督目标检测技术对模型进行训练, 最后回归这些检测响应幅度, 进行属性预测. 5) Wang 等^[27] 提出 JRL, 使用循环神经网络探索同一行人所有属性的关系、不同行人之间的属性相关性及特征相似的行人之间的上下文关系. 6) HP-net^[20] 为基于注意力机制的深度模型, 采用 Inception-V2 网络结构, 将多层注意力图多向映射到不同层的网络块, 使模型可以融合全局和局部信息. 7) Li 等^[33] 提出序列化的预测方式, 使用基于图像上下文信息的模型 (Contextual CNN-RNN, CTX CNN-RNN) 编码场景上下文信息, 并且探索行人之间的关系, 为同一幅图像中的多个行人建模. 8) Liu 等^[34] 提出语义正则化的模型 (Semantically Regularised CNN-RNN, SR CNN-RNN), 用于处理多标签图像分类问题, 选择更鲁棒的图像特征.

3.5 实验结果

各方法在 2 个数据集上的整体结果分析如表 1 和表 2 所示. 从表中可以看出, 对于传统方法, MPAR 精度远不如基于深度学习的方法, 这得益于深度学习方法, 尤其是深度卷积神经网络在图像分类上强

大的非线性拟合能力. 在所有对比方法中, ELF + SVM、CNN + SVM 及 DeepSAR 单独为每个属性设计分类器, 未考虑属性之间的关系, 而其它方法均联合训练所有属性, 而属性的联合学习可以提高模型预测精度. 在 2 个数据集上, MPAR 均有 3 个指标的精度达到最高值, 但 mA 指标精度相对不高, 说明 MPAR 更倾向于尽可能多地正确预测每个样本的所有属性, 而对于所有样本的同个属性的预测, 整体性能稍微偏弱.

表 1 RAP 数据集上不同方法的性能对比

Table 1 Performance comparison of different methods on RAP dataset

方法	<i>mA</i>	<i>Acc_{exam}</i>	<i>Prec_{exam}</i>	<i>Rec_{exam}</i>	<i>F1</i>
ELF+SVM ^[35]	69.94	29.29	32.84	71.18	44.49
CNN+SVM ^[25]	72.28	31.72	35.75	71.78	47.73
MRFr2 ^[11]	—	—	—	—	—
ACN ^[28]	69.66	62.61	80.12	72.26	75.98
DeepSAR ^[13]	—	—	—	—	—
DeepMAR ^[13]	73.79	62.02	74.92	76.21	75.56
DeepMAR* ^[25]	74.44	63.67	76.53	77.47	77.00
WPAL ^[21]	79.48	53.30	60.82	78.80	68.65
JRL ^[27]	77.81	—	78.11	78.98	78.58
HP-net ^[20]	76.12	65.39	77.33	78.79	78.05
CTX CNN-RNN ^[33]	70.13	—	71.03	71.20	70.23
SR CNN-RNN ^[34]	74.21	—	75.11	76.52	75.83
MPAR	75.64	66.17	78.29	79.23	78.76

表 2 PETA 数据集上不同方法的性能对比

Table 1 Performance comparison of different methods on PETA dataset

方法	<i>mA</i>	<i>Acc_{exam}</i>	<i>Prec_{exam}</i>	<i>Rec_{exam}</i>	<i>F1</i>
ELF+SVM ^[35]	75.21	43.68	49.45	74.24	59.36
CNN+SVM ^[25]	76.65	45.41	51.33	75.14	61.00
MRFr2 ^[11]	75.60	—	—	—	—
ACN ^[28]	81.15	73.66	84.06	81.26	82.64
DeepSAR ^[13]	81.30	—	—	—	—
DeepMAR ^[13]	82.89	75.07	83.68	83.14	83.41
DeepMAR* ^[25]	—	—	—	—	—
WPAL ^[21]	85.50	76.98	84.07	85.78	84.90
JRL ^[27]	85.67	—	86.03	85.34	85.42
HP-net ^[20]	81.77	76.13	84.92	83.24	84.07
CTX CNN-RNN ^[33]	80.13	—	79.68	80.24	79.68
SR CNN-RNN ^[34]	82.83	—	82.54	82.76	82.65
MPAR	83.57	77.80	86.19	85.07	85.63

MPAR 分为三个训练阶段, 每个阶段 (除第一阶段以外) 的属性预测性能都优于前一阶段. 表 3 为每一阶段得到的预测结果, 可以看出, 除第一阶段外,

每一阶段较前一阶段总体上均有三个或三个以上的指标精度有所提升. 另外,改进损失函数对模型性能的提升具有促进作用. 表 4 为不使用改进损失函数的精度. 对比表 3 和表 4 可以看出,改进的损失函数使三个或三个以上的指标精度更高.

表 1 表明 MPAR 在 RAP 数据集上表现优越. 对比其它方法,在 Acc_{exam} 、 Rec_{exam} 和 $F1$ 上均获得最高精度 66.17%、79.23% 和 78.76%. 在 mA 和 $Prec_{exam}$ 上获得 75.64% 和 78.29% 的精度,其中在 $Prec_{exam}$ 上的精度表现为次优,略低于 ACN 在 $Prec_{exam}$ 上取得的最高精度 (80.12%),这说明 ACN 在预测中相对保守,仅预测置信度非常高的属性.

表 3 使用改进损失函数时不同训练阶段得到的精度结果
Table 3 Accuracies obtained in different training stages with improved loss function

评价指标	RAP			PETA		
	第一阶段	第二阶段	第三阶段	第一阶段	第二阶段	第三阶段
mA	73.87	76.38	75.64	84.02	83.80	83.57
Acc_{exam}	64.42	65.51	66.17	77.47	77.70	77.80
Rec_{exam}	79.38	80.06	79.23	85.23	85.21	85.07
$Prec_{exam}$	75.84	76.69	78.29	85.51	85.88	86.19
$F1$	77.57	78.34	78.76	85.37	85.54	85.63

表 4 不使用改进损失函数时不同训练阶段得到的精度结果
Table 4 Accuracies obtained in different training stages without improved loss function

评价指标	RAP			PETA		
	第一阶段	第二阶段	第三阶段	第一阶段	第二阶段	第三阶段
mA	75.21	77.69	76.62	84.25	83.99	83.57
Acc_{exam}	63.19	64.37	65.55	77.12	77.54	77.74
Rec_{exam}	80.80	81.17	80.10	85.36	85.21	84.97
$Prec_{exam}$	73.06	74.21	76.61	84.97	85.64	86.20
$F1$	76.73	77.54	78.32	85.16	85.42	85.58

总之,对于 RAP 数据集,MPAR 在基于属性的评价标准中稍显不足,但在基于样本的评价中表现优异.

表 2 表明 MPAR 在 PETA 数据集上取得较优性能. 在 12 种对比方法中,在 Acc_{exam} 、 $Prec_{exam}$ 和 $F1$ 上均取得最高精度,分别为 77.80%、86.19%、85.63%. 但在 mA 和 $Prec_{exam}$ 上表现不足,分别取得 83.57% 和 85.07% 的精度,位列第三,略低于最高精度 85.67% 和 85.78%. 总之,对于 PETA 数据集,MPAR 在基于属性的评价标准中稍显不足,但在基于样本的评价中表现优异.

对多阶段的训练结果进行精度统计的结果如表 3 和表 4 所示. 对于 RAP 数据集,无论加入改进损失函数与否,除第一阶段外,相比前一阶段,每阶段至少有三个评价指标的精度有所提升,其中第二阶段所有指标精度均高于第一阶段. 在第三阶段,模型在 Acc_{exam} 、 $Prec_{exam}$ 和 $F1$ 上的精度均高于第二阶段. 对于 PETA 数据集,无论加入损失函数与否,在第二、第三阶段的结果中,在 Acc_{exam} 、 $Prec_{exam}$ 和 $F1$ 上的精度均高于前一阶段. 而对于 mA 、 Rec_{exam} ,第三阶段的性能稍逊于第二阶段. 分析原因如下:第三阶段的训练目的在于把属性预测的置信度增大,即正例标签的预测概率值更接近于 1,而负例标签的预测概率值更接近于 0. 但是对于训练集来说,属性的正、负样本数目严重失衡,部分属性的正例样本过少,使模型的预测更易于提升负例样本的置信度,而对正例样本的预测产生一定的负面影响,使模型对部分正例样本过少的属性的预测性能降低. 但总体上 MPAR 对于提升行人属性的预测能力具有明显效果.

4 结束语

目前,监控场景下行人属性识别的研究在挖掘属性的相关性方面存在不足,尤其缺乏对行人属性间负相关性的探索工作. 为了解决这个问题,本文提出基于深度卷积神经网络的多阶段行人属性识别方法,分不同阶段探索行人属性之间的关系,包括相互促进的正相关性及彼此抑制的负相关性. 在训练过程中把具有抑制作用的属性分成不同组,结合增大正负样本差异的改进损失函数进行多阶段训练,提升模型的预测性能. 在不同数据集上的实验表明,本文方法的预测性能较优.

然而,本文方法在基于属性的评价准则上性能稍逊,下一步计划探索行人与行人之间的相关性,利用行人之间的约束,对不同属性加以推理,提高模型对单个属性的预测精度. 另外,本文方法在第三阶段的预测中部分指标弱于第二阶段,这是数据集中正、负样本分布不均衡所致,下一步打算通过产生式对抗学习挖掘样本数据的分布规则,生成需要的样本,达到增广样本数据的目的.

参 考 文 献

[1] KUMAR N, BERG A, BELHUMEUR P N, et al. Describable Visual Attributes for Face Verification and Image Search. IEEE Tran-

- sactions on Pattern Analysis and Machine Intelligence, 2011, 33(10): 1962–1977.
- [2] LIU Z W, LUO P, WANG X G, *et al.* Deep Learning Face Attributes in the Wild // Proc of the IEEE International Conference on Computer Vision. Washington, USA: IEEE, 2015: 3730–3738.
 - [3] LUO P, WANG X G, TANG X O. A Deep Sum-Product Architecture for Robust Facial Attributes Analysis // Proc of the IEEE International Conference on Computer Vision. Washington, USA: IEEE, 2013: 2864–2871.
 - [4] LAYNE R, HOSPEDALES T M, GONG S G. Person Re-identification by Attributes // Proc of the 23rd British Machine Vision Conference. Berlin, Germany: Springer, 2012. DOI: 10.5244/C.26.24.
 - [5] ZHU J Q, LIAO S C, YI D, *et al.* Multi-label CNN Based Pedestrian Attribute Learning for Soft Biometrics // Proc of the International Conference on Biometrics. Washington, USA: IEEE, 2015: 535–540.
 - [6] REID D A, NIXON M S, STEVENAGE S V. Soft Biometrics; Human Identification Using Comparative Descriptions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2014, 36(6): 1216–1228.
 - [7] DENG Y B, LUO P, LOY C C, *et al.* Pedestrian Attribute Recognition at Far Distance // Proc of the 22nd ACM International Conference on Multimedia. New York, USA: ACM, 2014: 789–792.
 - [8] JAHA E S, NIXON M S. Soft Biometrics for Subject Identification Using Clothing Attributes // Proc of the IEEE International Joint Conference on Biometrics. Washington, USA: IEEE, 2014. DOI: 10.1109/BTAS.2014.6996278.
 - [9] LIU C X, GONG S G, LOY C C, *et al.* Person Re-identification: What Features Are Important? // Proc of the European Conference on Computer Vision. Washington, USA: Springer, 2012: 391–401.
 - [10] CHEN H Z, GALLAGHER A, GIROD B. Describing Clothing by Semantic Attributes // Proc of the European Conference on Computer Vision. Washington, USA: Springer, 2012: 609–623.
 - [11] DENG Y B, LUO P, LOY C C, *et al.* Learning to Recognize Pedestrian Attribute [C/OL]. [2018–08–25]. <https://arxiv.org/pdf/1501.00901.pdf>.
 - [12] SHI Z Y, HOSPEDALES T M, XIANG T. Transferring a Semantic Representation for Person Re-identification and Search // Proc of the IEEE International Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2015: 4184–4193.
 - [13] LI D W, CHEN X T, HUANG K Q. Multi-attribute Learning for Pedestrian Attribute Recognition in Surveillance Scenarios // Proc of the 3rd IARP Asian Conference on Pattern Recognition. Washington, USA: IEEE, 2015: 111–115.
 - [14] DONG Q, GONG S G, ZHU X T. Class Rectification Hard Mining for Imbalanced Deep Learning // Proc of the IEEE International Conference on Computer Vision. Washington, USA: IEEE, 2017: 1851–1860.
 - [15] DONG Q, GONG S G, ZHU X T. Multi-task Curriculum Transfer Deep Learning of Clothing Attributes // Proc of the IEEE Winter Conference on Applications of Computer Vision. Washington, USA: IEEE, 2017: 520–529.
 - [16] PENG P X, TIAN Y H, XIANG T, *et al.* Joint Semantic and Latent Attribute Modelling for Cross-Class Transfer Learning. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(7): 1625–1638.
 - [17] ZHU J Q, LIAO S C, LEI Z, *et al.* Multi-label Convolutional Neural Network Based Pedestrian Attribute Classification. Image and Vision Computing, 2017, 58: 224–229.
 - [18] WANG J Y, ZHU X T, GONG S G. Video Semantic Clustering with Sparse and Incomplete Tags // Proc of the 30th AAAI Conference on Artificial Intelligence. Palo Alto, USA: AAAI Press, 2016: 3618–3624.
 - [19] WANG J Y, ZHU X T, GONG S G. Discovering Visual Concept Structure with Sparse and Incomplete Tags. Artificial Intelligence, 2017, 250: 16–36.
 - [20] LIU X H, ZHAO H Y, TIAN M Q, *et al.* HydraPlus-Net: Attentive Deep Features for Pedestrian Analysis // Proc of the IEEE International Conference on Computer Vision. Washington, USA: IEEE, 2017: 350–359.
 - [21] YU K, LENG B, ZHANG Z, *et al.* Weakly-Supervised Learning of Mid-level Features for Pedestrian Attribute Recognition and Localization [C/OL]. [2018–08–25]. <https://arxiv.org/pdf/1611.05603.pdf>.
 - [22] LI Y N, HUANG C, LOY C C, *et al.* Human Attribute Recognition by Deep Hierarchical Contexts // Proc of the European Conference on Computer Vision. Washington, USA: Springer, 2016: 684–700.
 - [23] GKIOXARI G, GIRSHICK R, MALIK J. Contextual Action Recognition with R* CNN // Proc of the IEEE International Conference on Computer Vision. Washington, USA: IEEE, 2015: 1080–1088.
 - [24] ZHANG N, PALURI M, RANZATO M A, *et al.* PANDA: Pose Aligned Networks for Deep Attribute Modeling // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2014: 1637–1644.
 - [25] LI D W, ZHANG Z, CHEN X T, *et al.* A Richly Annotated Dataset for Pedestrian Attribute Recognition [C/OL]. [2018–08–25]. <https://arxiv.org/pdf/1603.07054.pdf>.
 - [26] FABBRI M, CALDERARA S, CUCCHIARA R. Generative Adversarial Models for People Attribute Recognition in Surveillance [C/OL]. [2018–08–25]. <https://arxiv.org/pdf/1707.02240.pdf>.
 - [27] WANG J Y, ZHU X T, GONG S G, *et al.* Attribute Recognition by Joint Recurrent Learning of Context and Correlation [C/OL]. [2018–08–25]. <https://arxiv.org/pdf/1709.08553.pdf>.
 - [28] SUDOWE P, SPITZER H, LEIBE B. Person Attribute Recognition with a Jointly-Trained Holistic CNN Model // Proc of the IEEE International Conference on Computer Vision. Washington, USA: IEEE, 2015: 87–95.
 - [29] SUDOWE P, LEIBE B. Patchit: Self-supervised Network Weight Initialization for Fine-Grained Recognition // Proc of the British Machine Vision Conference. Berlin, Germany: Springer, 2016: 75.1–75.12.

[30] HAND E M, CHELLAPPA R. Attributes for Improved Attributes: A Multi-task Network for Attribute Classification [C/OL]. [2018-08-25]. <https://arxiv.org/pdf/1604.07360.pdf>.

[31] SIMONYAN K, ZISSERMAN A. Very Deep Convolutional Networks for Large-Scale Image Recognition [C/OL]. [2018-08-25]. <https://arxiv.org/pdf/1409.1556.pdf>.

[32] YOSINSKI J, CLUNE J, BENGIO Y, *et al.* How Transferable Are Features in Deep Neural Networks? [C/OL]. [2018-08-25]. <https://arxiv.org/pdf/1411.1792.pdf>.

[33] LI Y, LIN G S, ZHUANG B H, *et al.* Sequential Person Recognition in Photo Albums with a Recurrent Network // Proc of the IEEE Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2017: 1338-1346.

[34] LIU F, XIANG T, HOSPEDALES T M, *et al.* Semantic Regularisation for Recurrent Image Annotation // Proc of the IEEE International Conference on Computer Vision and Pattern Recognition. Washington, USA: IEEE, 2017: 2872-2880.

[35] LAYNE R, HOSPEDALES T M, GONG S G. Attributes-Based Re-identification // GONG S G, CRISTANI M, YAN S C, *et al.*, eds. Person Re-identification. Berlin, Germany: Springer, 2014: 93-117.

作者简介



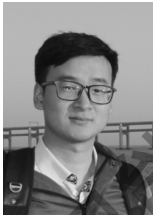
郑少飞, 硕士研究生, 主要研究方向为计算机视觉、机器学习. E-mail: shaofeizheng@foxmail.com.
(ZHENG Shaofei, master student. His research interests include computer vision and machine learning.)



汤进(通讯作者), 博士, 教授, 主要研究方向为计算机视觉、大数据处理. E-mail: jtang99029@foxmail.com.
(TANG Jin (Corresponding author), Ph. D., professor. His research interests include computer vision and big data processing.)



罗斌, 博士, 教授, 主要研究方向为模式识别、图匹配理论. E-mail: luobin@ahu.edu.cn.
(LUO Bin, Ph. D., professor. His research interests include pattern recognition and graph matching theory.)



王逍, 博士研究生, 主要研究方向为计算机视觉、模式识别. E-mail: Wangxiaocvpr@foxmail.com.
(WANG Xiao, Ph. D. candidate. His research interests include computer vision and pattern recognition.)



王文中, 博士, 讲师, 主要研究方向为计算机视觉、机器学习. E-mail: wenzhong@ahu.edu.cn.
(WANG Wenzhong, Ph. D., lecturer. His research interests include computer vision and machine learning.)

2018 中国自动化大会在古城西安盛大开幕

2018 年 12 月 1 日至 2 日, 由中国自动化学会(以下简称学会)主办, 西安交通大学承办的 2018 中国自动化大会在陕西省西安市盛大开幕。围绕“自动化创造智慧社会”会议主题, 大会设有与自动化、信息与智能科学技术相关的 7 个大会报告、200 余个专题特邀报告、180 余个会议论文报告以及 800 余篇张贴论文, 围绕自动化未来发展, 设有 1 个大会圆桌论坛和 2 个专题的圆桌论坛。来自海内外自动化、电气信息技术、先进制造、人工智能及相关领域的 13 位院士, 200 余位长江学者、国家杰青、千人计划学者等自动化领域专家学者们欢聚一堂, 共探自动化产业的未来发展。

大会开幕式由学会副理事长兼秘书长、中国科学院自动化研究所王飞跃研究员主持。学会理事长郑南宁院士、西安交通大学校长王树国教授分别致辞。

中国工程院周济院士带来开场报告《以智能制造为主攻方向, 加快建设制造强国》。中国工程院孙优贤院士作题为《大型高炉高性能运行控制与优化》的报告。任教于香港科技大学和上海交通大学的曹希仁教授为大会带来报告《Relative Optimization, Policy Iteration, and Reinforcement Learning》(《相对优化, 策略迭代以及加强学习》)。

中国自动化学会第四批会士荣誉称号授予仪式也在大会首日重磅举行。作为学会会员在学会的最高学术荣誉, 本次中国自动化学会会士共授予了 22 位自动化、信息与智能科学领域的杰出科技工作者, 以表彰其在上述领域的卓越成就和为学会做出的重大贡献。