



基于多级注意力跳跃连接网络的行人属性识别

王 林, 李聪会

(西安理工大学 自动化与信息工程学院, 西安 710048)

摘要: 行人属性识别易受视角、尺度和光照等非理想自然条件变化的影响,且某些细粒度属性识别难度较大。为此,提出一种多级注意力跳跃连接网络 MLASC-Net。在网络中间层,利用敏感注意力模块在通道及空间维度上对原特征向量进行筛选加权,设计多级跳跃连接结构来融合所提取的显著性特征。在网络顶层,改进多尺度金字塔池化以融合局部特征和全局特征。在网络输出层,结合验证损失算法自适应更新损失层,加速模型的收敛并提高精度。在 PETA 和 RAP 数据集上的实验结果表明,MLASC-Net 的识别准确率相较原基准网络分别提高约 4.62 和 6.54 个百分点,其在识别效果和模型收敛速度上有明显优势,同时在非理想自然条件下具有良好的泛化能力,可有效提高网络对细粒度属性的鲁棒性。

关键词: 行人属性识别;多级跳跃连接网络;敏感注意力;多尺度金字塔;残差网络

开放科学(资源服务)标志码(OSID):



中文引用格式: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.

英文引用格式:WANG Lin, LI Conghui. Pedestrian attribute recognition based on multi-level attention skip connection network[J]. Computer Engineering, 2021, 47(2): 314-320.

Pedestrian Attribute Recognition Based on Multi-Level Attention Skip Connection Network

WANG Lin, LI Conghui

(School of Automation and Information Engineering, Xi'an University of Technology, Xi'an 710048, China)

[Abstract] Pedestrian attribute recognition is susceptible to changes in non-ideal natural conditions, including perspective, scale and lighting, which increases the difficulty in fine-grained attribute recognition. To address the problem, this paper proposes a multi-level attention skip connection network, MLASC-Net. In the middle layer of the network, the sensitive attention module is used to filter and weight the original feature vectors in the channel and spatial dimensions, and a multi-level skip connection structure is designed to fuse the extracted salient features. In the top layer of the network, multi-scale pyramid pooling is improved to integrate local features and global features. In the output layer of the network, the verification of the loss trend algorithm is used to update the loss layer adaptively, accelerate the convergence of the model and increase accuracy. Experimental results on PETA and RAP datasets show that compared with the original benchmark network, MLASC-Net improves the recognition accuracy by about 4.62 and 6.54 percentage points, respectively. It has obvious advantages in recognition effect and model convergence speed, and has good generalization ability under non-ideal natural conditions, which can effectively improve the robustness of the network to fine-grained attributes.

[Key words] pedestrian attribute recognition; multi-level skip connection network; sensitive attention; multi-scale pyramid; Residual Network (ResNet)

DOI: 10.19678/j.issn.1000-3428.0057107

0 概述

行人属性识别是智能监控领域的研究热点之一,其主要任务是识别监控图像或视频中的行人视觉属性,如性别、年龄和着装等语义信息,可应用于安防监控、客群分析和实时广告精准投放等诸多任

务,具有广阔的市场应用前景。传统的行人属性识别方法侧重于从手工特征、分类器模型等角度来挖掘更具鲁棒性的特征表达,这类研究方法大多为各属性训练单独的二值分类模型,依赖于研究人员的经验,而忽略了属性间的语义关系,难以适应场景的复杂性和属性特征的多样性。

基金项目:陕西省科技计划重点项目(2017ZDCXL-GY-05-03)。

作者简介:王 林(1962—),男,教授、博士,主研方向为数据挖掘、图像处理;李聪会,硕士研究生。

收稿日期:2020-01-03 修回日期:2020-02-06 E-mail: 18829028767@163.com

随着计算机性能的提升和数据资源的丰富,深度学习在计算机视觉等领域快速发展。近年来,研究人员尝试使用卷积神经网络(Convolutional Neural Networks, CNN)作为表征行人属性特征的工具,并取得了突破性的进展。文献[1]将行人图像作为识别模型的输入,无需考虑人体姿态^[2]、图像块^[3-4]及上下文信息^[5]而对所有属性进行预测,该方法虽然简单直观,但缺乏对细粒度属性特征的考量。文献[6]将人体划分为15个可重叠的部分并为每个部分训练CNN模型,根据空间约束关系选择对应的模型进行属性预测。虽然利用部件信息^[7-8]可以提高模型的整体性能,但作为一种中间阶段的操作,部件定位需要更多的训练和推理时间,甚至需要人工标注部件位置。文献[9]提出的基于注意力机制的深度网络(HydraPlus-Net, HP-Net),通过整合从局部到整体等多级注意力特征来提升模型对细粒度特征的表达能。文献[10]从网络的多个阶段提取以弱监督方式学习到的视觉注意力掩码,并引入注意力损失函数以防训练失衡。文献[11]将视图预测作为行人属性推断的关键线索,并从网络的后续层中分离出特定视图的属性预测单元。

综上,结合注意力机制的神经网络属性分类方法能使网络在较小能耗下提取属性相关位置的关键信息,具有较好的映射表达能力,可突破传统方法的局限,但若将其应用于实际场景中仍存在问题,一是对于正样本尺寸较小的细粒度属性(如眼镜、围巾等)识别准确率较低,二是在实际场景中各属性类别的类内差异较大(如外观尺寸多样性和外观模糊性),三是不同属性的收敛速度不同,从而导致各行人属性在网络训练过程中相互影响。

针对上述问题,本文基于残差网络(Residual Network, ResNet)^[12],提出一种多级注意力跳跃连接(Multi-Level Attention Skip Connection, MLASC-Net)网络,分别在PETA^[13]和RAP^[14]经典数据集上联合训练所有属性,以有效传递和重用属性间的信息。在设计MLASC模块中,注意力模块主要用于增强网络对属性信息的敏感性,使得特征表示更加紧凑,提高网络对关键特征的提取筛选能力,而多级跳跃连接结构通过嵌套的跨层级连接,既保留了浅层的细粒度信息,又有利于梯度的反向传播。此外,采用改进的空间金字塔池化方法^[15]来替代网络顶层的池化层,从而更全面地集成不同尺度和抽象层次的属性信息。在网络输出层,利用自适应加权损失层加快模型的收敛速度,以进一步提升行人属性识别的准确率和效率。

1 改进的行人属性识别方法

1.1 网络模型

本文所采用的神经网络主体结构如图1所示,主要由ResNet50残差网络区块、MLASC模块、多尺度金字塔及自适应加权损失层四部分组成。首先,借助MLASC模块对输入的行人图像进行浅层及深层特征提取,通过敏感注意力模块将局部特征与其全局依赖关系相结合,突出目标属性对应的重点区域及关键特征,并将缩放的注意力加权特征进行多级跨越连接,从而将特征传递给后续网络层。在网络顶层利用改进的金字塔实现多尺度特征表达,同时利用自适应加权损失层来衡量和学习不同属性任务间的关系,动态调整各属性任务的权重,最终实现对行人属性的预测。

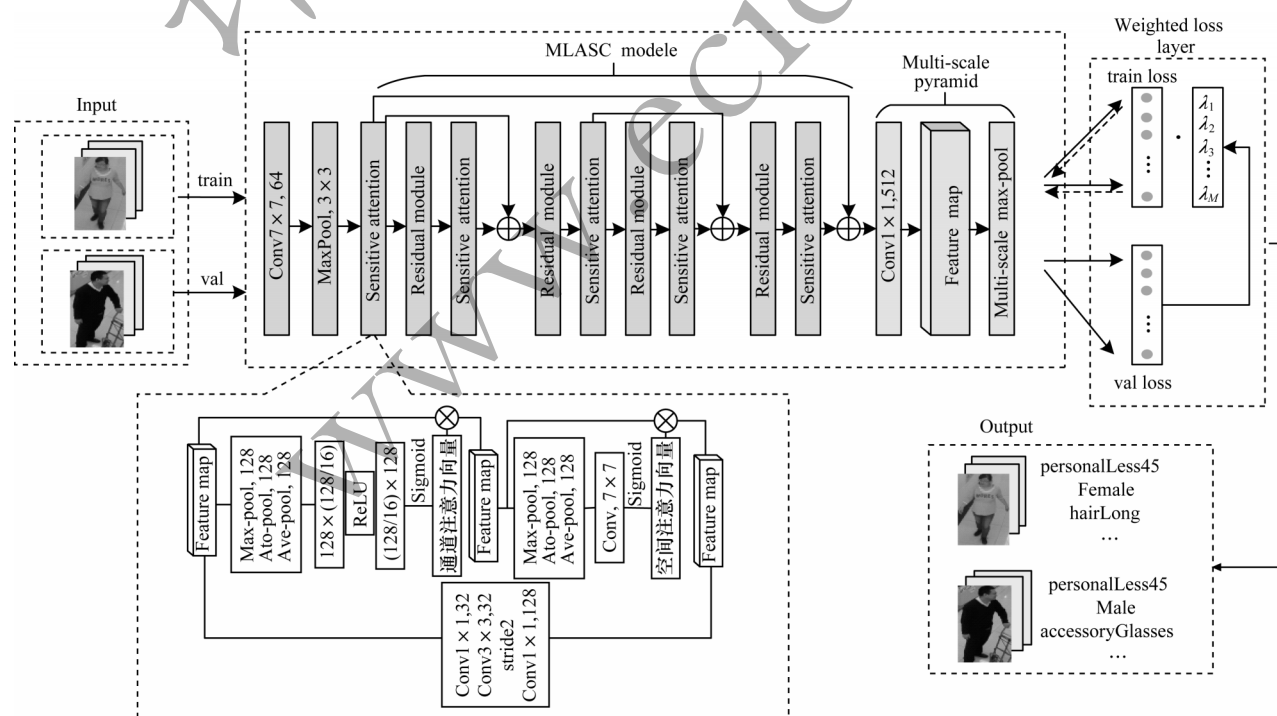


图1 本文行人属性识别网络结构

Fig.1 Structure of pedestrian attribute recognition network in this paper

1.2 多级注意力跳跃连接 MLASC 模块

1.2.1 敏感注意力模块

为了增强网络对不同属性信息的敏感性,有效促进网络的信息流动,本文借鉴并改进文献[16-17]中的注意力机制,先后对通道及空间的特征信息进行筛选,改进的敏感注意力模块网络结构如图2所示。

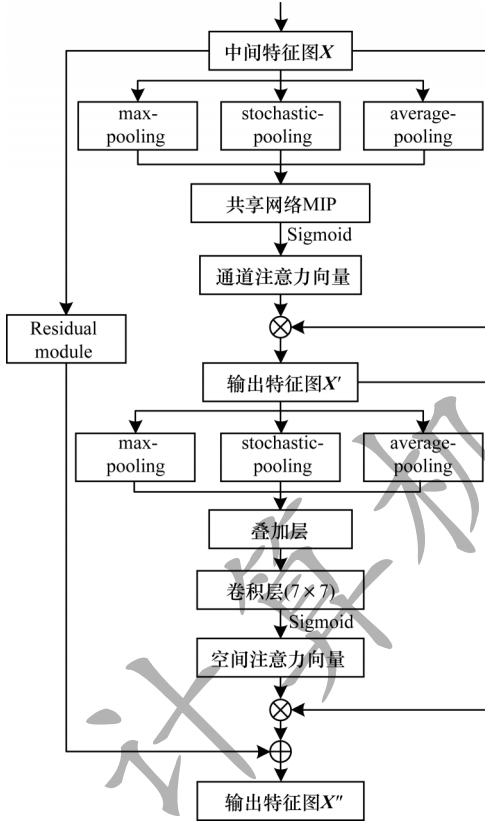


图2 敏感注意力模块

Fig.2 Sensitive attention module

改进的敏感注意力模块主要分为核心分支和辅助分支2个部分,核心分支由通道和空间注意力单元组成,在 average-pooling、max-pooling 的基础上设计添加 stochastic-pooling 操作以完善所提取的特征并减少关键信息的损失。此外,为了使网络达到更好的适应性,本文设计一组残差块作为辅助分支,降低2个分支在同一网络层聚合后的特征差异,从而更易与残差模块堆叠形成深度卷积网络。核心分支对于特征图X的处理过程可表示为:

$$X' = Y_c(X) \otimes X \quad (1)$$

$$X'' = Y_s(X') \otimes X' \quad (2)$$

其中, \otimes 表示对应矩阵相乘, X' 、 X'' 分别表示经过通道和空间注意力筛选加权处理后的特征图。中间特征图经过核心分支精修后,依次生成一维通道注意力图 $Y_c \in \mathbb{R}^{C \times 1 \times 1}$ 和二维空间注意力图 $Y_s \in \mathbb{R}^{1 \times H \times W}$, 其计算过程分别如式(3)、式(4)所示:

$$\begin{aligned} Y_c(X) &= \delta \left(\text{MLP}(\text{AvgPool}(X)) + \right. \\ &\quad \left. \text{MLP}(\text{MaxPool}(X)) + \right. \\ &\quad \left. \text{MLP}(\text{StoPool}(X)) \right) = \\ &\quad \delta \left(W_1 \left(W_0(X_{\text{avg}}^c) \right) + W_1 \left(W_0(X_{\text{max}}^c) \right) + \right. \\ &\quad \left. W_1 \left(W_0(X_{\text{sto}}^c) \right) \right) \\ Y_s(X) &= \delta \left(f^{7 \times 7} [\text{AvgPool}(X); \text{MaxPool}(X); \right. \\ &\quad \left. \text{StoPool}(X)] \right) = \\ &\quad \delta \left(f^{7 \times 7} \left([X_{\text{avg}}^s; X_{\text{max}}^s; X_{\text{sto}}^s] \right) \right) \end{aligned} \quad (3)$$

其中, δ 表示 Sigmoid 函数, W_0 和 W_1 分别为共享网络 MLP 的权重, $f^{7 \times 7}$ 表示卷积核为 7×7 的卷积操作, X_{avg}^c 、 X_{max}^c 和 X_{sto}^c 分别表示经过3种不同池化后所得到的通道注意力特征图,同理, X_{avg}^s 、 X_{max}^s 和 X_{sto}^s 则表示经过3种池化操作后所得到的空间注意力特征图。

1.2.2 多级跳跃连接

残差网络的跳跃连接结构已被证明在解决神经网络的梯度消失及梯度爆炸问题中具有有效性,不同于此类原始的跳跃连接结构,本文所讨论的多级跳跃连接主要针对所采用的敏感注意力模块,通过在每2个注意力模块之间增加一个跳跃连接,保留并重用来自浅层的细粒度信息和局部信息,并与下级注意力特征进行逐元素加和,以确保局部特征不会因为特征通道的变化而导致较多的信息损失。此外,对于首尾2个注意力模块再叠加一组跨度更广的跳跃连接,一方面有利于维持网络在反向传播过程中的梯度流,另一方面将注意力模块的显著性特征的补偿跨度进一步拉大,实现属性信息从低级到高级的层级提取。

1.3 多尺度金字塔池化

由于卷积残差网络在特征提取的过程中侧重于融合不同卷积层的全局特征,忽略了同一卷积层上多尺度局部区域特征的充分融合,缺乏对场景中小目标的预测能力,因此本文对空间金字塔池化方法进行改进:首先采用 1×1 的卷积将通道数减少到512,然后按照4种不同的尺度进行最大池化,以更好地保留注意力模块约束或增强后的特征表达。改进后的多尺度金字塔中的滑动窗口和特征图大小分别为 $S_{\text{pool}} \times S_{\text{pool}}$ 、 $S_{\text{fmap}} \times S_{\text{fmap}}$, 两者的关系可表示为:

$$S_{\text{pool}} = \left\lceil \frac{S_{\text{fmap}}}{n_i} \right\rceil \quad (5)$$

令 $n_i = 1, 2, 3, 4$, 将会得到4种不同尺度的滑动窗口,根据不同的滑动窗口来集成不同尺度的特征信息。其中,池化的步长为1,利用填充操作来保证输出特征图大小恒定,然后将得到的不同尺度的特征图进行叠加,最终输出2048维的特征向量。

1.4 自适应加权策略

由于多任务学习中的共享策略可能会带来负迁移问题,即当学习任务不同时,不充分的强制迁移学习可能会影响网络的学习效率。同时,网络在针对不同的学习任务时具有不同的学习难度和收敛速率,如行人衣服颜色比性别属性容易识别。因此,本文去掉原有的损失层,通过计算不同属性任务在验证集上的损失变化趋势及占比来更新下一轮训练中各属性任务的权重,并反向传播至前层网络,改进后的网络结构如图1所示。通过一种能够自适应学习不同任务间关系的内部共享机制,并结合验证损失算法^[18]实现各属性任务权重的动态调整,以兼顾任务间的相关性和模型的复杂性。

为了避免以相同的权重强制网络学习不同的属性任务,本文采取自适应加权策略来建模训练样本的预期损失,建模过程如式(6)所示:

$$P = \operatorname{argmin} \sum_{j=1}^M \sum_{i=1}^N \langle \lambda_j, L(\psi_j(x_i; P) - y_{ij}) \rangle \quad (6)$$

其中, P 为网络的优化参数集, M 、 N 分别表示属性数及样本数, $\langle \cdot \rangle$ 表示内积操作, λ_j 为属性 j 的权重, $L(\cdot)$ 为损失函数, x_i 为输入样本 i , y_{ij} 为样本 i 中属性 j 的实际值, ψ_j 为属性 j 的预测函数。

1.5 多尺度样本训练

在实际场景中,远处和近处的行人目标尺寸通常存在较大差异,结合相机的成像原理,同一目标处在与相机不同距离的区域时,成像效果也会有所区别。考虑到若采用单一尺度的行人样本进行训练,会出现行人属性的漏检或错检现象,因此,本文在网络训练时采用多尺度训练,利用双线性插值算法为每个样本设置3种尺寸,按比例将原图像缩放0.5倍、1.0倍和1.5倍。对于每个样本,随机选择其中一种尺寸输入至模型中进行训练,从而均衡训练目标的多样性分布,提高网络模型对于不同尺寸目标的鲁棒性。

2 实验结果与分析

2.1 实验数据集

为了验证本文所提 MLASC-Net 网络在行人属性识别任务中的有效性,采用 PETA 和 RAP 数据集进行实验,其中, PETA 数据集包含不同场景和不同条件下的 19 000 张行人图像,训练集包含 9 500 张行人图像,验证集为 1 900 张,其余 7 600 张作为测试集。RAP 数据集共有 41 585 个行人样本,收集于多摄像头所拍摄的室内场景。为了保证实验更具可靠性和说服力,本文分别选取 PETA 和 RAP 数据集中的通用属性进行评估,数据集中的部分行人样本如图3所示。



图3 数据集部分样本

Fig.3 Some samples of datasets

2.2 结果分析

为了评估 MLASC-Net 在解决行人属性分类问题时的有效性,采用以下2种评价指标来衡量模型性能:

1) 基于标签的评价指标——平均准确率(mA), 计算公式如下:

$$mA = \frac{1}{2N} \sum_{i=1}^L \left(\frac{TP_i}{P_i} + \frac{TN_i}{N_i} \right) \quad (7)$$

其中, N 为样本数, L 为属性个数, P_i 、 TP_i 分别表示属性 i 中的正样本数和被正确预测的正样本数, N_i 、 TN_i 分别表示属性 i 中的负样本数和被正确预测的负样本数。

2) 基于样本的评价指标——准确率(accuracy)、精确率(precision)、召回率(recall)、F1值,这种评估方式可以使用多属性联合学习模型。4个指标计算公式分别如下:

$$\text{accuracy} = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i \cap f(x_i)|}{|Y_i \cup f(x_i)|} \quad (8)$$

$$\text{precision} = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i \cap f(x_i)|}{|f(x_i)|} \quad (9)$$

$$\text{recall} = \frac{1}{N} \sum_{i=1}^N \frac{|Y_i \cap f(x_i)|}{|Y_i|} \quad (10)$$

$$F1 = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (11)$$

其中, Y_i 、 $f(x_i)$ 分别表示第 i 个样本真值中的正标签、第 i 个样本预测标签中的正标签, $|\cdot|$ 为集合基数。

本文算法采用 Pytorch 框架,在 GTX 1080Ti GPU (11 GB 显存)、64 位 Centos7 计算机系统下进行相关实验,并分析属性分类结果,实验主要分为以下4个部分:

实验1 不同属性识别方法性能对比

将本文 MLASC-Net 与深度单属性识别(DeepSAR)、多属性识别(DeepMAR)^[19]、弱监督行人属性定位网络(WPAL-network)^[20]、HP-net^[9]、联合循环学习模型(JRL)^[21]和分组循环学习模型(GRL)^[22]等方法进行对比,各方法在 PETA 和 RAP 数据集上的实验结果分别如表1、表2所示。

表1 PETA数据集上不同方法的性能对比结果

Table 1 Performance comparison results of different methods on PETA dataset %

方法	mA	准确率	精确率	召回率	F1
ACN ^[1]	81.15	73.66	84.06	81.26	82.64
DeepSAR	81.30	—	—	—	—
DeepMAR	82.89	75.07	83.68	83.14	83.41
WPAL	85.50	76.98	84.07	85.78	84.90
HP-net	81.77	76.13	84.92	83.24	84.07
JRL	85.67	—	86.03	85.34	85.42
Imbalanced-Learning* ^[10]	84.59	78.56	86.79	86.12	86.46
VeSPA ^[11]	83.45	77.73	86.18	84.81	85.49
GRL	86.70	—	84.34	88.82	86.51
MLASC-Net	85.28	79.10	87.90	85.39	86.63

表2 RAP数据集上不同方法的性能对比结果

Table 2 Performance comparison results of different methods on RAP dataset %

方法	mA	准确率	精确率	召回率	F1
ACN	69.66	62.61	80.12	72.26	75.98
DeepSAR	—	—	—	—	—
DeepMAR	73.79	62.02	74.94	76.21	75.56
WPAL	79.48	53.30	60.82	78.80	68.65
HP-net	76.12	65.39	77.33	78.79	78.05
JRL	77.81	—	78.11	78.98	78.58
Imbalanced-Learning*	77.39	68.16	81.57	78.59	80.05
VeSPA	77.70	67.35	79.51	79.67	79.59
GRL	81.20	—	77.70	80.90	79.29
MLASC-Net	80.34	68.71	79.63	81.25	80.43

从表1、表2可以看出,由于DeepSAR对每个属性优化单独的二值分类模型,未考虑各属性间的关系,而其他方法均联合训练所有属性,模型预测平均准确率均优于DeepSAR。对比其他方法,本文在残差网络的基础上设计MLASC模块和多尺度金字塔单元,同时结合加权损失层提高模型的整体性能。在2个数据集上,MLASC-Net均有3个基于样本的评价指标达到了最高值,而对于基于标签的度量指标mA,MLASC-Net略低于GRL方法,但相对其他多数方法仍有明显优势,该结果验证了MLASC-Net在行人属性识别任务中的有效性。

实验2 采用不同策略训练网络的识别效果

本文所采用的行人属性识别方法中包含3个新的模块:1) MLASC模块,可以提升网络对行人不同目标属性的针对性;2)改进的多尺度金字塔单元,用以提取行人多尺度局部区域特征;3)在损失层引入了自适应加权策略。为了进一步评估各模块对网络性能的影响,本次实验以未加入MLASC模块和多尺

度金字塔单元以及自适应加权策略的原网络(baseline网络)为基准,在此基础上,将注意力(即MLASC模块)、金字塔和加权损失视为3个实验变量进行一系列消融实验,在PETA与RAP数据集上的实验结果分别如表3、表4所示。其中,“√”表示加入该模块,“×”表示未加入该模块。

表3 PETA数据集上采用不同策略训练网络时的识别效果对比

Table 3 Comparison of recognition effect when training network with different strategies on PETA dataset %

策略	注意力	金字塔	加权损失	mA	准确率	精确率	召回率	F1
1	×	×	×	81.70	74.48	82.33	82.27	82.30
2	√	×	×	84.63	77.09	84.61	85.12	84.86
3	×	√	×	83.04	75.94	82.79	83.65	83.22
4	×	×	√	83.56	76.03	83.12	83.94	83.53
5	√	√	√	85.28	79.10	87.90	85.39	86.63

表4 RAP数据集上采用不同策略训练网络时的识别效果对比

Table 4 Comparison of recognition effect when training network with different strategies on RAP dataset %

策略	注意力	金字塔	加权损失	mA	准确率	精确率	召回率	F1
1	×	×	×	71.32	62.17	72.58	75.49	74.01
2	√	×	×	78.67	67.52	77.94	78.46	78.20
3	×	√	×	74.64	63.89	73.29	76.52	74.87
4	×	×	√	74.81	64.65	74.37	77.23	75.77
5	√	√	√	80.34	68.71	79.63	81.25	80.43

对比表3、表4中策略1、策略5的实验结果可以看出,相比于baseline网络,MLASC模块在mA、准确率、精确率、召回率和F1这5项指标上的分类性能均有所提高,其中,在PETA和RAP数据集上的识别准确率分别提高约4.62和6.54个百分点,这说明网络在抽取不同属性的相关特征时更具偏向性,模型的判别能力增强。对比策略1和策略3发现,采用改进的多尺度金字塔能够使得模型对不同尺度的行人属性更具鲁棒性,增强模型的泛化能力。对比策略1和策略4发现,由于网络通过衡量损失变化趋势为各属性赋予动态权重参数,从而调度不同的属性任务,使得模型的各项性能指标均有显著提升。通过策略1~策略4的对比可以看出,MLASC模块对于提升模型性能的贡献最大,损失加权策略的贡献次之。实验结果表明,本文所采用的MLASC模块、多尺度金字塔以及损失加权策略,均能有效提升属性识别模型的判别能力。

实验3 模型收敛速度对比

多信息模型自适应加权的思想就是在网络训练的过程中动态调度不同的目标任务,以得到优化的参

数据集合并加速模型的收敛。为了衡量自适应加权策略的有效性,本次实验将对未加权网络及动态加权网络在训练过程中loss值的变化趋势进行对比分析,在PETA与RAP 2个数据集上的实验结果分别如图4、图5所示。

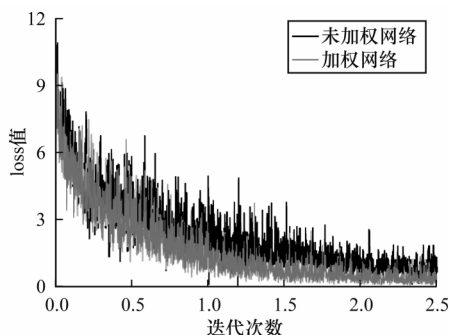


图4 PETA数据集上模型收敛速度对比

Fig.4 Comparison of model convergence rates on PETA dataset

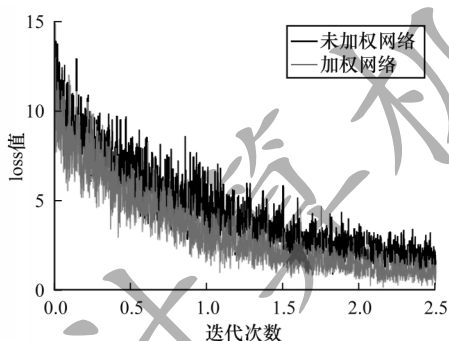


图5 RAP数据集上模型收敛速度对比

Fig.5 Comparison of model convergence rates on RAP dataset

从图4、图5可以看出,未加权网络为各属性任务赋予相同的权重参数,而加权网络在训练过程中可以阶段性地衡量不同任务的重要程度,从而动态地指导任务调度,因此,加权网络在PETA数据集上迭代至16 000次左右开始收敛,而未加权网络在迭代次数达到约20 000次时趋近收敛。同理,在RAP数据集上加权网络的收敛速度也快于未加权网络。此外,由于网络对具有不同学习难度及收敛速度的属性赋予不同的权重,缓解了多属性任务间的负迁移问题,因此从整体上来看,在相同迭代次数时加权网络的loss值相对较低。在2个数据集上的实验结果均表明,本文所采用的自适应加权策略可以有效加速模型的收敛并提高精度。

实验4 MLASC-Net的属性识别结果

为了进一步验证本文MLASC-Net的有效性,本次实验随机选取了测试数据集中不同视角、目标较模糊等不同场景下的单个行人样本进行属性识别,实验结果如图6所示,每个行人图像下方是其对应的属性预测结果,其中,浅色字体为细粒度属性,如帽子、头发和眼镜。从图6可以看出,第3个和第6个为同一行人在不同视角下的图像,性别、背包、帽子和服装属性的检测结果相同,但由于行人正面比背面的属性信息更为丰富全面,因此,第3个样本中的年龄范围、围巾、眼镜属性均被检测出。综合其他行人样本的属性识别结果得出,MLASC-Net可有效提高网络对细粒度属性的鲁棒性。

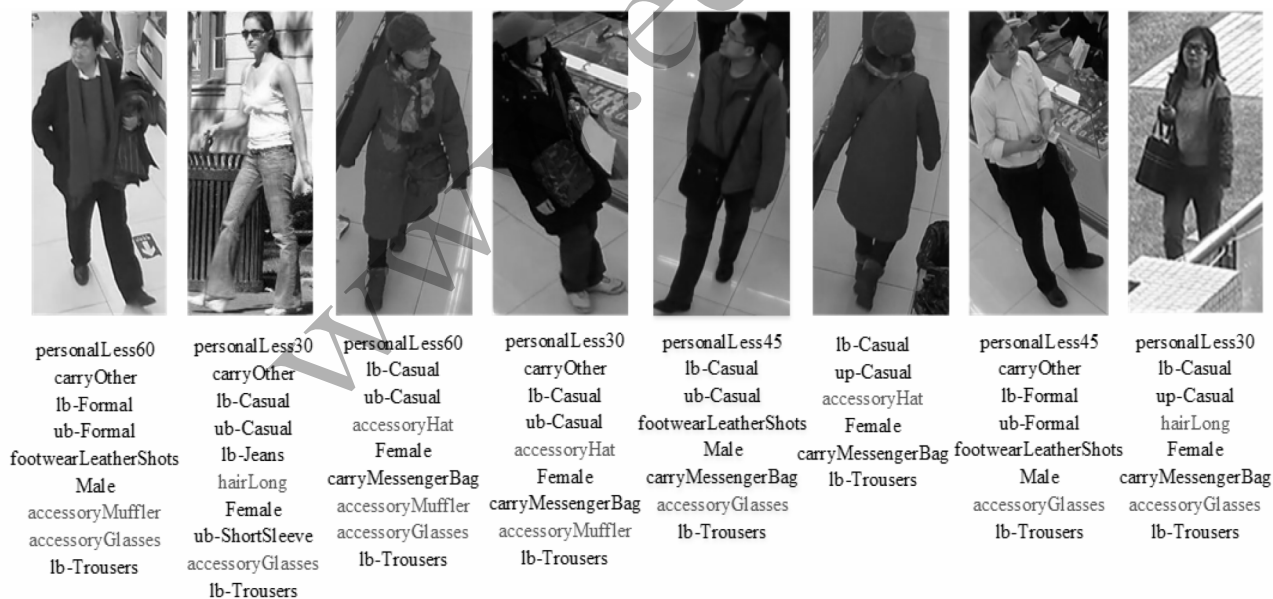


图6 MLASC-Net的属性识别结果

Fig.6 Attribute recognition results of MLASC-Net

3 结束语

本文提出多级注意力跳跃连接网络 MLASC-Net, 以进行行人属性识别。通过 MLASC 模块和多尺度金字塔克服图像编码表征存在的局限性, 增强网络对属性信息的敏感性。此外, 自适应加权损失层不仅有利于模型训练, 同时尽可能地规避了不同属性任务间的冲突。实验结果表明, MLASC-Net 能够自动提取行人属性的关键特征, 训练出的模型对于多尺度属性具有良好的鲁棒性和泛化能力。但是, 本文实验中还存在属性样本比例不均衡的问题, 下一步将采集复杂环境下的行人图像或视频来扩充数据集, 从而平衡属性样本并进一步提升识别效果。

参考文献

- [1] SUDOWE P, SPITZER H, LEIBE B. Person attribute recognition with a jointly-trained holistic CNN model[C]//Proceedings of 2015 IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2015: 87-95.
- [2] ZHANG N, PALURI M, RANZATO M, et al. PANDA: pose aligned networks for deep attribute modeling[C]//Proceedings of 2014 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2014: 1637-1644.
- [3] FABBRI M, CALDERARA S, CUCCHIARA R. Generative adversarial models for people attribute recognition in surveillance[C]//Proceedings of 2017 IEEE International Conference on Advanced Video and Signal Based Surveillance. Washington D. C., USA: IEEE Press, 2017: 1-6.
- [4] SUDOWE P, LEIBE B. PatchIt: self-supervised network weight initialization for fine-grained recognition[C]//Proceedings of British Machine Vision Conference. New York, USA: ACM Press, 2016: 24-30.
- [5] LI Y N, HUANG C, LOY C C, et al. Human attribute recognition by deep hierarchical contexts[M]. Berlin, Germany: Springer, 2016.
- [6] ZHU Jianqing, LIAO Shengcai, YI Dong, et al. Multi-label CNN based pedestrian attribute learning for soft biometrics [C]//Proceedings of 2015 International Conference on Biometrics. Washington D. C., USA: IEEE Press, 2015: 535-540.
- [7] LI Dangwei, ZHANG Zhang, CHEN Xiaotang, et al. A richly annotated dataset for pedestrian attribute recognition [EB/OL]. [2019-12-10]. <https://arxiv.org/abs/1603.07054>.
- [8] WANG Feng, LIU Qingshan. Fine pedestrians recognition based on double convolutional neural network[J]. China Sciencepaper, 2017(14): 17-21. (in Chinese)
王枫, 刘青山. 基于双卷积神经网络的行人精细化识别[J]. 中国科技论文, 2017(14): 17-21.
- [9] LIU Xihui, ZHAO Haiyu, TIAN Maoqing, et al. HydraPlus-Net: attentive deep features for pedestrian analysis [C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017: 350-359.
- [10] SARAFIANOS N, XU X, KAKADIARIS I A. Deep imbalanced attribute classification using visual attention aggregation[M]. Berlin, Germany: Springer, 2018.
- [11] SAQUIB M S, SCHUMANN A, WANG Y, et al. Deep view-sensitive pedestrian attribute inference in an end-to-end model[C]//Proceedings of British Machine Vision Conference. Washington D. C., USA: IEEE Press, 2017: 159-168.
- [12] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2016: 770-778.
- [13] DENG Y B, LUO P, LOY C C, et al. Pedestrian attribute recognition at far distance [C]//Proceedings of the 22nd ACM International Conference on Multimedia. New York, USA: ACM Press, 2014: 789-792.
- [14] LI Dangwei, ZHANG Zhang, CHEN Xiaotang, et al. A richly annotated pedestrian dataset for person retrieval in real surveillance scenarios [J]. IEEE Transactions on Image Processing, 2019, 28(4): 1575-1590.
- [15] HUANG Zhanchao, WANG Jianlin. DC-SPP-YOLO: dense connection and spatial pyramid pooling based YOLO for object detection [J]. Information Sciences, 2020, 522: 241-258.
- [16] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Washington D. C., USA: IEEE Press, 2018: 7132-7141.
- [17] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[M]. Berlin, Germany: Springer, 2018.
- [18] HE Keke, WANG Zhanxiong, FU Yanwei, et al. Adaptively weighted multi-task deep network for person attribute classification [C]//Proceedings of the 25th ACM International Conference on Multimedia. New York, USA: ACM Press, 2017: 1636-1644.
- [19] LI Dangwei, CHEN Xiaotang, HUANG Kaiqi. Multi-attribute learning for pedestrian attribute recognition in surveillance scenarios[C]//Proceedings of 2015 IAPR Asian Conference on Pattern Recognition. Washington D. C., USA: IEEE Press, 2015: 111-115.
- [20] LI Dangwei, YU Kai, LENG Biao, et al. Weakly-supervised learning of mid-level features for pedestrian attribute recognition and localization [C]//Proceedings of British Machine Vision Conference. Washington D. C., USA: IEEE Press, 2017: 15-26.
- [21] WANG Jingya, ZHU Xiatian, GONG Shaogang, et al. Attribute recognition by joint recurrent learning of context and correlation [C]//Proceedings of 2017 IEEE International Conference on Computer Vision. Washington D. C., USA: IEEE Press, 2017: 531-540.
- [22] ZHAO Xin, SANG Liufang, DING Guiguang, et al. Grouping attribute recognition for pedestrian with joint recurrent learning [EB/OL]. [2019-12-10]. <https://www.ijcai.org/Proceedings/2018/0441.pdf>.