

联合全局与局部外观特征的无人机行人属性识别

陈楠¹ 杨玻¹ 刘书羽¹ 尉嘉维¹
CHEN Nan YANG Bo LIU Shuyu YU Jiawei

摘要

针对无人机航拍图像下行人属性识别任务特征提取效率低问题,提出一种联合全局与局部外观特征学习的行人属性识别算法。首先为了提升算法的局部区域定位能力,提出一种新颖的基于注意力的属性定位模块,模块通过矩阵乘法来建模不同通道间的空间依赖关系,并利用全局平均池化产生的特征进一步增强提取到的局部区域信息,实现了更好的局部特征表达能力;其次,为了防止生成的局部区域信息出现冗余,设计一种注意力多样性损失,通过最小化特征通道的相似性来约束各个局部区域互不重叠;最后,在两个公开的行人属性识别数据集上的实验结果表明,上述设计可以有效提高属性识别准确率并且参数量仅轻微增加。

关键词

行人属性识别;属性分组与定位;注意力多样性;无人机

doi: 10.3969/j.issn.1672-9528.2024.02.025

0 引言

随着航空遥感技术的发展,无人机被广泛应用于军事侦察、环境监测及交通规划等领域。人作为图像的主体,对无人机航拍图像进行行人属性识别能够拓宽无人机的场景理解能力,在军事和民用领域具有重要的应用价值。与常规行人属性识别问题相比,无人机航拍图像具有图像视野广、小目标多等挑战,因此如何能从航拍图像小目标的有限特征中提高特征提取效率是解决问题的关键。

按照特征提取的方式,基于深度学习的行人属性识别的主流方法可分为基于全局图像^[1-3]、基于图区块^[4-6]、基于注意力^[7-9]和基于序列^[10-12]的方法。基于全局图像的行人属性识别方法使用整幅图像作为网络的输入,通过特征共享、多任务学习、端到端训练等方法提取特征,这类方法虽然简单高效,但未考虑到某些局部属性所需要的细粒度特征,识别精度较低。

为了综合细粒度特征和全局特征,已有工作通常采用额外的部件定位模块,如目标检测、姿态估计、人体分割以及注意力模块提取局部细粒度特征。前者需要额外的计算资源,并且如果直接使用其他数据集训练得到的检测模型,实际应用时可能会出现误差,此类方法也无法获取携带物的局部区域。后者仅使用全局身份标签作为监督,获得的不同属性对应的局部区域可能会出现冗余重叠,难以准确地定位多个不

同的局部人体区域。

基于序列的方法将不同的神经网络、属性组、多任务学习等模块集成到循环神经网络(recurrent neural network, RNN)框架中进行序列化的属性预测任务,以此来建模属性间的相关性。此类方法虽然取得较高精度,但由于计算量较大,使得实际工程应用时执行效率较低。针对上述不足,本文提出了一种联合全局与局部外观特征学习的人员属性识别算法,通过设计一种新颖的基于注意力的属性定位模块来获取更好的局部特征表达能力,并且引入一种注意力多样性损失约束生成的局部区域信息互不重叠。在多个公开的行人属性识别数据集上的实验结果表明,上述设计可以大幅度提高图像的特征提取效率,进而提高属性的识别准确率。

1 方法介绍

1.1 算法整体流程

如图1所示,本章所提出的联合全局与局部外观特征学习的人员属性识别算法共分为3个步骤,首先通过本章改进后的深度卷积骨干网络提取深度特征,然后经过全局外观特征学习分支和局部外观特征学习分支各自得到一个预测结果,最后将两个分支的估计以最大投票的方式进行融合,以改善最终的决策。具体地,全局外观特征学习分支和局部外观特征学习分支共享一个特征提取网络,为了增强骨干网络所提取的语义特征,使其更适合人员属性识别任务,对骨干网络进行以下改进:将第5个卷积阶段用于空间下采样的卷积层的滑动步长设置为1,以保留某些局部属性所需要的细节信息;第5个卷积阶段后引入高斯通道注意力模块来重新校准不同通道的能力,同时增加一个恒等映射使语义增强后

1. 中国航空工业集团公司西安航空计算技术研究所
陕西西安 710068

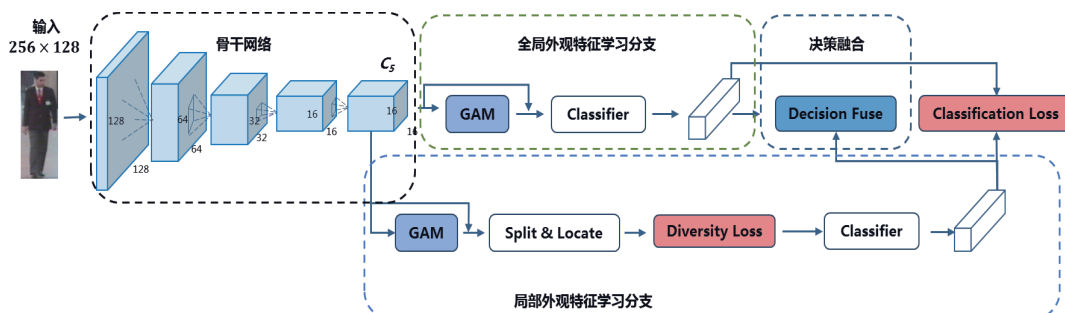


图1 算法整体流程图

的特征更具有适应性，让网络更关注于具有判别力的特征。

对于全局外观特征学习分支，本文使用通用的属性学习框架，即对于增强后的深度卷积网络提取到的高级语义特征，全局外观特征学习分支利用一个由全局平均池化层和全连接层组成的分类器来输出各个属性存在的概率，其中全局平均池化层用来提取人员的全局特征。

如果只使用全局特征，模型将无法关注某些属性所需的细粒度特征，因此设计一个局部外观特征学习分支来利用属性所处空间区域的规律性，以及属性之间的纠缠性提升性能，弥补只使用全局特征的局限性。整个局部外观特征学习分支包含两部分的内容：属性分组与定位以及注意力多样性损失。

通过全局外观特征学习分支和局部外观特征学习分支，各自获得一个预测向量。在推理过程中，与常用方法类似，对这两个分支的预测结果进行决策融合（逐元素求最大值），以获得两个分支的最佳预测结果。与推理过程不同，在训练过程中这两个分支产生的预测向量直接由真实标签监督，而不是对融合后的预测结果进行监督，此操作的目的是保证每一个分支都可以拥有足够的梯度，以免出现某一分支训练不足的情况。

1.2 局部外观特征学习分支

全局外观特征学习分支共用一个全局平均池化特征来分类多个属性，忽略了不同属性之间的空间差异和语义差异，本节针对该问题提出局部外观特征学习分支。具体来讲，首先根据空间分布将属性分成4组，然后为每一个组的属性设计一个属性定位模块来定位与属性相关的局部区域，最后同时利用多样性损失函数来迭代改善属性相关区域的定位。下面将详细介绍局部外观特征学习分支的算法细节。

1.2.1 属性分组 (split) 与定位 (locate)

由于行人属性识别任务需要一次性识别多个属性，本文根据各属性所处的空间位置对属性进行分组，具体来讲，采取多分支结构来关注不同区域的属性。多分支结构共享第5个卷积阶段的特征，并且第5个卷积阶段的特征将经过高斯通道注意力增强后分解成树状结构，从而产生各局部区域的子任务网络。如图2所示，参考文献[13]，将尺寸为 16×8

的人员特征图划分为互相重叠的三个区域：头肩、上半身以及下半身区域，三个区域尺寸依次为 4×8 、 7×8 和 7×8 。将上述三个局部特征图和全身特征图的尺寸统一调整到 8×8 后，输入

到本节设计的组属性定位模块 GAL (group attribute location module)，前三个分支依次输入头肩、上半身、下半身特征，第四个分支输入图像的整体特征。

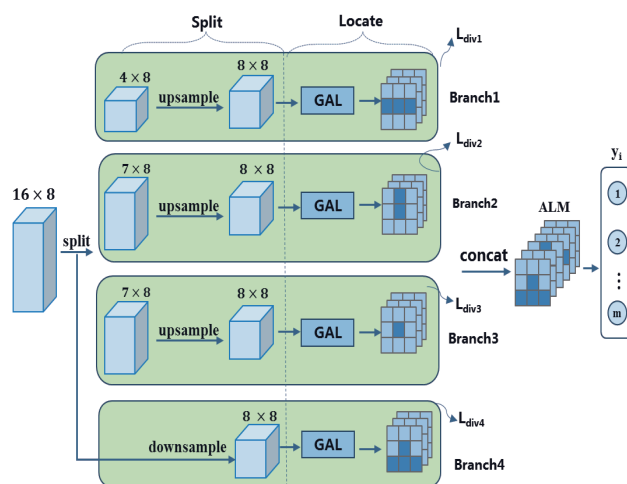


图2 局部外观特征学习分支流程示意图

在行人属性识别任务中，某些细粒度属性通常对应着一个局部身体区域（例如头部区域）或显著目标区域（例如背包），为了增强各分支的局部区域定位能力，将属性按照所属空间区域分配到四组属性中后，本节设计了一个简单有效的属性定位模块。该模块由3层卷积层、1个平均池化层、1个BN层以及 Sigmoid 激活层组成，充分利用属性学习过程中本身具有的局部感知能力，自适应地学习特征的空间依赖关系。具体地，本节在每一个分支中都引入一个属性定位模块，为每一个属性组生成各自的注意力定位掩膜。通常注意力定位掩膜个数等于最终要识别的属性个数，本节考虑到不同属性可能对应相似或者相同的人员局部区域，例如上身衣着款式中的“条纹”和“格子”对应人员相同的局部区域。为了减少冗余定位消息，手动进行属性组合，使拥有相似局部区域的属性共享一个属性定位掩膜。具体地，为每个属性定位分支生成 K 个属性定位掩膜 ($K \leq C$, C 为每组属性类别个数)：

$$M_i = GAL(F) \quad i=1,2,\dots,K \quad (1)$$

式中： $GAL()$ 表示组属性定位模块， F 表示输入的属性特征图，

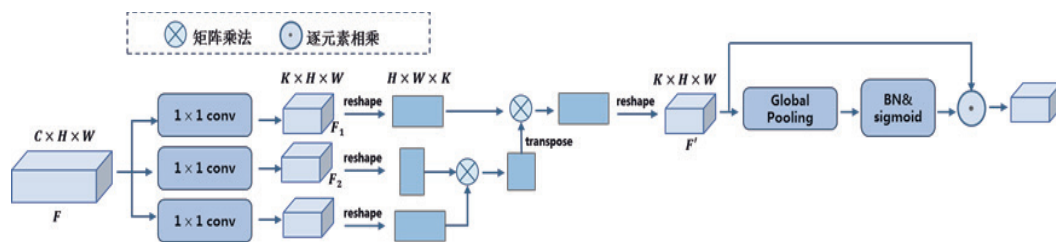


图3 GAL 流程图

具体步骤如图3所示，可分为以下几步。

(1) 首先对于特征图 $F_3 \in \mathbb{R}^{C \times H \times W}$ ，分别通过3个卷积层得到3个特征图 $F_1, F_2, F_3 \in \mathbb{R}^{K \times H \times W}$ ， K 为通道数，表示生成的属性定位图数目，通过 1×1 卷积生成属性定位掩膜，每个通道的特征代表对应属性的空间显著区域。

(2) 然后利用矩阵乘法建模 K 个属性定位掩膜之间的空间依赖关系，具体地，先将上述操作生成的 F_1, F_2, F_3 进行 reshape 操作使其维度均变为 $K \times N$ ，其中 $N=H \times W$ 。

再将经过 reshape 操作后的 F_2 特征的转置（维度为 $N \times K$ ）与 reshape 后的 F_3 特征（维度为 $K \times N$ ）进行矩阵乘得到关系注意力掩膜 $S (K \times K)$ ；将 S 转置后（ $K \times K$ ）与 reshape 后的 F_1 特征（ $K \times N$ ）进行矩阵乘法操作，并 reshape 为原来形状，得到特征图 F' ，此过程类似于自注意力机制，可以自主学习各个通道间的关系。

(3) 最后，对各个属性定位掩膜学习到的语义特征进行增强。对特征 F' 进行平均池化得到初始的注意力掩膜，然后利用 BN 层对该掩膜进行归一化处理，并同时学习两个缩放偏移参数使得 BN 归一化操作可被还原，经过 Sigmoid 激活函数得到最终的注意力权重，利用该权重对特征图 F' 增强，增强其定位能力。

得到每一个属性组的属性定位掩膜后，将4个属性组的属性定位掩膜拼接得到完整的属性定位掩膜。此步骤可以帮助模型利用属性之间的纠缠性，如果直接在每一个分支预测对应属性的类别，则无法考虑位于不同区域之间的属性纠缠性。例如，一个人上半身衣着休闲，穿着运动鞋，其下半身穿着也大概率是休闲风格，因此，当将所有属性组的定位掩膜拼接后，可以根据相关属性的定位掩膜来辅助判断该属性是否存在，从而提高属性的识别准确率。

1.2.2 多样性损失函数

上节所得到的各属性定位掩膜是基于属性的分类标签学到的，这些标签对局部区域的学习来说属于弱监督信息，因此在学习得到的多个注意力定位图中很可能出现相互重叠的区域，本节针对该问题设计一种用于定位多个不同局部区域的特征图多样性损失函数来指导属性定位掩膜的生成，期望每一个通道表示一个特定的人员局部区域。

多样性损失函数基于本节提出的多样性组件实现，其核

心思想是通过最小化特征通道的相似性来约束各个通道间的特征距离。图4是本节提出的多样性组件结构图，可以发现本节所提的多样性组件仅由 softmax 层、跨通

道最大池化（cross channel max pooling, CCMP）以及求和层（sum）组成，是一个无参数模块。

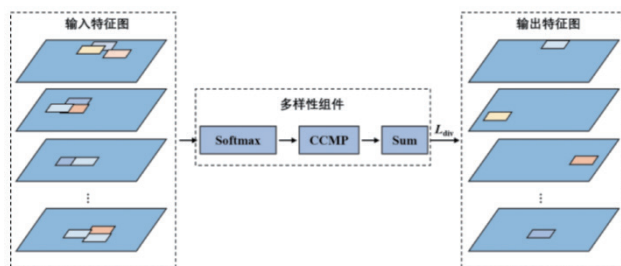


图4 多样性组件结构

具体步骤如下。首先在空间维度进行 Softmax 归一化操作，然后在上述操作后引入跨通道最大池化，通过在空间维度求和测量角度来对不同通道的特征实施监督，可表示为式

(2)。理论上，越具有判别力的区域，激活响应越大，因此将所有通道的最大响应值统一放于一张特征图上，如果各个通道特征图关注的区域不相同，则经过跨通道最大池化并进行求和操作后，得到的特征值 $h(F)$ 会比较大，反之则会比较小，具体如图5所示。由于使用随机梯度优化损失函数使其最小化，故将多样性损失函数表示为 $1 - \text{特征值 } h(F)$ ，具体如式(3)所示，则当特征值 $h(F)$ 最大，即各通道定位区域各不相同，损失函数最小。为每一个分支引入一个多样性损失函数，因此总的注意力多样性损失函数为4个分支之和，具体如式(4)所示。

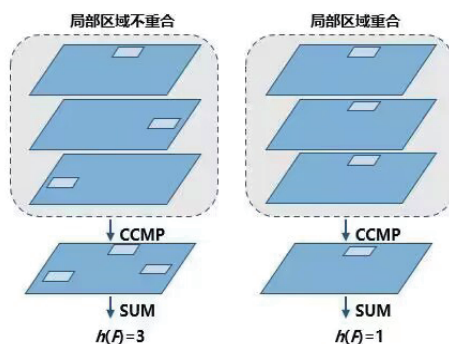


图5 跨通道最大池化（CCMP）示意图

$$h(F) = \sum_{i=1}^{WH} \max_{j=1,2,\dots,K} \left(\frac{e^{F_{i,j}}}{\sum_{i'=1}^{WH} e^{F_{i',j}}} \right) \quad (2)$$

$$L_{divi} = 1 - h(F_i) \quad (3)$$

$$L_{div} = L_{div1} + L_{div2} + L_{div3} + L_{div4} \quad (4)$$

2 实验结果与分析

ResNet50 在多项任务中均表现较优秀，是目前使用最广泛的深度卷积神经网络，因此以 ResNet50 作为骨干网络依次添加本文所提出的各个模块，实验结果如表 1 所示，可以发现所提出的全局分支、属性分组 & 定位模块以及注意力多样性损失仅轻微增加参数量，但在 PETA 和 PA-100K 数据集上均可以有效提升两种评价指标（基于标签和基于实例），且对基于标签的指标 mA 提升效果更明显。

表 1 模块消融实验

方法	PETA					PA-100K					Params/M
	mA	Acc	Prec	Rec	F_1	mA	Acc	Prec	Rec	F_1	
Resnet50	85.42	79.58	86.42	87.37	86.89	79.96	77.94	85.70	87.64	86.61	23.561
全局分支	85.62	79.81	86.24	88.15	87.18	80.43	78.52	86.04	88.17	87.09	23.561
+ 属性分组 & 定位	86.26	79.57	85.53	88.53	87.00	81.06	78.44	84.91	89.49	87.14	23.669
+ 多样性损失	86.61	79.47	85.08	89.05	87.02	81.27	78.54	85.16	89.24	87.15	23.669

3 结论

为了更好地处理无人机航拍图像下的行人属性识别任务，本文提出一种联合全局与局部外观特征学习的行人属性识别算法，该算法由全局外观特征学习分支和局部外观特征学习分支组成，通过融合两个分支的预测结果来使模型更具鲁棒性。为了更好地提取属性所需要的局部细粒度特征，本文在局部外观特征学习分支中按照属性所处空间区域将属性分为 4 组，并设计一个轻量的属性定位模块来辅助模型感知属性的局部区域，同时引入注意力多样性损失以约束各个局部区域不重叠。在 PETA 和 PA-100K 数据集上测试本文所提算法性能，实验结果表明所提算法可以有效提高特征提取效率且参数量没有明显变化，具有较强实用价值。

参考文献:

- [1] SUDOWE P, SPITZER H, LEIBE B, et al. Person attribute recognition with a jointly-trained holistic CNN model[C]// Proceedings of the 2015 IEEE International Conference on Computer Vision.Piscataway:IEEE, 2015: 329-337.
- [2] ABDULNABI A H, WANG G, LU J W, et al. Multi-task CNN model for attribute prediction[J]. IEEE transactions on multimedia, 2015, 17(11): 1949-1959.
- [3] LI D, CHEN X, HUANG K, et al. Multi-attribute learning for pedestrian attribute recognition in surveillance scenarios[C]// Proceedings of the 3rd IAPR Asian Conference on Pattern Recognition.Piscataway:IEEE,2015: 111-115.
- [4] ZHU J, LIAO S, LEI Z, et al. Multi-label convolutional neu-

ral network based pedestrian attribute classification[J]. Image and Vision Computing, 2017, 58: 224-229.

- [5] ZHANG N, PALURI M, RANZATO M A, et al. PANDA: pose aligned networks for deep attribute modeling[C]//Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition.Piscataway:IEEE,2014: 1637-1644.
- [6] GKIOXARI G, GIRSHICK R, MALIK J, et al. Actions and attributes from wholes and parts[C]//Proceedings of the 2015 IEEE International Conference on Computer Vision.Piscataway:IEEE,2015: 2470-2478.
- [7] LIU X, ZHAO H, TIAN M, et al. Hydraplus-net: attentive deep features for pedestrian analysis[C]//Proceedings of the 16th IEEE International Conference on Computer Vision.Piscataway:IEEE, 2017: 350-359.
- [8] SARAFIANOS N, XU X, KAKADIARIS I A. Deep imbalanced attribute classification using visual attention aggregation[C]//Proceedings of the 15th European Conference on Computer Vision.Piscataway:IEEE,2018: 708-725.
- [9] TAN Z, YANG Y, WAN J, et al. Attention-based pedestrian attribute analysis[J]. IEEE transactions on image processing, 2019, 28(12): 6126-6140.
- [10] WANG J, ZHU X, GONG S, et al. Attribute recognition by joint recurrent learning of context and correlation[C]//Proceedings of the 16th IEEE international conference on computer vision.Piscataway:IEEE,2017: 531-540.
- [11] WANG J, YANG Y, MAO J, et al. CNN-RNN: A unified framework for multi-label image classification[C]// Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition.Piscataway:IEEE,2016: 2285-2294.
- [12] ZHAO X, SANG L, DING G, et al. Grouping attribute recognition for pedestrian with joint recurrent learning[C]//Proceedings of the 27th international joint conference on artificial intelligence.New York:Curran Associates,2018: 3177-3183.
- [13] SUN Y, ZHENG L, YANG Y, 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.Berlin:Springer,2018: 501-518.

【作者简介】

陈楠(1997—)，女，陕西西安人，硕士研究生，工程师，研究方向：图像处理。

(收稿日期：2023-12-01)