

基于深度学习的非实验室场景人脸属性识别

葛宏孔 罗恒利 董佳媛

(南京航空航天大学计算机科学与技术学院 南京 211106)

摘要 非实验室场景下的人脸图片数量巨大,更加贴近生活,对其进行识别具有较大的研究价值。文中对非实验室环境下的人脸属性识别问题进行了研究,提出了一种人脸属性识别网络(Regional Multiple Layer Attributes Related Net, RMLARNet),不仅对人脸特征的提取方式进行了研究,还挖掘了人脸属性间的关系。该网络由 3 个部分组成:1)将人脸图像分割成包含属性部位的多个局部区域,并将这些局部区域作为输入提取特征信息;2)以 Inception V3 为迁移模型,采取多个不相邻卷积层迁移方式提取人脸特征;3)搭建了一个以人脸属性关系为约束的属性识别网络。实验结果表明,对 CelebA 数据集进行筛选处理,创建属性样本较平衡的 CelebA-数据集,并在该数据集上设计实验将取得优于现有方法的实验效果。

关键词 人脸属性识别,多标签任务,深度学习,迁移学习,属性约束

中图分类号 TP391.4 **文献标识码** A

Face Attributes in Wild Based on Deep Learning

GE Hong-kong LUO Heng-li DONG Jia-yuan

(School of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China)

Abstract Faces in the wild are huge in number and more close to life, and the recognition of facial attributes is a valuable research. A face attributes recognition method named RMLARNet (Regional Multiple Layer Attributes Related Net) was proposed for faces in the wild, which explores a new feature extraction method and attributes relationship. The processing steps of this method are as follows: 1) Feature extraction is based on the regional parts of image. 2) Features are extracted from different layer of Inception V3, and they are concatenated to get the final face feature. 3) An attributes relationship related network is used for attributes recognition. The experiment is conducted on a balanced CelebA-data set which is a subset of CelebA, and this method outperforms state-of-the-art methods.

Keywords Face attributes recognition, Multi-label task, Deep learning, Transfer learning, Attributes constraint

1 引言

人脸属性表征人脸特征的一系列生物特性,具有很强的自身稳定性和个体差异性,标识了人的身份,包括年龄、人种、性别、表情等。人脸属性识别是一个具有研究价值的课题,可以应用到多个人脸识别领域,如人脸确认^[1-2]和人脸鉴别。识别非实验室自然环境下的人脸属性有一定难度,因为这样条件下的人脸受到各种因素干扰,图 1 是受到遮挡等干扰因素影响的图片。



图 1 非实验室场景下的干扰因素

根据人脸图像的使用区域,将人脸属性识别方法分为全局特征方法和局部特征方法。全局方法对一整张人脸图像提取特征,识别人脸属性。局部方法对人脸图像进行分割,提取

局部特征,已有一些局部方法^[1,3-4]可以侦测人脸关键点,但 these methods for faces in the wild are not accurate enough. In recent years, there has been research on local methods based on deep learning methods^[5,19], using convolutional neural networks to identify face regions in non-laboratory environments, and achieved good recognition results. The focus of this paper is not on face detection, but on using face keypoint information as prior knowledge, focusing on designing face attribute recognition models.

近年来关于提取特征的研究不断发展。一种传统的方法是提取人工算子,如 Kumar 等^[1]在进行人脸识别时提取的是 HOG-like 算子,这些人工算子存在活力不强的弱点,会受到主观设计的影响。随着深度学习在计算机视觉领域的快速发展,已经可以通过搭建深层精妙的神经网络来自动学习活力强的特征。Donahue 等^[6]利用多重姿势归一化 CNNs 提取的集成学习特征,与传统人工算子相比可以带来更好的实验结果。Srinivas 等^[7]通过搭建设计的 CNNs 提取人脸特征进行性别、年龄和人种的预测。虽然通过搭建深层卷积网络可以提高最终的识别效果,但过深网络的训练速度较慢并且需要大量训练数据作为支撑。为了解决大型网络训练时间过长的问

本文受国家自然科学基金(61772268)资助。

葛宏孔(1994-),男,硕士生,主要研究方向为人脸属性识别, E-mail: 18549830154@163.com(通信作者);罗恒利(1994-),男,硕士生,主要研究方向为人脸聚类;董佳媛(1995-),女,硕士生,主要研究方向为图像分类。

题,许多学者利用迁移学习方法来处理多实例多标签问题的研究^[8-9]。Dong 等^[9]将 VGGNet16 模型迁移到他们的研究中。虽然迁移学习可提速,但最终的属性预测正确率与自主搭建神经网络稍微差了一些。

本文提出了一种将迁移学习与自搭建神经网络结合,并且加属性约束的人脸属性识别网络模型。该网络模型的设计重点体现在如下方面。

首先,人脸属性与人脸局部区域之间存在联系,考虑到这种关联性,本文将人脸图像分割成多个局部区域。其次,在人脸特征提取阶段,探究了新的迁移学习迁移方式。考虑到卷积神经网络中各层的不同抽象级别信息,将从神经网络的多层提取联合特征,而不只是迁移最后一个全连接层。最后,探究人脸属性之间的关联性,并将属性关系约束加入属性识别网络。最终,该论文搭建了一个多属性网络模型,并加入了辅助属性关系约束。Ghosh 等^[10]在他们的研究中也考虑到了多人脸属性关系,但仅在特征提取阶段,与属性识别阶段分离。在训练的过程中,在学习不同人脸属性之间关系的基础上训练每一种人脸属性。

在人脸数据集的选择上,该论文创建了名为 CelebA-的数据集,其图片来自拥有约 20 万张照片、40 种属性的数据集 CelebA^[6]。CelebA-的图片数目(约 6 万张)更少、关注的人脸属性(8 种)更少,并且各属性标签样本更加平衡。

本文第 2 节介绍了 RMLARNet 网络的框架和设计原理;第 3 节介绍了 CelebA-数据集的创建过程,并进行了对照实验的实验结果分析;最后总结全文并展望未来。

2 人脸属性识别网络模型

CelebA-数据集共有 8 种人脸属性(Bald, Bangs, Blond Hair, Eyeglasses, Goatee, Male, Smiling 和 Wearing Hat),每种属性有正反(1,0)两个真值。将每种人脸属性的识别视为一个二元分类问题,通过搭建神经网络来处理这个问题。图 2 展示了 RMLARNet 模型框架。

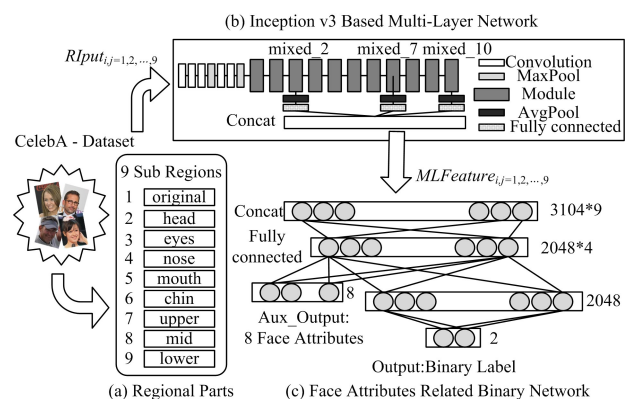


图 2 RMLARNet 模型框架图

图 2(a)所示的 Regional Parts 部分利用人脸五点信息将人脸划分为 9 个区域。图 2(b)所示的 Inception V3 Based Multi-Layer Network(简称 I3MLNet)部分,改进了 Inception V3^[12]的迁移结构,采取多层迁移的方式得到人脸特征。图 2(c)所示的 Face Attributes Related Binary Network(简称 FABNet)部分,加入了辅助人脸属性关系约束,此处是整个 RMLARNet 中唯一需要训练参数的部分。输入层的是人脸特征,输出层有两个,辅助输出层 Aux_Output 是 8 个人脸属性真值,识别输出层 Output 是指定识别人脸属性的二元值。

2.1 局部分割

本节研究分割原理及 Regional Parts 的分割方法。将人脸图像分割成包含人脸多个部位的局部区域是有理可寻的。人脸属性较多,如是否佩戴眼镜、是否微笑、是否有胡子、是否为男性等。对于一张人脸图片来说,很多时候仅需关注某一部局区域就可以进行相关属性的识别,如判断是否佩戴眼镜只需关注眼睛附近的区域、判断是否微笑只需关注嘴巴附近的区域,因此重点关注这些与属性相关的局部区域,排除不相关对象的影响,能够提升属性的识别效果。一些人脸属性的识别需要多个不连续的局部区域,如性别识别,很难说究竟是关注哪个区域,可能与胡子、头发、眼睛、眉毛等多个区域相关。

本文用迁移学习来提取特征,既然选择了这种方式,就要尽可能解决迁移模型不能很准确地捕捉跨数据集上的局部特征这一问题。一种可行的方式就是对人脸图片进行预处理,将人脸图片分割成与属性相关的多个局部区域。本文不采用神经网络捕捉局部区域的方式,而是利用人脸五点信息对人脸图片进行分割,分割的原则是确保各个局部区域尽可能地包含属性信息。例如,图 3(b)是以人脸五点信息为依据划分的局部区域,图 3(c)是经过神经网络捕捉到的局部区域。从多实例多标签分类问题角度出发,图 3(c)区域就是实例区域,而图 3(b)区域是包含实例的大实例区域,相当于对实例进行初步预处理,这一处理过程有利于之后进行人脸属性识别。设计这样的分割方式的原因是人脸图像的特殊性,人脸属性识别的判定区域都是围绕人脸五点,利用人脸五点信息就可以将人脸图像分割成多个区域。



图 3 两种人脸图像局部区域分割方式

分割方式如图 4 所示,这些区域关注了不同级别的特征,既有局部特征也有全局特征。人脸图片被分割成头部、眼部、鼻部、嘴部、下巴部、上半面部、中间面部以及下半面部,这些局部区域的分割利用了人脸五点坐标(左眼、右眼、鼻子、左嘴角以及右嘴角)。



图 4 以人脸五点为依据的人脸局部区域分割示意图

图 4(a)所示的原始人脸图像尺寸是 178×218 ,图 4(b)所示的头部区域以左眼为底往上取到第 20 像素处,图 4(c)所示的眼睛区域以鼻子为底往上取到右眼上方 30 像素处,图 4(d)所示的鼻部区域取左嘴角和左眼之间的区域,图 4(e)所示的嘴部区域以鼻子为顶向下取到左嘴角下方 30 像素处,图 4(f)所示的下巴区域以左嘴角为顶向下取到第 208 像素处,图 4(g)所示的上半面部取鼻子以上区域,图 4(h)所示的中间面部取左嘴角下方 30 像素到左眼上方 30 像素之间的区域,图 4(i)所示的下半面部取鼻子以下区域。由于每张人脸图片中的五点坐标不同,因此最终切出的局部区域尺寸不统一。

2.2 多层迁移特征提取网络

本节研究多层迁移的原理以及 I3MLNet 的具体迁移方式。人脸属性识别是多实例多标签分类问题,近年来的研究表明 CNNs 很适合用来处理这类问题。从多实例出发,卷积层的过滤器滑动扫描一张图片并创建备选实例对象,再由池化层选出最具有表征性的实例对象;从多标签出发,深层网络结构让所有标签的不同抽象级别特征都被嵌入其中,通过这些共享的神经元可以深入研究这些标签之间的关系。因此,本文选择用 CNNs 来处理人脸属性识别问题,为了更快地进行特征提取,这里迁移了已经训练好的网络模型。可迁移的模型很多,如 VGGNet16^[12]和 Inception V3 等,这些模型都是活力很强的,不仅在 Imagenet 上的分类效果很好,而且实验证明将它们迁移到其他数据集上的分类效果也较好。

一种通用的迁移方式是在新的数据集上进行参数微调,或者固定住模型参数,将 softmax 层改成多个二元分类层,仅训练这一层的参数。本文不采用这两种方式,而是以 Inception V3 为基准网络,提出了一种全新的迁移方式,尽可能地提高迁移模型提取跨数据集图片特征的活力。

文献[9,13-14]表明,神经网络中所有层提取的特征都对最终的分类有帮助。网络模型的层次结构设计,嵌入了不同抽象级别的特征。离输出层近的网络层相当于语义侦测器,它们更加贴近标签的语义描述空间,与标签的关系更密切。离输入层近的网络层是关于标签的低级抽象,更加关注输入层的某些特征,因此可以充分利用这些不同抽象级别的特征。Dong 等^[9]提取了 VGGNet16 中的每一层,然而这种方式不适合 Inception V3,因为这两个模型在结构上的差异很大。

Inception V3 是一个拥有非常精妙设计和构造的极深的网络结构,如图 2(b)所示,前面有 5 个卷积层和 2 个池化层,后面紧接着 11 个 Inception 模块,每个模块中又有很多的卷积层和池化层。I3MLNet 网络提取的是第 3、8 和 11 个模块(即 mixed_2, mixed_7, mixed_10),这样设计是有原因的。首先,若将 11 个模块都做提取,特征的尺寸会变得巨大,过大的尺寸将增加分类模型的参数,降低模型训练速度,这与选择迁移学习的初衷相悖。其次,深层层结构的网络中,距离太近的网络层之间的特征语义水平差别不是很大,表征性接近的特征对对最终的分类效果提升并不大,但会带来参数增加、速度降低等缺陷。最后,选择这三层与 Inception V3 的网络结构有关。将这 11 个模块分为 3 个模块组,前 3 个为第 1 模块组,中间 5 个为第 2 模块组,最后 3 个为第 3 模块组,每个模块组内部的模块设计基本一致,因此选择每个模块组的输出层。具体的迁移方式如下:

$$Concat_i \rightarrow Pool_i \rightarrow FC_i \rightarrow Concat_i, i=1,2,3 \quad (1)$$

其中, $Concat_i$ 依次表示图 2 中(b)的 mixed_2, mixed_7, mixed_10 模块组的输出层(尺寸分别为 $35 \times 35 \times 288, 17 \times 17 \times$

$768, 8 \times 8 \times 2048$); $Pool_{i=1,2,3}$ 是平均池化层; $FC_{i=1,2,3}$ 是全连接层(尺寸依次为 $288, 768, 2048$), $Concat_i$ 是将 $FC_{i=1,2,3}$ 拼接在一起的拼接层(尺寸为 3104),是经过 I3MLNet 网络提取出来的特征。Regional Parts 中的每一个区域都要依次进入 I3MLNet 提取人脸特征。

2.3 人脸属性识别网络

本节探究属性关系对 FABNet 人脸属性识别的影响。在多实例多标签分类问题中,探索标签之间的隐含关系将有助于分类学习。很明显,在人脸属性识别问题中,各种人脸属性的关系对属性预测起到重要作用。依据人脸属性的特点把这些属性分成两大类:第一类是可以追究到具体某个局部区域的属性,如微笑、涂口红对应到嘴部区域,戴眼镜、化眼妆对应到眼部区域,秃头、刘海对应到头部周围区域,胡子对应到脸颊下巴区域等;第二类是无法直接利用某个区域,需要利用与其他属性的关系做辅助来判断的属性,如性别的判定可能就要去关注长发、秃头、胡子、涂口红、化眼妆等属性,化妆就要去关注涂口红、化眼妆、打腮红等属性。属性关系不仅对第二类属性识别很重要,对于第一类属性也很重要,如一旦判定为秃头,就可以判定不为长发。各个人脸属性之间的共生关系是本文的重点专注关系,这里的共生关系就是属性倾向于同时出现或者彼此排斥。

多标签关系的研究^[11,15-16]展示了研究多标签关系的重要性。Zhang 等^[15]提出使用贝叶斯网络结构去探索训练集数据的标签关系,将得到的标签关系信息应用于在同一个训练集上训练的分类模型,这样就存在过拟合的风险。在人脸属性关系的捕获上,本文将采用神经网络来学习属性间的关系,挖掘属性间的隐含关系。为了避免人脸属性关系的捕获阶段与人脸属性的识别阶段的分离,也就是对于同一个训练集数据先捕获属性关系再训练属性识别模型造成的过拟合问题,将这两个阶段放在一起进行。

设计原则是在二元人脸属性识别网络中加入多属性关系约束,保证在考虑属性关系的同时训练属性识别网络。如图 2(c)所示, FABNet 网络是加入辅助人脸属性关系约束的网络,为了防止过拟合,设计的网络结构较小,设计如下:

$$MLFeature_{i,j} \rightarrow Concat \rightarrow FC_1 \rightarrow FC_2 \parallel (RELU_1 \rightarrow FC_3 \rightarrow FC_4), j=1,2,3,\dots,9 \quad (2)$$

其中, $MLFeature_{i,j}$ 是第 i 张图片第 j 个局部区域经过 I3MLNet 网络提取的多层特征; Concat 是拼接层,将 9 个多层特征拼接成一个联合特征。 $FC_{1,2,3,4}$ 是全连接层, $RELU_1$ 是激活函数。各层的尺寸在图 2(c)中标出。 FC_1 层之后有两条走向,一条是输出层(Aux_Output)尺寸为 8 的 FC_2 ,称为 Aux_Output 走向传播网络(AOPart);另一条是输出层(Output)尺寸为 2 的 $RELU_1 \rightarrow FC_3 \rightarrow FC_4$,称为 Output 走向传播网络(简称 OPart)。辅助输出 Aux_Output 的尺寸就是人脸属性的个数,即 8 种人脸属性,这一部分的作用是加入多属性关系约束,保证 FC_1 提出的特征能学习到属性关系。属性输出 Output 对应的是属性的二元真值,这一部分才是实现人脸属性识别的主体。这两个部分组成 FABNet 网络,训练时每一层的参数都需要更新,当训练完成时可以删去 FC_2 层,固定剩下的参数。

FARBNet 网络的训练过程中涉及到的损失函数来自两个方面:1)MARLoss 属性关系损失函数(Multiple Attr Relationship Loss);2)ABCLoss 属性二元分类损失函数(Attr Binary Classification Loss)。

MARLoss 的表达式如下:

$$MARLoss = -\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^Q y_{i,k} \log(p_{i,k}) \quad (3)$$

$$p_{i,k} = \frac{\exp(f_k(MLFeature_i))}{\sum \exp(f_k(MLFeature_i))}$$

其中, N 是训练集样本数, Q 是人脸属性个数, $y_{i,j} \in \{0, 1\}$ 表示第 i 张图片是否拥有第 k 个属性, $p_{i,k}$ 是第 i 张图片拥有第 k 个属性的概率值。 $f_k(MLFeature_i)$ 是以 $MLFeature_i$ 为输入, 经过 AOPart 得到的最后一层第 k 个神经元取值。

$ABCLoss_k$ 表示训练第 k 个属性分类器用到的损失函数, 表达式如下:

$$ABCLoss_k = -\frac{1}{N} \sum_{i=1}^N (y_{i,k} \log(q_{i,k}) + (1 - y_{i,k}) \log(1 - q_{i,k})) \quad (4)$$

$$q_{i,k} = \frac{\exp(g_k(MLFeature_i))}{\sum \exp(g_k(MLFeature_i))}$$

其中, $g_k(MLFeature_i)$ 是指以 $MLFeature_i$ 为输入, 经过 OPart 得到的第 k 个属性的预测值; $q_{i,k}$ 表示第 i 张图片拥有第 k 个属性的概率值, $q_{i,k}$ 和 $p_{i,k}$ 是经过不同走向传播网络得到的, 概率值是不同的。

最终使用的损失函数将这两个损失函数相结合, 并加上权重, 表达式如下:

$$TLoss_k = \alpha MARLoss + (1 - \alpha) ABCLoss_k \quad (5)$$

其中, 权重 α 是通过设计 FARBNet 网络框架实现的, 将 MARLoss 约束的 Aux_Output 接在 FC_1 后, 将 $ABCLoss_k$ 约束的 Output 接在 FC_3 后, 避免 MARLoss 对模型有过大的影响。 FABNet 网络实质上就是一个加了约束的二元分类器, 因此对于每种人脸属性的识别, 都要训练一个模型, 损失函数用到的就是这里的 $TLoss_k$ 。

3 实验

3.1 实验数据

CelebA^[5] 是一个非实验室环境下的人脸图片数据集, 共拥有 20 多万张人脸图片, 每张人脸标注 40 种人脸属性和 5 个人脸关键点。 CelebA 存在属性正负样本不平衡的现象, 有 29 种属性的正反样本数目差距超过 3 倍, 其中有 14 种属性正反样本数目差距 9 倍之多。

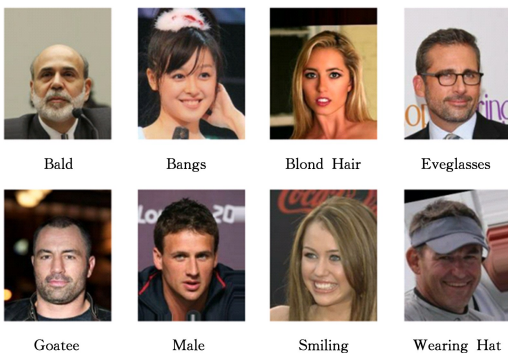


图 5 8 个属性对应的人脸样本

表 1 CelebA-数据集的样本分布

| 人脸属性 | 正样本数目 | 负样本数目 |
|-------------|--------|--------|
| Bald | 4 522 | 61 760 |
| Bangs | 11 402 | 54 880 |
| Blond Hair | 10 969 | 55 313 |
| Eyeglasses | 8 006 | 58 276 |
| Goatee | 7 882 | 58 400 |
| Male | 32 508 | 33 774 |
| Smiling | 31 242 | 35 040 |
| Wearing Hat | 6 458 | 59 825 |

本文创建了一个样本数目相对较小并且各属性标签样本数目更加平衡的 CelebA-数据集。选取了 8 种人脸属性 (Bald, Bangs, Blond Hair, Eyeglasses, Goatee, Male, Smiling 和 Wearing Hat), 图 5 中 8 个属性值都为 1。

在每种人脸属性中随机抽取 5 000 张正负样本 (Bald 抽取 4 500 张), 删除重复样本。该数据集共有 66 282 张人脸图片, 各属性的正反样本数目如表 1 所列。最后, 以 8:1:1 的比例将 CelebA-划分成训练集、验证集和测试集, 并采用 10 次交叉验证的测试方法。

3.2 对比方法

本文选择 FaceTracer^[17], PANDA-w^[4], PANDA-l^[4] 以及 HLLBNet 与 RMLARNet 进行对比。 FaceTracer 先是提取重要脸部区域 (这些区域是根据人脸五点坐标得到的) 的 HOG 算子和颜色柱状图, 然后训练 SVM 进行人脸属性识别。 PANDA-w 和 PANDA-l 是在 PANADA^[4] 基础上拓展的方法, 通过集成多个 CNNs (用来提取多个对齐人脸局部区域的特征) 来提取集成特征, 然后训练 SVM 进行人脸属性识别。 PANDA-w 和 PANDA-l 的不同在于获取对齐人脸局部区域的方法, 前者利用先进的人脸检测^[18] 和人脸对齐技术^[19], 后者用的是人脸五点坐标。 HLLBNet 仅以原始不分割人脸图片作为输入, 人脸特征仅迁移 Inception V3 最后一个全连接层, 并且在属性二元分类模型的训练过程中不加入辅助多人脸属性关系约束。 为了防止数据集不同带来的差异, 这里所有的方法都是在 CelebA-数据集上训练的。 使用这些方法的原因是这些方法都是近年来在人脸识别领域取得巨大成果的方法。 本文利用 tensorflow (GUP 版本) 实现神经网络模型的搭建, 代码的编写选择 python, 开发环境为 CUDA9.0 CuDNN7.0。

3.3 结果分析

各方法的人脸属性识别结果如表 2 所列, 可以看到在 8 个人脸属性平均识别率上, RMLARNet 为 96, 比 FaceTracer (90)、PANDA-w (89)、PANDA-l (95) 和 HLLBNet (89) 都高。 在 2.3 节中人脸属性被分为两类, 结果显示 RMLARNet 不仅可以识别全局人脸属性 (如 Male), 对于局部属性 (如 Bald, Eyeglasses 等) 的识别也取得了很高的准确性, 可见其对于全局人脸属性和局部人脸属性的识别效果都很好。

表 2 各方法中 8 种人脸属性识别准确率的对比

| 方法 | Bald | Bangs | Blond Hair | Eyeglasses | Goatee | Male | Smiling | Wearing Hat | Average |
|------------|------|-------|------------|------------|--------|------|---------|-------------|---------|
| FaceTracer | 89 | 88 | 80 | 98 | 93 | 91 | 89 | 89 | 90 |
| PANDA-w | 92 | 89 | 81 | 94 | 86 | 93 | 89 | 91 | 89 |
| PANDA-l | 96 | 92 | 93 | 98 | 93 | 97 | 92 | 96 | 95 |
| HLLBNet | 94 | 84 | 90 | 91 | 86 | 95 | 81 | 91 | 89 |
| RMLARNet | 98 | 94 | 95 | 99 | 94 | 98 | 89 | 98 | 96 |

进一步观察这8个人脸属性的识别率,有一些人脸属性(如 Bald, Eyeglasses, Male 和 Wearing Hat)的预测值能达到98和99,有一些属性(如 Bangs, Blond Hair 和 Goatee)稍微低一些,达到94和95,还有如 Smiling 属性的识别要更低,为89,造成这种识别阶梯现象的原因如下。首先,数据集样本的人脸属性存在认知偏差标记,由于采取人工标注的方式,因此会与标注者对于属性的认知有关。如图6(a)所示,对于 Smiling 属性的识别,在没有咧嘴笑的情况下很难区分究竟是微笑还是不笑,这种特征的无区分性造成 Smiling 的识别率较低。然后,存在错误标记,即因失误而错误判断的属性,如图6(b)所示(本文剔除错误标记)。最后,人脸图片质量较低,如图6(c)所示,存在极端姿态(如90°侧脸就将很难捕获某些人脸属性)、关键区域遮挡(如一个光头戴帽子和一个短发戴帽子是一样的,这对于光头的识别有影响;又或者遮挡住嘴部无法识别关于嘴部的属性,如微笑、涂口红等)。不同属性之间的遮挡问题是无法避免的,但是可以通过修改错误标记和认知偏差标记,以及过滤噪声过大的人脸图片,进而使得人脸属性识别效果得到进一步提升。



图6 影响属性识别率的样本

结束语 本文提出了一种非实验室环境下人脸属性识别网络模型,是一种基于局部区域的多层迁移与人脸属性关系约束的网络模型。该模型关注了不同级别的特征,既关注到了局部特征也考虑到了全局特征,将人脸图片分割成包含人脸属性的多个局部区域;I3MLNet 多层迁移特征提取网络改进了 Inception V3 的迁移结构,采取多个不连续层的迁移方式;FABNet 人脸属性识别网络加入了辅助多人脸属性关系约束。实验结果证明,本文提出的模型有投入现实生活的可行性。未来的工作会考虑将自动捕捉非实验室场景下的人脸加入模型,进一步完善人脸识别模型,同时也将考虑增加属性的数目,进一步探究各属性之间的关系约束。

参考文献

[1] KUMAR N, BERG A C, BELHUMEUR P N, et al. Attribute and simile classifiers for face verification[C]// IEEE International Conference on Computer Vision. IEEE, 2009: 365-372.

[2] SONG F, TAN X, CHEN S. Exploiting relationship between attributes for improved face verification [J]. Computer Vision & Image Understanding, 2014, 122(4): 143-154.

[3] BOURDEV L, MAJI S, MALIK J. Describing people: A poselet-based approach to attribute classification[C]// IEEE International Conference on Computer Vision. IEEE, 2011: 1543-1550.

[4] ZHANG N, PALURI M, RANZATO M, et al. PANDA: Pose A-

igned Networks for Deep Attribute Modeling[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2014: 1637-1644.

[5] LIU Z, LUO P, WANG X, et al. Deep Learning Face Attributes in the Wild[C]// IEEE International Conference on Computer Vision. IEEE, 2015: 3730-3738.

[6] DONAHUE J, JIA Y, VINYALS O, et al. DeCAF: a deep convolutional activation feature for generic visual recognition[C]// International Conference on International Conference on Machine Learning. JMLR. org, 2014: I-647.

[7] SRINIVAS N, ATWAL H, ROSE D C, et al. Age, Gender, and Fine-Grained Ethnicity Prediction Using Convolutional Neural Networks for the East Asian Face Dataset[C]// IEEE International Conference on Automatic Face & Gesture Recognition. IEEE, 2017: 953-960.

[8] ZENG T, JI S. Deep Convolutional Neural Networks for Multi-instance Multi-task Learning[C]// IEEE International Conference on Data Mining. IEEE, 2016: 579-588.

[9] DONG M, PANG K, WU Y, et al. Transferring CNNs to multi-instance multi-label classification on small datasets[C]// IEEE International Conference on Image Processing. Beijing: IEEE, 2017: 1332-1336.

[10] GHOSH S, LAKSANA E, SCHERER S, et al. A multi-label convolutional neural network approach to cross-domain action unit detection[C]// International Conference on Affective Computing and Intelligent Interaction. IEEE, 2015: 609-615.

[11] SZEGEDY C, VANHOUCHE V, IOFFE S, et al. Rethinking the Inception Architecture for Computer Vision[C]// Computer Vision and Pattern Recognition. IEEE, 2016: 2818-2826.

[12] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J]. arXiv: 1409. 1556, 2014.

[13] ZOU W Y, ZHU S, NG A Y, et al. Deep learning of invariant features via simulated fixations in video[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc, 2012: 3203-3211.

[14] COATES A, NG A Y. Selecting receptive fields in deep networks[C]// International Conference on Neural Information Processing Systems. Curran Associates Inc, 2011: 2528-2536.

[15] ZHANG M L, ZHANG K. Multi-label learning by exploiting label dependency[C]// ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM, 2010: 999-1008.

[16] HUANG S J, ZHOU Z H. Multi-label learning by exploiting label correlations locally[C]// Twenty-Sixth AAAI Conference on Artificial Intelligence. AAAI Press, 2012: 949-955.

[17] KUMAR N, BELHUMEUR P, NAYAR S. FaceTracer: A Search Engine for Large Collections of Images with Faces [C]// European Conference on Computer Vision. Springer-Verlag, 2008: 340-353.

[18] LI J, ZHANG Y. Learning SURF Cascade for Fast and Accurate Object Detection[C]// Computer Vision and Pattern Recognition. IEEE, 2013: 3468-3475.

[19] SUN Y, WANG X, TANG X. Deep Convolutional Network Cascade for Facial Point Detection[C]// IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2013: 3476-3483.