



(12)发明专利

(10)授权公告号 CN 103824054 B

(45)授权公告日 2018.08.07

(21)申请号 201410053852.7

(22)申请日 2014.02.17

(65)同一申请的已公布的文献号
申请公布号 CN 103824054 A

(43)申请公布日 2014.05.28

(73)专利权人 北京旷视科技有限公司
地址 100080 北京市海淀区海淀大街3号1
幢1001-011室

(72)发明人 印奇 曹志敏 姜宇宁 杨东

(74)专利代理机构 北京君尚知识产权代理事务
所(普通合伙) 11200

代理人 余长江

(51)Int.Cl.

G06K 9/00(2006.01)

G06N 3/08(2006.01)

(56)对比文件

CN 102750824 A,2012.10.24,

CN 101615248 A,2009.12.30,

Erjin Zhou et al..“Extensive Facial
Landmark Localization with Coarse-to-fine
Convolutional Network Cascade”.《2013 IEEE
Conference on Computer Vision and Pattern
Recognition》.2013,第2013年卷第386-391页.

Zhimin Cao et al..“Face Recognition
with Learning-based Descriptor”.《2010
IEEE Conference on Computer Vision and
Pattern Recognition》.2010,第2010年卷第
2707-2714页.

审查员 丁园园

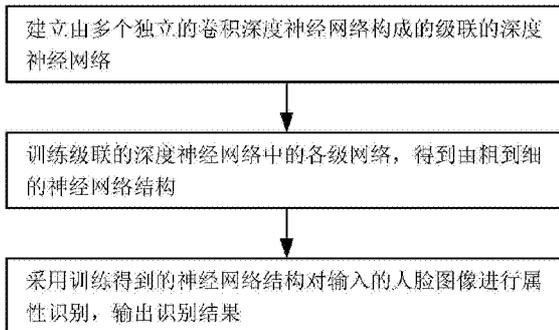
权利要求书1页 说明书6页 附图1页

(54)发明名称

一种基于级联深度神经网络的人脸属性识别方法

(57)摘要

本发明涉及一种基于级联深度神经网络的人脸属性识别方法,其步骤包括:1)建立由多个独立的卷积深度神经网络构成的级联的深度神经网络;2)采用大量人脸图像数据逐级训练所述级联的深度神经网络中的各级网络,将前一级网络的输出作为后一级网络的输入,得到由粗到细的神经网络结构;3)采用所述由粗到细的神经网络结构对输入的人脸图像进行属性识别,并输出最终的识别结果。本发明在深度学习的基础上引入了级联算法体系,加快训练时间,并且通过级联的由粗到细的处理过程,每层利用上层网络的信息提高了最终网络的性能,能够有效提高人脸属性识别的速度和准确率。



1. 一种基于级联深度神经网络的人脸属性识别方法,其步骤包括:

- 1) 建立由多个独立的卷积深度神经网络构成的级联的深度神经网络;
- 2) 采用大量人脸图像数据逐级训练所述级联的深度神经网络中的各级网络,将前一级网络的输出作为后一级网络的输入,得到由粗到细的神经网络结构;
- 3) 采用所述由粗到细的神经网络结构对输入的人脸图像进行属性识别,并输出最终的识别结果;

其中,所述级联的深度神经网络中每一级独立的卷积深度神经网络包含多层,包括:卷积层,最大采样层,unshared卷积层,全连接层,soft-max层;

其中,所述人脸属性为年龄,所述级联的深度神经网络为包括第一级网络和第二级网络的两级级联深度卷积神经网络;所述采用所述由粗到细的神经网络结构对输入的人脸图像进行属性识别并输出最终的识别结果包括:

将所述人脸图像输入到所述第一级网络进行处理以得到所述人脸图像的年龄段;

将所述人脸图像输入到所述第二级网络中与所述年龄段对应的深度神经网络进行处理,输出所述年龄段内的精确的年龄。

2. 如权利要求1所述的方法,其特征在于:对输入所述级联的深度神经网络的人脸图像进行预处理,包括标定和归一化处理。

3. 如权利要求2所述的方法,其特征在于:所述标定采用多个关键点,对输入的人脸图像进行对齐仿射变换以减少不同姿态对属性的影响;所述归一化的方法为:

$$I_{ij} = \tanh\left(\frac{I_{ij} - I_{mean}}{I_{std.}}\right),$$

其中 I_{ij} 为标定后的人脸矩形框 I 的位置 (i, j) 处的像素值, I_{mean} 为整个矩形框像素平均值, $I_{std.}$ 为像素标准差。

4. 如权利要求1所述的方法,其特征在于:所述级联的深度神经网络包括两级独立的卷积深度神经网络。

5. 如权利要求1所述的方法,其特征在于:所述年龄段为0~6岁,6~18岁,18~40岁,40~60岁,60+岁五个年龄段之一,所述第一级网络采用三层卷积神经网络。

一种基于级联深度神经网络的人脸属性识别方法

技术领域

[0001] 本发明属于图像处理和人脸识别技术领域,具体涉及一种基于级联深度神经网络的人脸属性识别方法。

背景技术

[0002] 人脸属性即从人的面部特征能够得到的人的性别,年龄,种族等属性。对人脸属性的识别,能够帮助人脸识别更加精确,且单独的人脸属性识别也有着很多的应用场景。传统的人脸属性识别方法采用人工设计的纹理算子加上SVM等传统分类器的浅层结构,往往得不到较为精确的预测效果。

[0003] 深度神经网络是近几年来比较热的一个研究方向,它从仿生学的角度模拟人脑的分多层计算架构体系,是最接近人工智能(AI)的一个方向,相对于SVM等传统的浅层机器学习架构,它更能表征一些复杂的模式与函数。近几年在语音识别及图像处理领域,深度学习均取了state-of-the-art的结果。但深度学习存在训练困难,训练周期长等缺点,虽然已经在人脸属性识别及分类方面得到应用,但人脸属性识别的精度和处理速度方面都不能很好的满足实际需求。

发明内容

[0004] 本发明针对上述的问题,提供一种基于级联深度神经网络的人脸属性识别方法,在深度学习的基础上引入了级联算法体系,加快训练时间,并且通过级联的由粗到细(coarse-to-fine)的处理过程,每层利用上层网络的信息提高了最终网络的性能。

[0005] 为实现上述目的,本发明采用的技术方案如下:

[0006] 一种基于级联深度神经网络的人脸属性识别方法,其步骤包括:

[0007] 1) 建立由多个独立的卷积深度神经网络构成的级联的深度神经网络;

[0008] 2) 采用大量人脸图像数据逐级训练所述级联的深度神经网络中的各级网络,将前一级网络的输出作为后一级网络的输入,得到由粗到细的神经网络结构;

[0009] 3) 采用所述由粗到细的神经网络结构对输入的人脸图像进行属性识别,并输出最终的识别结果。

[0010] 进一步地,所述级联的深度神经网络中每一级独立的卷积深度神经网络包含多层,包括:卷积层,最大采样层,unshared卷积层,全连接层,soft-max层。

[0011] 进一步地,对输入所述级联的深度神经网络的人脸图像进行预处理,包括标定和归一化处理。

[0012] 进一步地,所述人脸属性为下列中的一种:性别,年龄,种族。

[0013] 一种采用上述方法对人脸图像进行年龄识别的方法,其步骤包括:

[0014] 1) 对输入的人脸图像数据进行对齐仿射变换和归一化预处理,将人的年龄分为多个年龄段;

[0015] 2) 建立两级级联的深度卷积神经网络,将预处理后的人脸图像输入第一级网络,

通过多层卷积神经网络、全连接网络以及soft-max分类器,得到输入的人脸图像的年龄段;

[0016] 3)将第一级网络得到的人脸图像的年龄段输入第二级网络,通过卷积神经网络层、全连接网络层、线性回归层,输出所述年龄段内精确的年龄。

[0017] 进一步地,上述方法将人的年龄分为0~6岁,6~18岁,18~40岁,40~60岁,60+岁五个年龄段,所述第一级网络采用三层卷积神经网络。

[0018] 一种采用上述方法对人脸图像进行性别识别的方法,其步骤包括:

[0019] 1)建立由两级独立的卷积深度神经网络构成的级联的深度神经网络;

[0020] 2)采用大量包含不同性别的人脸图像数据训练所述级联的深度神经网络;

[0021] 3)采用训练得到两级级联的神经网络结构对输入的人脸图像进行性别识别,并输出最终的识别结果

[0022] 进一步地,所述第一级网络采用三层卷积神经网络,所述第二级网络与所述第一级网络的结构相同。

[0023] 本发明采用多层级联深度卷积神经网络模型,每个Level是一个独立的卷积深度神经网络,按照Level的递增,后续Level在前面Level基础上完成更精细的计算,完成又粗到细(coarse-to-fine)的过程;每个Layer为独立卷积深度神经网络的一层,它由卷积神经网络层,全连接层,soft-max层等组成,共同完成单个Level的工作。采用本发明的方法,能够加快训练时间,有效提高人脸属性识别的速度和准确率。

附图说明

[0024] 图1是本发明的基于级联深度神经网络的人脸属性识别方法的步骤流程图。

[0025] 图2是实施例中深度神经网络的结构示意图。

具体实施方式

[0026] 下面通过具体实施例和附图,对本发明做进一步说明。

[0027] 本发明的基于级联深度神经网络的人脸属性识别方法,其步骤流程如图1所示,首先建立由多个独立的卷积深度神经网络构成的级联的深度神经网络;然后采用大量人脸图像数据逐级训练所述级联的深度神经网络中的各级网络,得到由粗到细的神经网络结构;然后采用所述由粗到细的神经网络结构对输入的人脸图像进行属性识别,并输出最终的识别结果。下面进行具体说明。

[0028] 1. 预处理

[0029] 为了减少噪声及人类pose(姿态)等因素对人脸属性的较大影响,在进行分层的深度级联神经网络之前,我们对输入图像进行标定、归一化等处理,提升后续网络的性能。

[0030] 标定采用多个关键点,不失一般性地,本实施例使用5个关键点,输入人脸图像进行对齐仿射变换,减少不同pose对属性的影响。

[0031] 归一化的过程为:

$$[0032] \quad I_{ij} = \text{tanh} \left(\frac{I_{ij} - I_{\text{mean}}}{I_{\text{std}}} \right)$$

[0033] 其中 I_{ij} 为标定后的人脸矩形框I的位置(i, j)处的像素值, I_{mean} 为整个矩形框像素平均值, I_{std} 为像素标准差。

[0034] 2. 卷积深度神经网络

[0035] 卷积深度神经网络受生物学的启发,特别是Hubel和Wiesel早期关于猫视觉神经的研究结果而模仿的视觉系统,是一个自下而上的网络结构。它采用多层网络,逐层抽象,每一层在上层的基础上抽象出更高阶的应对各种不变性的纹理表示,达到在视觉识别或分类过程中应对各种图像变化更加鲁棒的结构。它每一层卷积网络的组成一般为多个二维平面组成(feature map),每个feature map由以下几个部分组成:

[0036] (1) shared或unshared卷积

[0037] i). shared卷积采用共享的权值对整个可视域进行扫描构成一个单独的映射(feature map),这样大大减少了网络的参数且能带来一定的平移,缩放不变性。

[0038] ii). unshared卷积在不同的可视域位置采用不同的卷积核,我们在较为上层的网络层中使用这种卷积,以区别对待各个不同区域高级抽象表达对后续网络的影响。

[0039] (2) 非线性变换

[0040] 非线性变换模仿神经元的非线性作用,产生刺激,常用的有sigmoid,tanh函数等。

以tanh为例,第k个feature map非线性变换的输出为:

$$[0041] \quad h_{ij} = \tanh((W^k * x)_{ij} + b_k),$$

[0042] 其中,x为卷积神经网络层的输入, W^k 为第k个feature map对应的卷积核, b_k 为第k个feature map的偏差, h_{ij} 为输入x经过卷积和非线性变换后的输出。

[0043] (3) 聚合池化(pooling/sub-sampling)

[0044] pooling的过程是一个下采样的过程,常用的pooling过程有max pooling,average pooling等。pooling的目的在于在减少下层网络的输入,从而减少计算量的同时,保证网络具有一定的平移不变性,更加鲁棒。

[0045] 通过上述的卷积网络训练得到的图像纹理表达输入linear regression或者soft-max等算法即可完成各种分类及识别任务。一个典型的卷积神经网络如图2所示,包括:卷积层(convolutional,简称为conv.,或者称为卷积神经网络层),最大采样层(maxpooling,简称为maxp.),unshared卷积层(unshared conv.),全连接层(fully connected),后续还可以级联soft-max层(soft-max分类器)等等。

[0046] 3. 级联算法

[0047] 深度卷积神经网络由于其网络的复杂度,存在训练困难,训练周期长,较难优化等缺点。为了克服这些缺点,本发明在深度卷积神经网络的基础上引入了级联算法。级联算法的出发点是将模型分为多层,每层在上层较为模糊的结果的基础上进一步提高性能,逐级提升,由粗到精提升整个模型的性能。

[0048] 鉴于深度学习训练的复杂性以及人脸属性识别的特殊性。本发明结合卷积神经网络与级联算法,创造一种新的算法架构,将原始的深度神经网络分解为多个复杂度较低的深度神经网络(即一个单独的Level)。在训练的过程中,我们逐级训练各级网络。由这个逐步逼近,由粗到细训练过程,最后得到较原始复杂度很高的网络更优的网络结构,且减少了训练时间,加快了工程应用。我们将这种架构用于人脸属性的鉴别,取得较单独卷积神经网络更好的性能。不失一般性,这种架构在人脸属性以外的应用场景也能大大提升系统的性能。

[0049] 本发明采用的多层级联深度卷积神经网络模型,其包括以下结构:

[0050] 1) Level: 每个Level是一个独立的卷积深度神经网络,按照Level的递增,后续Level在前面Level基础上完成更精细的计算,完成coarse-to-fine的过程;

[0051] 2) Layer: 每个Layer为独立卷积深度神经网络的一层,它由卷积神经网络层,全连接层,soft-max层等组成,共同完成单个Level的工作。

[0052] 模型训练过程:

[0053] 因为采用了分级(Level)的结构,本发明在训练时,先训练第一级Level1。如果Level1不能继续收敛,且模型效果未达到技术指标的情况下,在Level1基础上增加一级Level1,继续训练Level1。以此类推,直到训练的某层达到技术指标要求不再增加下一级。

[0054] 4. 采用级联深度卷积神经网络进行人脸属性分析和识别

[0055] 1) 年龄预测:

[0056] 在年龄的预测中我们对比了浅层网络SVM+FEATURE SELECT,深层卷积神经网络(CNN LEARNING),以及本发明的级联深度卷积神经网络(CNN LEARNING+CASCADE)三种算法的计算复杂度与性能。

[0057] 1.1) 网络结构

[0058] 我们采用的级联深度卷积神经网络的结构为:

[0059] (1) 预处理使用alignment算法得到60x60的人脸图片,然后进行归一化处理;

[0060] (2) 两级级联深度卷积神经网络:

[0061] i) Level1:

[0062] 将人的年龄分为0~6岁,6~18岁,18~40岁,40~60岁,60+岁五个年龄段,网络采用3层卷积神经网络,每层神经网络的配置如下:

[0063] layer1: 输入: 预处理后的人脸图像; 输出: 5x5shared卷积+max pooling生成20个feature map;

[0064] layer2: 输入: layer1的20个feature map; 输出: 5x5shared卷积+max pooling生成40个feature map;

[0065] layer3: 输入: layer2的40个feature map; 输出: 3x3shared卷积+max pooling生成60个feature map。

[0066] 上述三层卷积神经网络后,将layer3的输出进行串行化(flatten),作为后续级联的全连接网络的输入,全连接网络输出500维数据,再传给后续级联的soft-max分类器进行分类,得到最后输入图像的年龄段。

[0067] ii) Level2

[0068] 在Level1输出的五个年龄段,各级联一个较为简单的深度神经网络:

[0069] layer1: 卷积神经网络层,输入: 预处理后的人脸图像; 输出: 5x5shared卷积+max pooling生成20个feature map;

[0070] layer2: 全连接网络层,输入: layer1flatten后数据; 输出: 200维数据;

[0071] layer3: 线性回归(linear regression),输入: layer2输出; 输出: 该年龄段内精确的年龄。

[0072] 1.2) 实验结果

[0073] 我们使用以上介绍的级联深度卷积神经网络(CNN LEARNING+CASCADE)与非级联的深度卷积神经网络(CNN LEARNING),基于SVM分类器的特征选择算法(SVM+FEATURE

SELECT) 进行对比测试, 结果如表1所示。

[0074] 表1.age预测实验结果对比列表

[0075]

方法	Data Tang错误率	CAS错误率
SVM+FEATURE SELECT	0.24	0.40
CNN LEARNING	0.20	0.30
CNN LEARNING+CASCADE	0.15	0.25

[0076] 由表1可见, 级联深度卷积神经网络在我们内部测试集上的表现均比SVM+FEATRUE这种一般的浅层网络好, 而本发明中使用的级联算法在CNN LEARNING的基础上有所提升, 且因为级联的coarse-to-fine过程, 神经网络的复杂度较非级联的网络低很多, 训练时间大大减少。

[0077] 2) 性别预测

[0078] 性别检测实验中, 我们同样对比了SVM+FEATURE SELECT, 深层卷积神经网络(CNN LEARNING), 以及本发明的级联深度卷积神经网络(CNN LEARNING+CASCADE) 三种算法的计算复杂度与性能。

[0079] 2.1) 网络结构

[0080] 我们采用的级联深度卷积神经网络的结构为:

[0081] (1) 预处理使用alignment算法得到60x60的人脸图片, 然后进行归一化处理;

[0082] (2) 两级级联深度卷积神经网络:

[0083] i) Level1:

[0084] Level1采用三层的深度卷积神经网络, 每层神经网络的配置如下:

[0085] layer1: 输入: 预处理后的人脸图像; 输出: 5x5shared卷积+max pooling生成20个feature map;

[0086] layer2: 输入: layer1的20个feature map; 输出: 5x5shared卷积+max pooling生成40个feature map;

[0087] layer3: 输入: layer2的40个feature map; 输出: 3x3unshared卷积+max pooling生成80个feature map。

[0088] 上述三层卷积神经网络后, 将layer3的输出进行串行化(flatten), 作为后续级联的全连接网络的输入, 再传给后续级联的soft-max分类器进行分类。

[0089] ii) Level2: 采用与Level1相同的结构。

[0090] 2.2) 训练过程

[0091] 本实验使用的男女各10w张图片作为训练样本, 首先训练Level1, 当Level1不再收敛后, 根据Level1分类结构, 增加分类错误的样本的在样本集里的权重, 将所有样本送入Level2继续训练。

[0092] 2.3) 实验结果

[0093] 本实验对比上述3种方法在内部测试集上的性能, 结果如表2所示, 发现在性别检测上, 本发明算法仍然具有明显的优势。

[0094] 表2: 性别检测实验结果

[0095]

方法	错误率
SVM+FEATURE SELECT	0.06
CNN LEARNING	0.04
CNN LEARNING+CASCADE	0.03

[0096] 以上实施例仅用以说明本发明的技术方案而非对其进行限制,本领域的普通技术人员可以对本发明的技术方案进行修改或者等同替换,而不脱离本发明的精神和范围,本发明的保护范围应以权利要求所述为准。

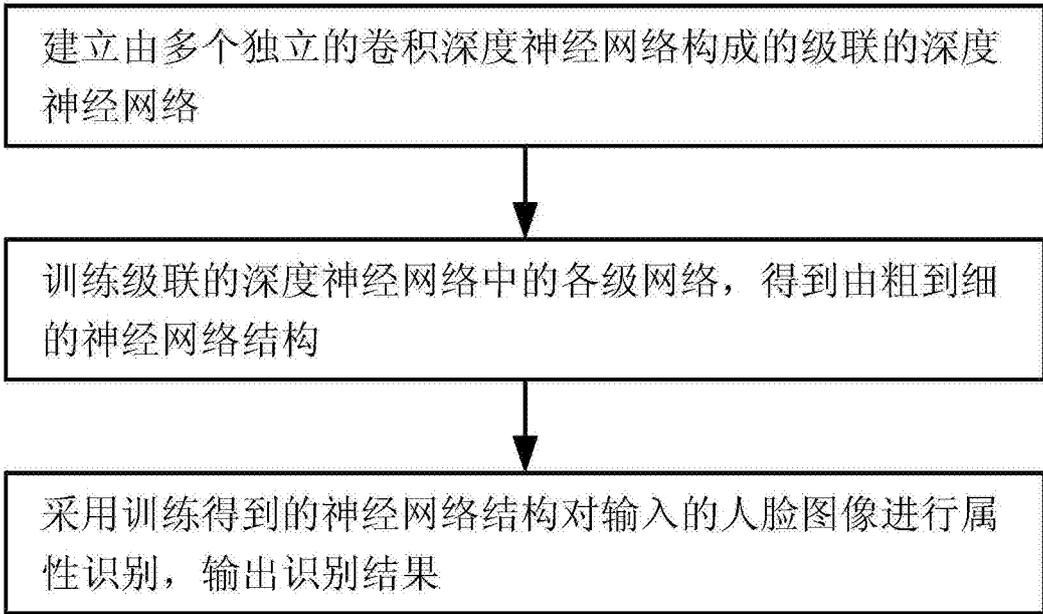


图1

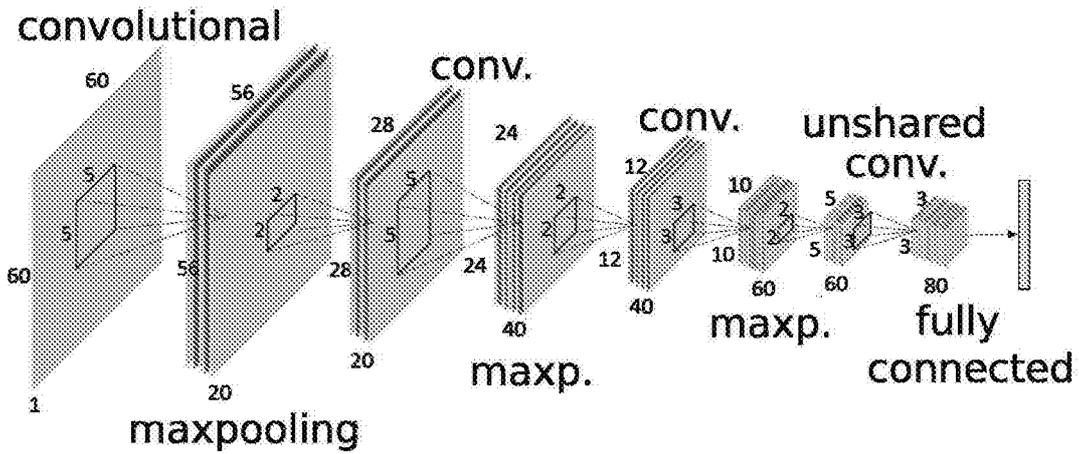


图2