



(12)发明专利

(10)授权公告号 CN 105981051 B
(45)授权公告日 2019.02.19

(21)申请号 201480075091.3
(22)申请日 2014.10.10
(65)同一申请的已公布的文献号
 申请公布号 CN 105981051 A
(43)申请公布日 2016.09.28
(85)PCT国际申请进入国家阶段日
 2016.08.08
(86)PCT国际申请的申请数据
 PCT/CN2014/088285 2014.10.10
(87)PCT国际申请的公布数据
 W02016/054802 EN 2016.04.14
(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
(72)发明人 印奇 曹志敏 周以苏
(74)专利代理机构 北京市柳沈律师事务所
 11105
 代理人 于小宁

(51)Int.Cl.
 G06K 9/62(2006.01)
(56)对比文件
 CN 103345774 A,2013.10.09,
 US 2004183451 A1,2004.09.23,
 CN 101425177 A,2009.05.06,
 Erjin Zhou etc..“Extensive Facial
 Landmark Localization with Coarse-to-fine
 Convolutional Network Cascade”.《2013 IEEE
 International Conference on Computer
 Vision Workshops》.2014,第386-391页.
 Mojtaba Seyedhosseini etc..“Image
 Segmentation with Cascaded Hierarchical
 Models and Logistic Disjunctive Normal
 Networks”.《2013 IEEE International
 Conference on Computer Vision》.2014,第
 2168-2175页.

审查员 李玥

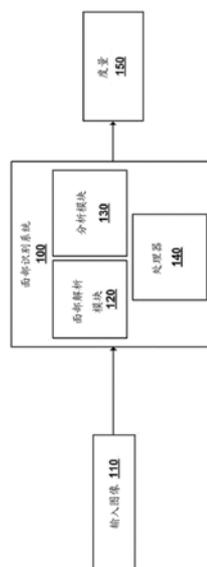
权利要求书3页 说明书10页 附图9页

(54)发明名称

用于图像解析的分层互连多尺度卷积网络

(57)摘要

所公开的面部识别系统(及方法)包括面部解析。在一种方法中,面部解析基于分层互连多尺度卷积神经网络(HIM)来识别面部图像的组成部分的位置和/或印迹。HIM从面部图像的不同分辨率图像生成多个级别的图像块,其中不同级别的图像块具有不同的分辨率。而且,HIM合并不同级别的图像块以生成不同级别的互连图像块,其中不同级别的互连图像块具有不同的分辨率。而且,HIM组合互连图像块以识别组成部分的精确化的位置和/或印迹。



1. 一种将图像解析成组成部分的系统,该系统包含
用于从图像定位组成部分的分层互连多尺度卷积神经网络(HIM),该HIM包含:
级别生成器,被配置为接收图像以及从所述图像生成N个级别的图像块,其中 $N > 2$,其中,不同级别n的图像块具有不同的分辨率 $R(n)$,并且级别n的图像块从被重新采样至分辨率 $R(n)$ 的所述图像生成;
互连组合器,被配置为接收来自所述级别生成器的所述N个级别的图像块以及从所述N个级别的图像块生成M个级别的互连图像块,其中 $2 < M \leq N$,其中,不同级别m的互连图像块具有不同的分辨率 $R(m)$,并且级别m的互连图像块从图像块的输入分组m生成,所述图像块的输入分组m包含(i)来自 $R(n) = R(m)$ 的级别n的图像块,以及(ii)来自 $R(n) \neq R(m)$ 的一个或多个级别n的图像块,其中这样的图像块已经被重新采样至分辨率 $R(m)$;以及
聚合器,被配置为通过组合所述M个级别的互连图像块来定位组成部分,
其中,所述聚合器被配置为接收所述M个级别的互连图像块以及生成L个级别的输出图像块,其中 $L < M$,其中,不同级别l的输出图像块具有不同的分辨率 $R(l)$,每个 $R(l)$ 对应于 $R(m)$ 中的一个,并且级别l的输出图像块从图像块的输入分组1生成,图像块的输入分组1包含(i)来自 $R(m) = R(l)$ 的级别m的互连图像块,以及(ii)来自具有小于 $R(l)$ 的分辨率的一个或多个级别的输出图像块,其中这样的输出图像块已经被重新采样至分辨率 $R(l)$ 。
2. 根据权利要求1所述的系统,其中,所述级别生成器通过对被重新采样至分辨率 $R(n)$ 的图像执行卷积和非线性变换来生成级别n的图像块。
3. 根据权利要求1所述的系统,其中,所述互连组合器通过对图像块的输入分组m执行卷积和非线性变换来生成级别m的互连图像块。
4. 根据权利要求1所述的系统,其中,所述聚合器通过对图像块的输入分组1执行卷积和非线性变换来生成级别1的输出图像块。
5. 根据权利要求1所述的系统,其中,所述聚合器基于最高分辨率的输出图像块来定位组成部分。
6. 根据权利要求1所述的系统,其中, $M = N$ 并且 $L = M - 1$ 。
7. 根据权利要求6所述的系统,其中,级别n=1具有与所述图像的分辨率相同的分辨率 $R(n=1)$,并且每个其他级别n具有相对于具有分辨率 $R(n-1)$ 的前一个级别n-1按照常量因子被向下采样的分辨率 $R(n)$ 。
8. 根据权利要求7所述的系统,其中,常量因子是2倍或以下的向下采样。
9. 根据权利要求7所述的系统,其中,所述级别生成器生成具有较低分辨率 $R(n)$ 的级别n的更多的图像块。
10. 根据权利要求6所述的系统,其中,所述互连组合器从图像块的输入分组m生成级别m的互连图像块,图像块的输入分组m包含(i)来自 $R(n) = R(m)$ 的级别n的图像块,(ii)存在的来自具有比 $R(m)$ 次高的分辨率的级别n以及具有比 $R(m)$ 次低的分辨率的级别n的图像块,其中这样的图像块已经被重新采样至分辨率 $R(m)$ 。
11. 根据权利要求6所述的系统,其中,所述聚合器从图像块的输入分组1生成级别1的输出图像块,图像块的输入分组1包含(i)来自 $R(m) = R(1)$ 的级别m的互连图像块,以及(ii)来自具有比 $R(1)$ 次低的分辨率的级别m的输出图像块,其中这样的输出图像块已经被重新采样至分辨率 $R(1)$ 。

12. 根据权利要求1所述的系统,还包含:

分层互连多尺度卷积神经网络的第二级,被配置为接收由所述HIM产生的组成部分位置,所述第二级中的每个分层互连多尺度卷积神经网络被配置为进一步精确化所述组成部分中的一个的位置和/或印迹。

13. 根据权利要求1所述的系统,其中,所述图像是面部图像,并且所述组成部分是包括眼睛、眉毛、鼻子和嘴的面部组成部分。

14. 根据权利要求13所述的系统,其中,所述聚合器产生指示所述面部组成部分的位置的二进制标签图。

15. 根据权利要求13所述的系统,还包含:

分层互连多尺度卷积神经网络的第二级,被配置为接收由所述HIM产生的面部组成部分位置,所述第二级中的分层互连多尺度卷积神经网络中的至少一个被配置为分别进一步精确化眼睛、鼻子和嘴的位置和/或印迹。

16. 根据权利要求15所述的系统,还包含:

图像提取器,耦合在所述HIM与所述第二级之间,所述图像提取器用于基于所述面部图像和由所述HIM产生的所述面部组成部分的所述位置来提取面部组成部分图像,所提取的面部组成部分图像被提供给所述第二级。

17. 根据权利要求13所述的系统,还包含:

输出生成器,被配置为在面部图像上重叠面部组成部分的位置。

18. 根据权利要求1所述的系统,其中:

$M=N$ 并且 $L=M-1$;

级别 $n=1$ 具有与图像的分辨率相同的分辨率 $R(n=1)$,并且每个其他级别 n 具有相对于具有分辨率 $R(n-1)$ 的前一个级别 $n-1$ 按照2或以下的常量因子被向下采样的分辨率 $R(n)$;

所述级别生成器通过对被重新采样至分辨率 $R(n)$ 的图像执行卷积和非线性变换来生成级别 n 的图像块,并且所述级别生成器生成具有较低分辨率 $R(n)$ 的级别 n 的更多的图像块;

所述互连组合器通过对图像块的输入分组 m 执行卷积和非线性变换来生成级别 m 的互连图像块,图像块的输入分组 m 包含(i)来自 $R(n)=R(m)$ 的级别 n 的图像块,以及(ii)存在的来自具有比 $R(m)$ 次高的分辨率的级别 n 以及具有比 $R(m)$ 次低的分辨率的级别 n 的图像块,其中这样的图像块已经被重新采样至分辨率 $R(m)$;

所述聚合器被配置为接收所述 M 个级别的互连图像块以及生成 L 个级别的输出图像块,其中 $L<M$,其中,不同级别 l 的输出图像块具有不同的分辨率 $R(l)$,每个 $R(l)$ 对应于 $R(m)$ 中的一个,并且级别 l 的输出图像块通过对图像块的输入分组 l 执行卷积和非线性变换来生成,图像块的输入分组 l 包含(i)来自 $R(m)=R(l)$ 的级别 m 的互连图像块,以及(ii)来自具有比 $R(l)$ 次低的分辨率的级别 m 的输出图像块,其中这样的输出图像块已经被重新采样至分辨率 $R(l)$;以及

所述聚合器基于最高分辨率的输出图像块来定位所述组成部分。

19. 一种将图像解析成组成部分的方法,该方法包含:

从图像生成 N 个级别的图像块,其中 $N>2$,其中,不同级别 n 的图像块具有不同的分辨率 $R(n)$,并且级别 n 的图像块从被重新采样至分辨率 $R(n)$ 的所述图像生成;

从所述N个级别的图像块生成M个级别的互连图像块,其中 $2 < M \leq N$,其中,不同级别m的互连图像块具有不同的分辨率 $R(m)$,并且级别m的互连图像块从图像块的输入分组m生成,所述图像块的输入分组m包含(i)来自 $R(n) = R(m)$ 的级别n的图像块,以及(ii)来自 $R(n) \neq R(m)$ 的一个或多个级别n的图像块,其中这样的图像块已经被重新采样至分辨率 $R(m)$;以及通过组合所述M个级别的互连图像块来定位组成部分,

其中,所述通过组合所述M个级别的互连图像块包括接收所述M个级别的互连图像块以及生成L个级别的输出图像块,其中 $L < M$,其中,不同级别1的输出图像块具有不同的分辨率 $R(1)$,每个 $R(1)$ 对应于 $R(m)$ 中的一个,并且级别1的输出图像块从图像块的输入分组1生成,图像块的输入分组1包含(i)来自 $R(m) = R(1)$ 的级别m的互连图像块,以及(ii)来自具有小于 $R(1)$ 的分辨率的一个或多个级别的输出图像块,其中这样的输出图像块已经被重新采样至分辨率 $R(1)$ 。

20. 一种非临时性计算机可读介质,被配置为存储程序代码,该程序代码包含用于将图像解析成组成部分的指令,所述指令在被处理器执行时使处理器:

从图像生成N个级别的图像块,其中 $N > 2$,其中,不同级别n的图像块具有不同的分辨率 $R(n)$,并且级别n的图像块从被重新采样至分辨率 $R(n)$ 的图像生成;

从所述N个级别的图像块生成M个级别的互连图像块,其中 $2 < M \leq N$,其中,不同级别m的互连图像块具有不同的分辨率 $R(m)$,并且级别m的互连图像块从图像块的输入分组m生成,所述图像块的输入分组m包含(i)来自 $R(n) = R(m)$ 的级别n的图像块,以及(ii)来自 $R(n) \neq R(m)$ 的一个或多个级别n的图像块,其中这样的图像块已经被重新采样至分辨率 $R(m)$;以及通过组合所述M个级别的互连图像块来定位组成部分,

其中,所述通过组合所述M个级别的互连图像块包括接收所述M个级别的互连图像块以及生成L个级别的输出图像块,其中 $L < M$,其中,不同级别1的输出图像块具有不同的分辨率 $R(1)$,每个 $R(1)$ 对应于 $R(m)$ 中的一个,并且级别1的输出图像块从图像块的输入分组1生成,图像块的输入分组1包含(i)来自 $R(m) = R(1)$ 的级别m的互连图像块,以及(ii)来自具有小于 $R(1)$ 的分辨率的一个或多个级别的输出图像块,其中这样的输出图像块已经被重新采样至分辨率 $R(1)$ 。

用于图像解析的分层互连多尺度卷积网络

技术领域

[0001] 本发明一般地涉及图像处理,更具体地,涉及将图像解析成组成部分。

背景技术

[0002] 图像解析是将图像标注成组成部分(或部位)的列表的任务。图像解析是计算机视觉方面的基本问题。关于图像解析,通常采用关键点提取(或者加关键点)或分割。关键点提取被用于定位组成部分,而分割被用于获得组成部分的印迹(footprint)或形状。面部图像的图像解析具有挑战性,因为从关键点提取获得的关键点可能不是界限分明的,并且可能难以对关键点(例如鼻梁)中的不确定性进行编码。另外,因为面部部位是可变形的,所以图像解析是具有挑战性的。

[0003] 关键点提取和分割的结果取决于信息的尺度。以较低(亦即,较精细)尺度的局部外观提供组成部分的精确轮廓,而以较高(亦即,较粗糙)尺度的全局特征提供组成部分的可能的外观或者所在位置。因此,多尺度信息的结合使得能够对图像进行准确的解析。

[0004] 因此,需要良好的方法以利用以不同尺度存在的信息来有效地执行图像解析。

发明内容

[0005] 本发明通过采用执行图像解析的框架来克服现有技术的限制。在一种方法中,该框架基于用于定位和/或分割图像的组成部分的分层互连多尺度卷积神经网络(HIM)的结构。HIM使得能够在任何方向上整合不同尺度(亦即,分辨率)的数据,以便更好地使用每个尺度的数据。

[0006] 一个方面涉及对图像执行图像解析的系统。在一个示例中,该系统通过解析面部图像在面部识别中使用。该系统包括用于从图像中定位和/或生成组成部分的印迹的HIM。HIM有效地整合不同尺度(亦即,分辨率)的信息。

[0007] 在一种架构中,HIM包括级别生成器、互连组合器和聚合器。级别生成器接收图像并且从图像生成N个级别的图像块。优选地, $N > 2$,并且不同级别n的图像块具有不同的分辨率 $R(n)$ 。级别n的图像块从被重新采样至分辨率 $R(n)$ 的图像生成。在示例性实现方式中,级别生成器为具有较低分辨率 $R(n)$ 的级别n生成更多的图像块。

[0008] 互连组合器接收来自级别生成器的N个级别的图像块,并且从N个级别的图像块生成M个级别的互连图像块,其中优选地, $2 < M \leq N$ 。不同级别m的互连图像块具有不同的分辨率 $R(m)$,并且级别m的互连图像块从图像块的输入分组m生成。图像块的输入分组m包括:(i)来自 $R(n) = R(m)$ 的级别n的图像块;以及(ii)来自 $R(n) \neq R(m)$ 的一个或多个级别n的图像块,其中这样的图像块已经被重新采样至分辨率 $R(m)$ 。

[0009] 聚合器通过组合M个级别的互连图像块来定位组成部分。在示例性实现方式中,聚合器接收M个级别的互连图像块并且生成L个级别的输出图像块,其中 $L < M$ 。不同级别l的输出图像块具有不同的分辨率 $R(l)$,其中每个 $R(l)$ 对应于 $R(m)$ 中的一个。级别l的输出图像块从图像块的输入分组1生成,图像块的输入分组1包含:(i)来自 $R(m) = R(l)$ 的级别m的互连图

像块;以及(ii)来自具有小于 $R(1)$ 的分辨率的一个或多个级别的输出图像块,其中这样的输出图像块已经被重新采样至分辨率 $R(1)$ 。

[0010] 另一方面涉及被配置为接收由HIM产生的组成部分位置的分层互连多尺度卷积神经网络的第二级。第二级中的每个分层互连多尺度卷积神经网络被配置为获得所述组成部分中的一个的经精确化的位置和/或印迹。

[0011] 其他方面包括与前述有关的组成部分、设备、系统、改进、方法、处理、应用和其他技术。

附图说明

[0012] 本发明具有其他优点和特征,根据下面结合附图进行的对本发明的详细描述以及所附的权利要求书,这些其他优点和特征将更加显而易见,附图中:

[0013] 图1是面部识别系统的简化图;

[0014] 图2是面部解析模块的图;

[0015] 图3是由面部解析模块解析输入图像的示例性处理的图;

[0016] 图4是分层互连多尺度卷积神经网络(HIM)的示例性架构的图;

[0017] 图5是解析面部图像的流程图;

[0018] 图6是识别面部部位的位置或印迹的方法的流程图;

[0019] 图7例示根据一个实施例的用于比较的原始图像和具有所识别的面部部位的经处理的图像;

[0020] 图8例示具有通过各种方法获得的所识别的面部部位的图像;

[0021] 图9例示具有通过使用常规加关键点的方法获得的所识别的面部部位的面部图像;

[0022] 图10例示具有通过面部解析模块获得的所识别的面部部位的面部图像。

[0023] 附图图示本发明的实施例,其仅用于例示的目的。本领域的技术人员从下面的讨论中将容易认识到,可以采用在本文中所例示的结构和方法的替代实施例,而不脱离在本文中所描述的本发明的原理。

具体实施方式

[0024] 附图和下面的描述仅涉及作为例示的优选实施例。应当注意,根据下面的讨论,在本文中所公开的结构和方法的替代实施例将容易被认为是可以被采用而不背离所要求保护的原理的可行的替代。

[0025] 所公开的面部识别系统(以及方法)包括面部解析。在一种方法中,面部解析基于分层互连多尺度卷积神经网络(HIM)以识别面部图像的组成部分的位置和/或印迹。HIM从面部图像的不同分辨率图像中生成多个级别的图像块,其中不同级别的图像块具有不同的分辨率。而且,HIM合并不同级别的图像块以从具有不同分辨率的不同级别中生成互连图像块。而且,HIM将互连图像块组合在一起以识别组成部分的位置和/或印迹。

[0026] 面部识别概述

[0027] 图1是根据一个实施例的面部识别系统100的简化图。面部识别系统100接收输入图像110并且尝试识别面部(亦即,将面部与特定个体相关联)。面部识别系统100从输入图

像110中识别组成部分(例如,面部部位),包括但不限于眼睛、眉毛、鼻子和嘴。面部识别系统100基于所识别的组成部分来执行分析以生成度量150。度量150可以指示输入图像110中的所识别的组成部分或者个体的特征。

[0028] 面部识别系统100包括面部解析模块120、分析模块130和处理器140。每个模块可以实施为硬件、软件、固件或者前者的组合。这些模块一起执行面部解析以分析输入图像110中的对象。

[0029] 面部解析模块120接收输入图像110作为输入,并且识别输入图像110的组成部分的位置和/或印迹。在一种方法中,面部解析模块120实现分层互连多尺度卷积神经网络(HIM)。HIM从输入图像110获得不同分辨率(亦即,尺度)的图像块,并且通过使用HIM中的神经网络来合并图像块。另外,面部解析模块120生成所识别的组成部分的位置和/或印迹的指示。

[0030] 分析模块130对来自面部识别模块120的所识别的组成部分执行分析。在一个方面中,分析模块130确定输入图像110中的对象与参考面部图像(未示出)中的对象是否匹配。分析模块130可以获得所识别的组成部分的特征,并且比较所识别的组成部分与参考面部图像中的对应的组成部分。面部识别系统100可以具有针对不同个体的参考面部图像的大型数据库,并且可以比较输入图像110与许多不同的参考面部图像,以识别关于输入图像110的个体。另外,分析模块130基于输入图像110的表示,来生成关于输入图像110是否属于同一对象(人)的度量150。

[0031] 处理器140执行指令以在面部识别系统100上执行面部识别。处理器140从存储器(未示出)或者外部电路系统接收指令。指令也可以完全或者至少部分地驻留在处理器140内(例如在处理器的高速缓冲存储器内)。处理器140根据指令来变换或选择输入图像110的一部分,以提供给面部解析模块120。另外,处理器140根据指令来操作面部解析模块120和分析模块130以执行面部识别的功能。

[0032] 面部解析

[0033] 图2是示例性面部解析模块120的图。面部解析模块120包括一个或多个HIM。面部解析模块120训练每个HIM中的神经网络,以识别输入图像110中的组成部分的位置和/或印迹。每个HIM聚合不同分辨率的信息,以有效地利用组成部分的局部外观和全局特征。

[0034] 如图2所示,在一种实现方式中,面部解析模块120包括第一图像提取器210、估计HIM 220、第二图像提取器230、精确化HIM 240和输出生成器250。在一个实施例中,面部解析模块120包括两级HIM:估计HIM 220和精确化HIM 240。估计HIM 220获得输入图像110中的组成部分的位置的估计,而精确化HIM 240进一步精确化组成部分的所识别的位置。在其他实施例中,面部解析模块120可以仅包括估计HIM 220和精确化HIM 240中的一个,或者可以包括串行或并行配置的多于两级的HIM。

[0035] 第一图像提取器210接收输入图像110,并且生成面部图像215。优选地,第一图像提取器210从输入图像中提取64x64像素的面部图像215,但是可以提取其他大小(例如,80x80像素)的面部图像215。

[0036] 估计HIM 220接收面部图像215并且识别面部部位的位置。在一种方法中,估计HIM 220识别面部部位的位置,包括眉毛、眼睛、鼻子、上嘴唇、嘴内(inner mouth)、下嘴唇和背景,并且生成对应的二进制标签图(label map) 225以指示对应的组成部分的位置。优选地,

估计HIM接收64x64x3像素(其中,x3代表三个色彩通道)的面部图像215。在其他实施例中,估计HIM 220可以识别或多或少数量的组成部分。例如,估计HIM 220可以忽略背景,并且替代地生成八个二进制标签图。

[0037] 第二图像提取器230接收面部部位的位置并且生成对应的部分面部图像。第二图像提取器230接收来自估计HIM 220的二进制标签图225中的面部部位的位置,以便提取部分面部图像235A-E和235M。在一种实现方式中,第二图像提取器230基于其对应的二进制标签图225来生成分别包括左眼、右眼、左眉、右眉和鼻子的部分面部图像235A-E中的每一个。另外,第二图像提取器230基于上嘴唇、嘴内和下嘴唇的二进制标签图225来生成包括嘴(包括上嘴唇、嘴内和下嘴唇)的部分面部图像235M。在一种方法中,第二图像提取器230生成64x64像素的部分面部图像235A-E以及80x80像素的部分面部图像235D。第二图像提取器230可以基于面部部位的位置来生成任何大小的、任何数量的部分面部图像235。在一种实现方式中,第一图像提取器210可以被用于执行第二图像提取器230的功能。

[0038] 精确化HIM 240接收部分面部图像235A并且识别面部部位的精确化的位置。在一种方法中,精确化HIM 240识别包括眉毛、眼睛、鼻子、上嘴唇、嘴内、下嘴唇和背景的面部部位的经精确化的位置,并且生成对应的精确化的二进制标签图245A-E和245M以指示对应的组成部分的位置。每个精确化的二进制标签图245被用于指示面部部位的精确化的位置和/或印迹。面部部位的精确化的位置和/或印迹可以由分析模块130使用以便对面部部位进行分析。

[0039] 在一种实现方式中,精确化HIM 240包括四个另外的HIM 242A-D作为用于精确化面部部位的位置的第二级。另外的HIM 242A接收包括眼睛(例如左眼)的部分面部图像235A,并且生成精确化的二进制标签图245A以指示眼睛的精确化的位置。因为眼睛基本上是对称的,所以另外的HIM 242A还接收包括另一只眼睛(例如右眼)的部分面部图像235B,并且生成精确化的二进制标签图245B以指示另一只眼睛的经精确化的位置。在一种方法中,另外的HIM 242A翻转部分面部图像235B以便定位另一只眼睛,并且翻转结果以生成精确化的二进制标签图245B。另外的HIM 242B基于部分面部图像235C和235D,类似地识别眉毛的精确化的位置,并且生成每个眉毛的精确化的二进制标签图245C和245D。另外,另外的HIM 242C接收包括鼻子的部分面部图像235E,并且生成精确化的二进制标签图245E以指示鼻子的精确化的位置。而且,另外的HIM 242D接收包括嘴的部分面部图像235M,并且生成精确化的二进制标签图245M以指示包括上嘴唇、嘴内和下嘴唇的嘴的精确化的位置。在其他实施例中,精确化HIM 240可以包括任何数量的另外的HIM。

[0040] 输出生成器250接收面部部位的精确化的位置,并且生成经解析的输出图像255。在一个方面中,输出生成器250在输入图像110上重叠面部组成部分的位置。经解析的输出图像255可以被分析模块130用于对输入图像110进行分析或者被提供给用户以便显示。

[0041] 图3是使用面部解析模块120解析面部图像215的示例性过程的图。在该示例中,估计HIM 220接收64x64像素的面部图像215,并且生成二进制标签图225A-I以识别分别包括左眼、右眼、左眉、右眉、鼻子、上嘴唇、嘴内、下嘴唇和背景的面部部位的位置。

[0042] 基于面部部位的所识别的位置,第二图像提取器230提取面部部位的部分面部图像235。在该示例中,第二图像提取器230分别提取左眉、右眉、左眼、右眼和鼻子的64x64像素的部分面部图像235A-E。另外,第二图像提取器230提取包括上嘴唇、嘴内和下嘴唇的嘴

的80x80像素的部分面部图像235M。

[0043] 精确化HIM 240接收部分面部图像235,并且生成精确化的二进制标签图245以分别获得左眉、右眉、左眼、右眼、鼻子和嘴的精确化的位置和/或印迹。嘴的精确化的二进制标签图245M可以包括上嘴唇、嘴内和下嘴唇的二进制标签图。

[0044] 输出生成器250接收精确化的二进制标签图245,并且生成指示面部图像215的所识别的面部部位的位置和/或印迹的经解析的输出图像255。

[0045] 面部解析

[0046] 图4是估计HIM 220的示例性架构的图。HIM 220接收图像作为输入,并且生成目标区域的一个或多个标签图。在一个实施例中,HIM 220包括级别生成器410、互连组合器450和聚合器480。在一种实现方式中,级别生成器410从面部图像215生成N个级别的图像块。互连组合器450整合不同级别的图像块,以生成M个级别的互连图像块。聚合器480组合互连图像块,以生成L个级别的输出图像块。优选地, $M=N$ 并且 $L=M-1$ 。第二级中的另外的HIM 242可以具有与HIM 220相同或相似的架构。取决于被监督的信号,聚合器480可以生成不同数量的输出或标签图。

[0047] 级别生成器410接收面部图像215,并且从面部图像215生成N个级别的图像块,其中优选地, $N>2$ 。每个神经网络执行向下采样、卷积和非线性变换(例如,双曲正切)。不同的CNN级别通常具有不同的深度和输入大小。在一种实现方式中,级别 $n=1$ 具有与面部图像215的分辨率相同的分辨率 $R(n=1)$,并且每个其他级别 n 具有相对于具有分辨率 $R(n-1)$ 的前一个级别 $n-1$ 按照2或更小的常量因子向下采样的分辨率 $R(n)$ 。级别生成器410生成具有较低分辨率 $R(n)$ 的级别 n 的更多的图像块。不同级别 n 的图像块具有不同的分辨率 $R(n)$,并且级别 n 的图像块通过对被重新采样至分辨率 $R(n)$ 的图像执行卷积和非线性变化来生成。

[0048] 在图4中的示例中, $N=4$,并且不同的CNN级别被标注为“级别1”至“级别4”。在该示例中,64x64或者80x80像素的面部图像215被向下采样成分别具有1/2、1/4和1/8大小的图像402、404和406。另外,对每个级别应用不同的卷积滤波器和双曲正切激活函数。级别1的卷积的输入可以是RGB的。级别生成器410对于级别1以原始尺度生成8个图像块412,对于级别2以1/2尺度生成16个图像块414,对于级别3以1/4尺度生成24个图像块416,而对于级别4以1/8尺度生成32个图像块418。

[0049] 互连组合器450接收来自级别生成器410的N个级别的图像块,并且从N个级别的图像块生成M个级别的互连图像块,其中优选地, $2<M\leq N$ 。不同级别 m 的互连图像块具有不同的分辨率 $R(m)$,并且级别 m 的互连图像块从图像块的输入分组 m 生成,图像块的输入分组 m 包含:(i)来自 $R(n)=R(m)$ 的级别 n 的图像块;以及(ii)来自 $R(n)\neq R(m)$ 的一个或多个级别 n 的图像块,其中这样的图像块已经被重新采样至分辨率 $R(m)$ 。优选地,级别 m 的互连图像块来自图像块的输入分组 m ,包括:(i)来自 $R(n)=R(m)$ 的级别 n 的图像块;(ii)来自具有比 $R(m)$ 次高的分辨率的级别 n 以及具有比 $R(m)$ 次低的分辨率的级别 n 的图像块(如果存在的话)。互连组合器450通过对图像块的输入分组 m 执行卷积和非线性变换来生成级别 m 的互连图像块。

[0050] 在 $N=M$ 的一种实现方式中,互连组合器450生成级别 m 的互连图像块,以具有 $R(n)=R(m)$ 情况下的级别 n 的相同数量的图像块。在该实施例中,可以级联地使用多个互连组合器450,或者可以递归地多次使用互连组合器450。

[0051] 在图4中的示例中, $N=M=4$, 并且互连组合器450基于图像块412、414、416和418来生成4个级别的互连图像块452、454、456和458。对于具有 $R(n) = R(m)$ 的级别 m , 互连组合器450通过整合来自级别 n 和级别 n 的相邻级别的图像块作为图像块的输入分组 m 以及对图像块的输入分组 m 执行卷积和非线性变换, 来生成来自级别 n 的相同数量的图像块。为了整合具有不同分辨率的图像块, 互连组合器450将来自级别 $R(n) \neq R(m)$ 的图像块重新采样至分辨率 $R(m)$ 。

[0052] 例如, 为了生成级别1的互连图像块452, 互连组合器450将来自级别1的8个图像块412和来自级别2的16个图像块414组合(或者堆叠)成图像块的输入分组1(451), 以具有24个图像块。为了匹配分辨率, 互连组合器450将来自级别2的16个图像块414向上采样至分辨率 $R(m=1)$ 。另外, 互连组合器450对图像块的输入分组1(451)执行卷积和非线性变换, 以生成级别1的8个互连图像块452。

[0053] 类似地, 为了生成级别2的互连图像块454, 互连组合器450将来自级别1的8个图像块412、来自级别2的16个图像块414以及来自级别3的24个图像块组合(或者堆叠)成图像块的输入分组2(453), 以具有48个图像块。为了匹配分辨率, 互连组合器450将来自级别3的24个图像块416向上采样至分辨率 $R(m=2)$, 并且将来自级别1的8个图像块412向下采样至分辨率 $R(m=2)$ 。另外, 互连组合器450对图像块的输入分组2(453)执行卷积和非线性变换, 以生成级别2的16个互连图像块454。

[0054] 为了生成级别3的互连图像块456, 互连组合器450将来自级别2的16个图像块414、来自级别3的24个图像块以及来自级别4的32个图像块418组合(或者堆叠)成图像块的输入分组3(455), 以具有72个图像块。为了匹配分辨率, 互连组合器450将来自级别4的32个图像块418向上采样至分辨率 $R(m=3)$, 并且将来自级别2的16个图像块414向下采样至分辨率 $R(m=3)$ 。另外, 互连组合器450对图像块的输入分组3(455)执行卷积和非线性变换, 以生成级别3的24个互连图像块456。

[0055] 为了生成级别4的互连图像块458, 互连组合器450将来自级别3的24个图像块416以及来自级别4的32个图像块418组合(或者堆叠)成图像块的输入分组4(457), 以具有56个图像块。为了匹配分辨率, 互连组合器450将来自级别3的24个图像块416向下采样至分辨率 $R(m=4)$ 。另外, 互连组合器450对图像块的输入分组4(457)执行卷积和非线性变换, 以生成级别4的32个互连图像块458。

[0056] 聚合器480接收来自互连组合器450的 M 个级别的互连图像块, 并且生成 L 个级别的输出图像块, 其中 $L < M$ 。不同级别1的输出图像块具有不同的分辨率 $R(1)$ 。每个 $R(1)$ 对应于 $R(m)$ 中的一个, 并且级别1的输出图像块486通过对图像块的输入分组1执行卷积和非线性变换来生成, 图像块的输入分组1包括: (i) 来自 $R(m) = R(1)$ 的级别 m 的互连图像块; 以及 (ii) 来自具有比 $R(1)$ 次低的分辨率的级别 m 的输出图像块, 其中, 这样的输出图像块已经被重新采样至分辨率 $R(1)$ 。另外, 聚合器480基于最高分辨率的输出图像块486来生成一个或多个标签图488, 以定位组成部分。

[0057] 在图4中的示例中, $N=M=4$ $L=3$, 并且聚合器480顺序地组合(或者堆叠)来自不同级别的互连图像块。首先, 聚合器480对来自级别4的32个互连图像块458进行向上采样, 并且将它们与来自级别3的24个互连图像块456组合(或者堆叠)在一起, 以形成56个图像块的输入分组3(481)。聚合器480对图像块的输入分组3(481)执行卷积和非线性变换。以生成级

别3的24个输出图像块482。

[0058] 另外,聚合器480对来自级别3的24个输出图像块482进行向上采样,并且将它们与来自级别2的16个互连图像块454组合(或者堆叠)在一起,以形成40个图像块的输入分组2(483)。聚合器480对图像块的输入分组2(483)执行卷积和非线性变换,以生成级别2的16个输出图像块484。

[0059] 类似地,聚合器480对来自级别2的16个输出图像块484进行向上采样,并且将它们与来自级别1的8个互连图像块452组合(或者堆叠)在一起,以形成24个图像块的输入分组1(485)。聚合器480对图像块的输入分组1(485)执行卷积和非线性变换,以生成级别1的输出图像块486。在一个方面中,输出图像块486包括8个输出图像块以及另外的输出图像块,优选为标签图的数量两倍(例如,2*标签图+8)。聚合器480还对来自级别1的输出图像块486执行另外的卷积,以生成二进制标签图488。

[0060] HIM 220的一个方面涉及执行卷积和非线性变换。执行卷积和非线性变换允许深度神经网络拟合来自输入的块的非线性函数与深度结构。将具有 L_1 个通道的2d输入数据标记为 $In_{x,y}^k$, 卷积滤波器为 $W_{k,i,j}^L$, 2d输出, 关于每个输出层的偏置 B^L , 由HIM 220执行的卷积和非线性变换可以表征如下:

$$[0061] \quad Out_{x,y}^L = \tanh \left[\left(\sum_{i=0}^{Rw} \sum_{j=0}^{Rw} \sum_{k=0}^{Rw} In_{x+i \times pad, y+j \times pad}^k W_{k,j}^L \right) + B^L \right] \quad (1)$$

[0062] 关于被用于获得二进制标签图488的线性卷积,可以省略双曲正切运算符。在一种实现方式中,优选相同大小的输入和输出,因此,输入的可见范围的外部用一圈零来填充(对于可见范围外部的 x 和 y , $In_{x,y}^k = 0$)。填充位移 pad 可以被设置成 $(Rw-1)/2$, 使得输出的各个字段以输入的中心位置为中心。可以在卷积之后、在双曲正切之前应用偏置参数 B^L 。

[0063] HIM 220的另一个方面涉及执行向下采样,以使得神经元能够具有更大的相应字段。在一种方法中,最大池化(max pooling)技术用于向下采样。给定 $N \times N$ 输入 $In_{x,y}^k$, 最大池化取得来自每个 $M \times M$ 子补片的最大值作为其输出。能够设置步幅 K , 使得子补片可以具有重叠的空间局部性。

$$[0064] \quad Out_{x,y}^L = \max_{i,j=0}^M In_{x \times k+i, y \times k+j}^L \quad (2)$$

[0065] 然而,HIM 220的另一个方面涉及执行向上采样。在一种方法中,向上采样通过添加冗余性来执行。向上采样以较粗糙的尺度来填充响应图(response map), 以使其适合较精细的尺度。例如,2x2向上采样可以如下定义:

$$Out_{2x+1,2y+1}^L = Out_{2x+1,2y+2}^L = Out_{2x+2,2y+1}^L = Out_{2x+2,2y+2}^L = In_{x,y}^L \quad (3A)$$

$$[0066] \quad Out_{x,y}^L = In_{\frac{x}{2}, \frac{y}{2}}^L \quad (3B)$$

[0067] 另外,HIM 220的另一个方面涉及直接生成逐像素(pixel-wise)的标签图。在一种方法中,作为损失函数的softmax和最大对数似然可以被用于如下那样地生成标签图:

$$Soft \max_{x,y}^L = \frac{e^{In_{x,y}^L}}{\sum_{i=1}^{L_{in}} e^{In_{x,y}^i}} \quad (4)$$

[0068]

$$Loss = \sum_{i,j=1}^{i \leq W, j \leq H} \sum_{k=1}^{L_{out}} \log(Soft \max_{i,j}^k) L_{i,j}^k \quad (5)$$

[0069] 其中, L_{xy} 是监督信号, 其标注仅 $1 \cdots L_{out}$ 中的 L 的正确类别具有值 1 而其他具有值 0。这里, W 和 H 分别表示输出标签的宽度和高度。优选地, 聚合器 480 将应用 softmax, 并且使用对数似然来计算损失函数。

[0070] 面部解析的运算

[0071] 图 5 是根据一个实施例的解析面部图像的方法的流程图。面部解析模块 120 采用一个或多个 HIM 来生成面部图像的代表。面部解析模块 120 使用 HIM 来识别面部部位的位置的估计值 (510)。面部解析模块 120 使用图像提取器 230, 基于部分面部图像的所估计的位置来提取部分面部图像 (520)。面部解析模块 120 使用另外的 HIM 作为眼睛、眉毛、鼻子和嘴的第二级, 基于所提取的部分面部图像, 来精确化面部部位的位置 (530)。面部解析模块 120 组合面部图像与所定位的面部部位 (540), 以指示面部部位的位置和/或印迹。

[0072] 图 6 是识别面部部位的位置和/或印迹的方法的流程图。HIM 从面部图像生成图像块的级别 (610)。HIM 通过组合与不同级别相关联的图像块来生成互连图像块 (620)。不同级别 m 的互连图像块具有不同的分辨率 $R(m)$, 并且级别 m 的互连图像块从图像块的输入分组 m 生成, 图像块的输入分组 m 包含: (i) 来自 $R(n) = R(m)$ 的级别 n 的图像块; 以及 (ii) 来自 $R(n) \neq R(m)$ 的一个或多个级别 n 的图像块, 其中这样的图像块已经被重新采样至分辨率 $R(m)$ 。优选地, 级别 m 的互连图像块来自图像块的输入分组 m , 图像块的输入分组 m 包括: (i) 来自 $R(n) = R(m)$ 的级别 n 的图像块; (ii) 来自具有比 $R(m)$ 次高的分辨率的级别 n 以及具有比 $R(m)$ 次低的分辨率的级别 n 的图像块 (如果存在的话)。HIM 通过组合互连图像块以及生成一个或多个二进制标签图, 来识别面部部位的位置和/或印迹 (630)。

[0073] 模拟结果

[0074] 为了验证面部识别系统 100 和面部解析模块 120 的性能, 测试不同的图像解析方法, 包括最大池化完全连接 (MPFC)、纯卷积双曲正切 (TANH)、不具有互连多尺度的分层 CNN (HNOIM) 以及包括 HIM 220 的面部解析模块 120。

[0075] 为了比较, 应用来自 Helen 数据库的 2330 个面部图像。将 2330 个面部图像分到 3 个分组中, 其中, 2000 个图像用于训练, 230 个用于验证/调谐, 并且 100 个用于测试。因为 Helen 数据库中的注释是基于关键点的, 所以, 为了更好地比较, 将 Helen 数据库中的图像重新调节大小并且修改为在像素级地面实况数据上粗略地对齐。为了比较, 通过使用手动注释的轮廓作为分割边界, 来自动地生成地面实况眼睛、眉毛、鼻子、嘴内、上嘴唇和下嘴唇片段。

[0076] 图 7 例示原始图像和具有所识别的面部部位的地面实况的经处理的图像的示例, 以便比较。来自 Helen 数据库的原始图像 710 在顶行中示出, 而具有所识别的面部部位的地面实况图像 720 在底行中示出。顶行中的图像具有围绕每个组成部分轮廓的点标记。这些点以标记眼睛、眉毛等的预先定义的次序。处理底行中的地面实况图像, 以将这些点柔和地链接, 从而生成逐像素的标签区域。

[0077] 对于训练集和调谐集中的每个图像,提取每个组成部分的块图像。优选地,块图像大小为64x64像素以在数据集中包括眉毛、眼睛和鼻子中的一个,以及80x80像素以包括嘴。

[0078] 为了防止过度拟合并且增强面部解析模块120,应用数据论证(data argumentation)。每当块图像被选择作为输入时,应用随机±15度旋转、随机0.9-1.1x缩放以及每个方向上的随机±10像素位移。

[0079] 图8例示包括由各种方法获得的所识别的面部部位中的一个的图像块,以便比较。第一列包括包含鼻子、眉毛、眼睛和上嘴唇的地面实况的图像块。第二列包括具有使用MPFC识别的组成部分的图像块。第三列包括具有使用TANH识别的组成部分的图像块。第四列包括具有使用HNOIM识别的组成部分的图像块。第五列包括如上所述地使用HIM识别的组成部分的图像块。MPFC、TANH和HNOIM是替代方法。

[0080] 如图8中所示,尽管MPFC良好地捕获全局特征,但是MPFC在阴影、非刚性变换和看不见的大形变方面表现较差。相比之下,TANH和HNOIM更多地关注于局部外观,但是在全局特征的合并方面欠缺。与MPFC模型不同,TANH和HNOIM在响应中的随机位置处具有随机噪声。因此,TANH和HNOIM产生组成部分的较差识别。HIM(如上所述的方法)利用全局特征与局部外观之间的美好平衡,从而实现接近第一列中的地面实况的出色结果。

[0081] 在表格1中列出MPFC、TANH、HNOIM和HIM在图像块上的F1得分。如表格1中所示,在识别眉毛、眼睛、鼻子和上嘴唇方面,HIM胜于MPFC、TANH和HNOIM。在识别嘴内和下嘴唇方面,HNOIM表现稍微好于HIM。然而,对于嘴内唇和下嘴唇,HIM仍然表现良好。

[0082] 表格1. 每个模型在每个组成部分上的F1得分比较

[0083]

模型	眉毛	眼睛	鼻子	上嘴唇	嘴内	下嘴唇
MPFC	0.812	0.875	0.930	0.662	0.786	0.798
TANH	0.723	0.794	0.920	0.636	0.681	0.726
HNOIM	0.717	0.854	0.919	0.729	0.842	0.842
HIM	0.877	0.894	0.961	0.778	0.836	0.833

[0084] 图9例示具有由使用常规加关键点各种方法所获得的所识别的面部部位的面部图像。使用常规加关键点代替估计HIM 220来定位面部部位,以便比较MPFC、TANH、HNOIM和HIM的性能。图像910是由MPFC获得的结果,图像920是由TANH获得的结果,图像930是由HNOIM获得的结果,而图像940是由HIM获得的结果。

[0085] 在表格2中列出使用常规加关键点的MPFC、TANH、HNOIM和HIM在面部图像上的F1得分。如图9中所示,HIM产生令人印象深刻的结果。

[0086] 表格2. 每个模型使用常规加关键点在每个组成部分上的F1得分比较

[0087]

模型	眉毛	眼睛	鼻子	上嘴唇	嘴内	下嘴唇	总体
MPFC	0.73	0.84	0.80	0.67	0.75	0.74	0.772
TANH	0.76	0.69	0.88	0.57	0.55	0.64	0.741
HNOIM	0.70	0.81	0.80	0.49	0.58	0.57	0.708
HIM	0.82	0.88	0.92	0.75	0.83	0.80	0.858

[0088] 图10例示具有由面部解析模块120获得的所识别的面部部位的面部图像。来自

Helen数据库的原始图像1010在顶行中示出,而具有所识别的面部部位的图像1020在底行中示出。

[0089] 如图10中所示,使用HIM的面部解析模块120在捕捉眉毛、鼻梁以及高度可形变的嘴唇的复杂形状时表现良好。面部解析模块120不需要任何预处理,并且在估计HIM 220中自动地获得头部位置和粗略分割。

[0090] 虽然详细描述包含许多细节,但是这些不应当被解释为限制本发明的范围,而是应当被解释为仅仅是例示本发明的不同示例和方面。应当领会到,本发明的范围包括未在上文详细讨论的其他实施例。例如,除了面部识别和表示之外,在HIM 220或者面部解析模块120中所公开的原理还可以应用于其他区域或对象,例如解析其他类型的图像。可以在本文所公开的本发明的方法和装置的布置、操作和细节方面做出对于本领域的那些技术人员将是显然的各种其他修改、改变和变化,而不背离如在所附的权利要求书中所限定的本发明的精神和范围。因此,本发明的范围应当由所附的权利要求书及其法律等同物来确定。

[0091] 在替代的实施例中,本发明实现为计算机硬件、固件、软件和/或它们的组合。本发明的装置能够实现为有形地实施在机器可读的存储设备中以便由可编程处理器执行的计算机程序产品;并且本发明的方法步骤能够由通过操作输入数据并且生成输出来执行指令程序以执行本发明的功能的可编程处理器来执行。本发明能够有利地实现为可在可编程系统上执行的一个或多个计算机程序,可编程系统包括至少一个可编程处理器、至少一个输入设备以及至少一个输出设备,所述至少一个可编程处理器被耦合以从数据存储系统接收数据和指令以及向数据存储系统传送数据和指令。每个计算机程序能够以高级程序或者面向对象的编程语言来实现,或者如果期望的话,能够以汇编或机器语言来实现;并且在任何情况下,语言可以是编译或解释语言。作为示例,适当的处理器包括通用和专用微处理器。一般地,处理器将从只读存储器 and/或随机存取存储器接收指令和数据。一般地,计算机将包括用于存储数据文件的一个或多个海量存储设备;这样的设备包括诸如内部硬盘和可移动盘这样的磁盘、磁光盘以及光盘。适合于有形地实施计算机程序指令和数据的存储设备包括所有形式的非易失性存储器,例如包括诸如EPROM、EEPROM和闪存设备这样的半导体存储设备、诸如内部硬盘和可移动盘这样的磁盘、磁光盘以及CD-ROM盘。前述中的任何均可由ASIC(专用集成电路)和其他形成的硬件补充或者合并到ASIC以及其他形成的硬件中。

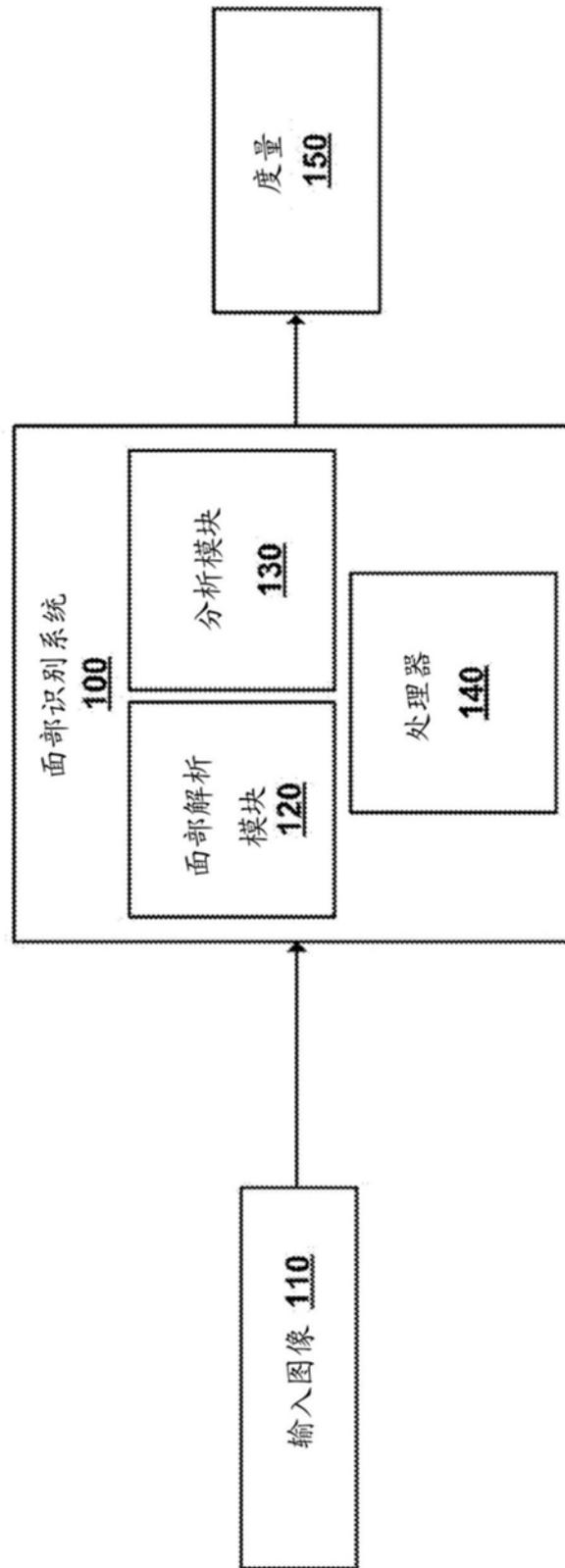


图1

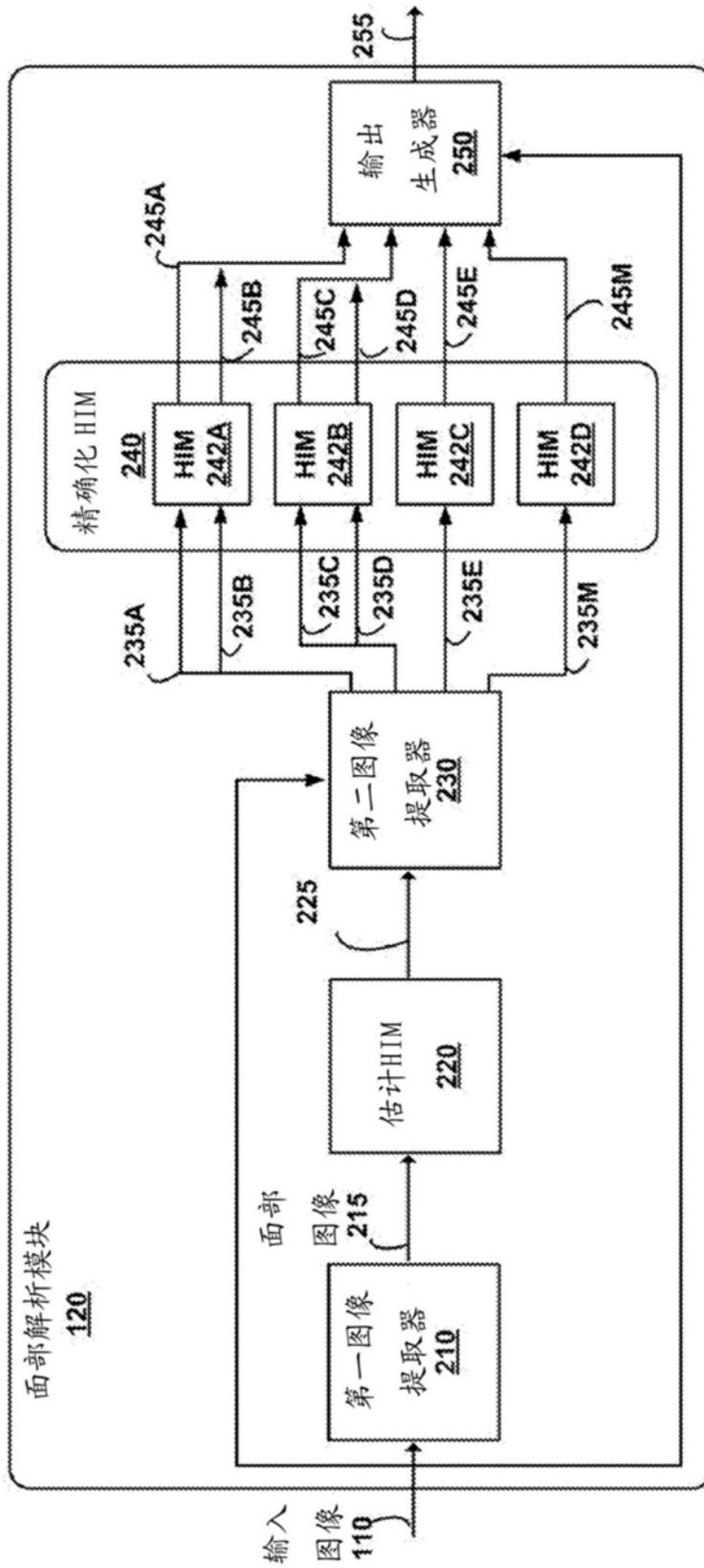


图2

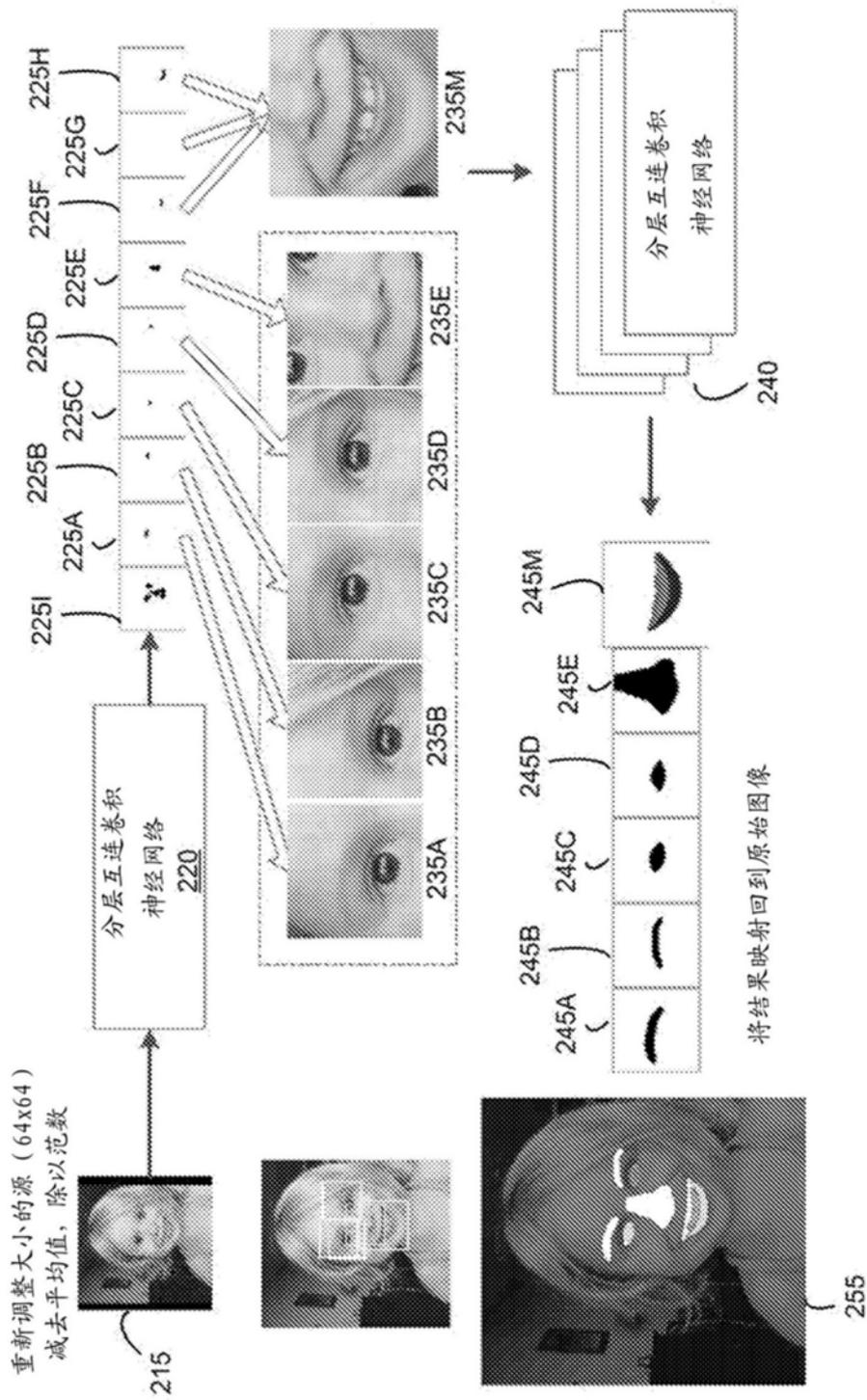


图3

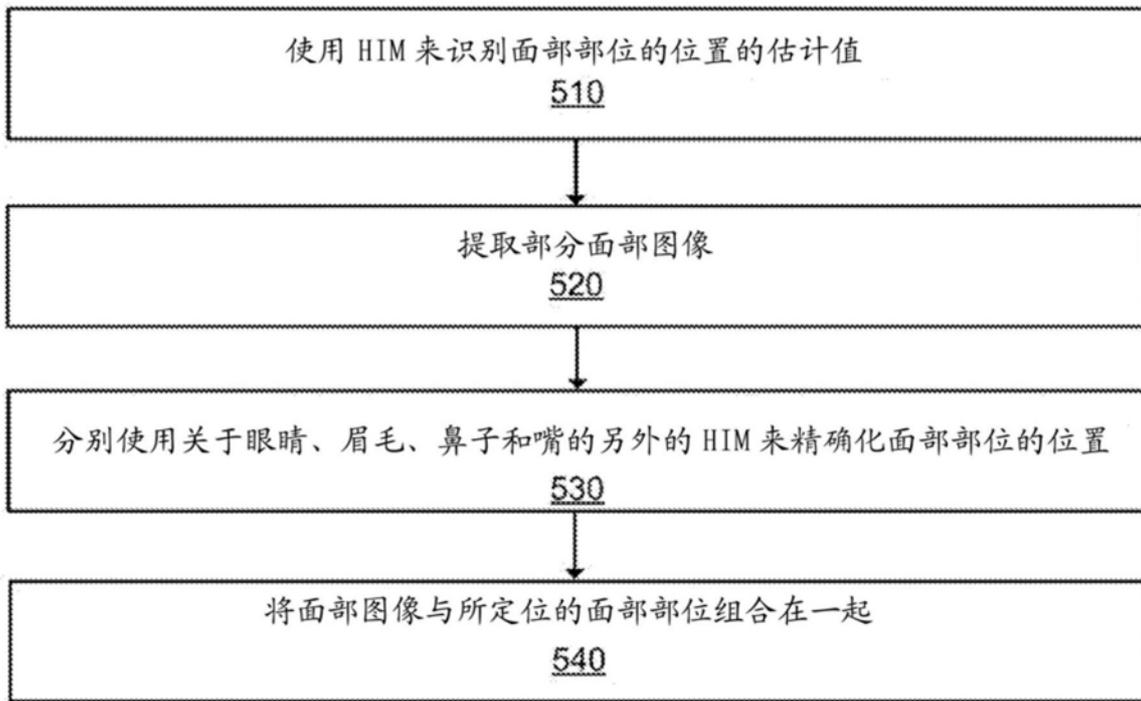


图5

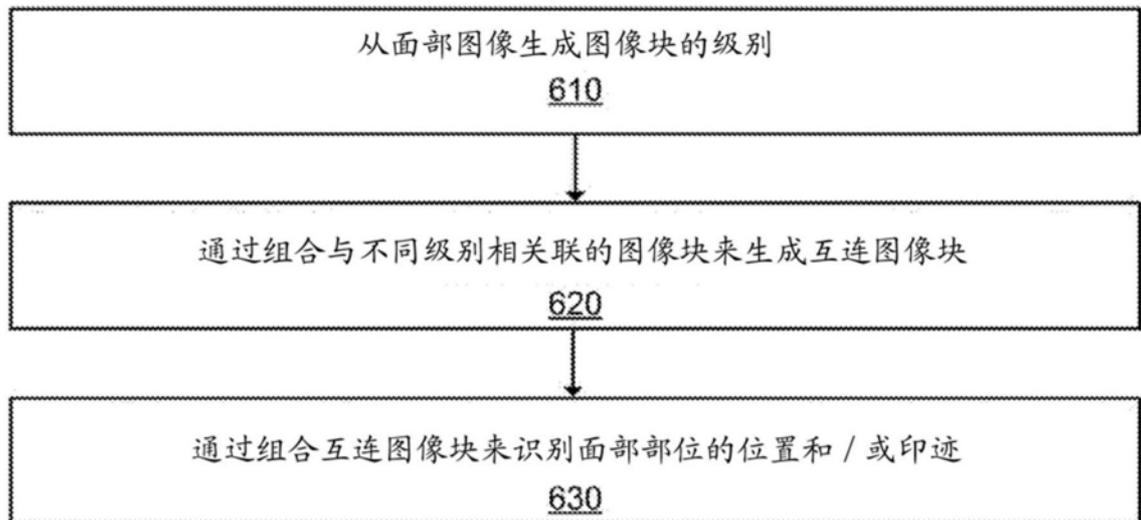


图6

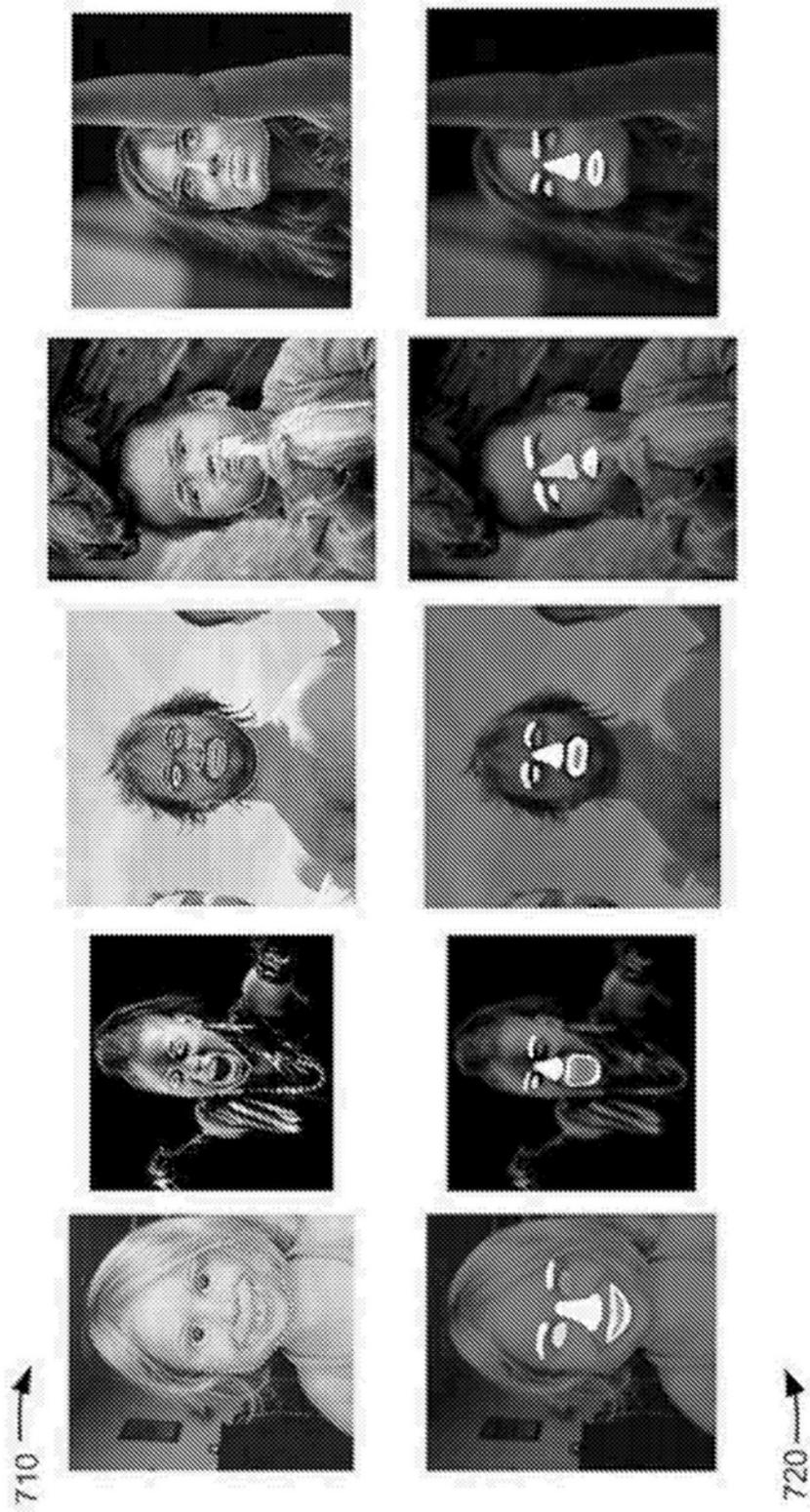


图7

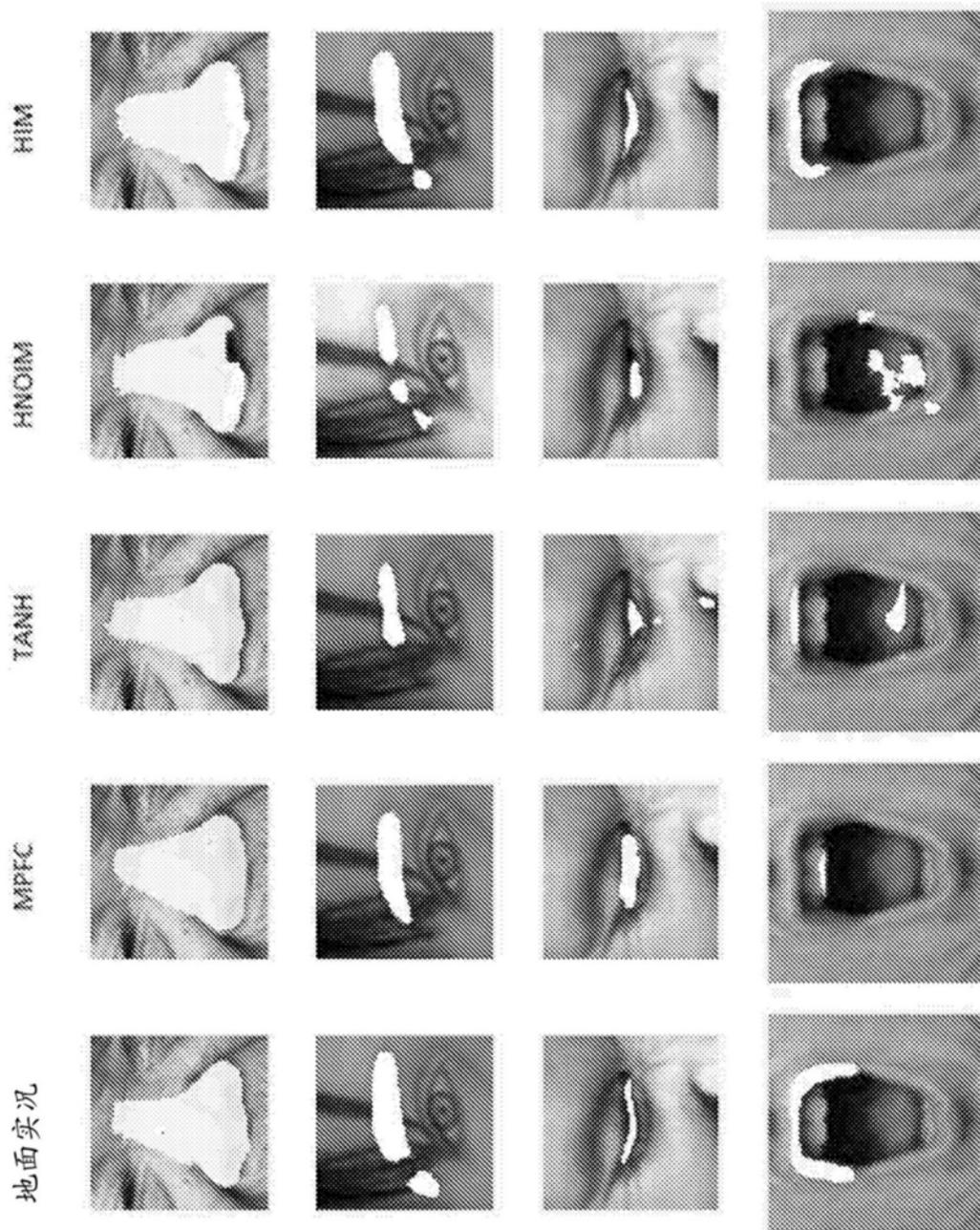


图8

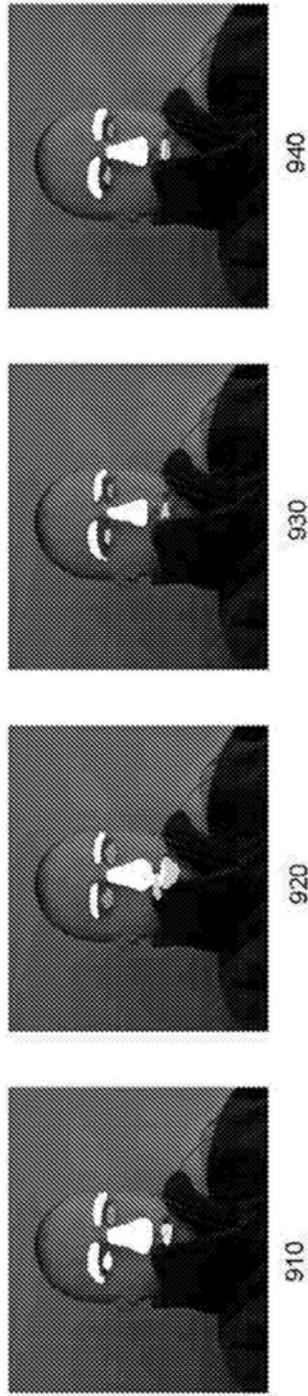


图9

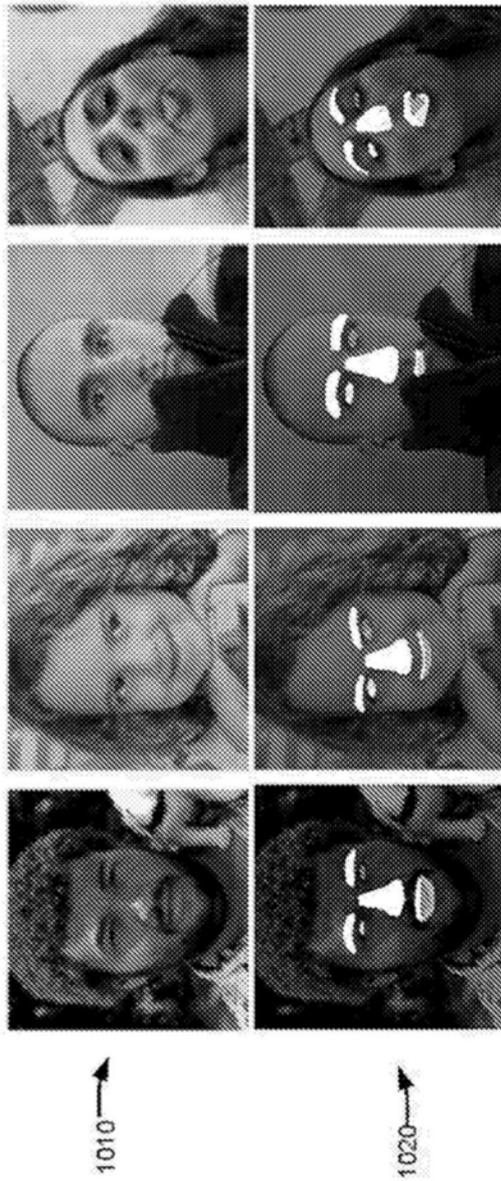


图10