



(12)发明专利申请

(10)申请公布号 CN 107155360 A

(43)申请公布日 2017.09.12

(21)申请号 201580043139.7

(74)专利代理机构 中国专利代理(香港)有限公司 72001

(22)申请日 2015.07.31

代理人 胡莉莉 刘春元

(30)优先权数据

14/457381 2014.08.12 US

(51)Int.Cl.

G06K 9/00(2006.01)

(85)PCT国际申请进入国家阶段日

G06K 9/46(2006.01)

2017.02.10

G06K 9/62(2006.01)

(86)PCT国际申请的申请数据

PCT/US2015/043057 2015.07.31

(87)PCT国际申请的公布数据

W02016/025189 EN 2016.02.18

(71)申请人 西门子保健有限责任公司

地址 德国埃朗根市

(72)发明人 H.阮 V.K.辛格 郑治枫

B.乔治斯库 D.科马尼丘 周少华

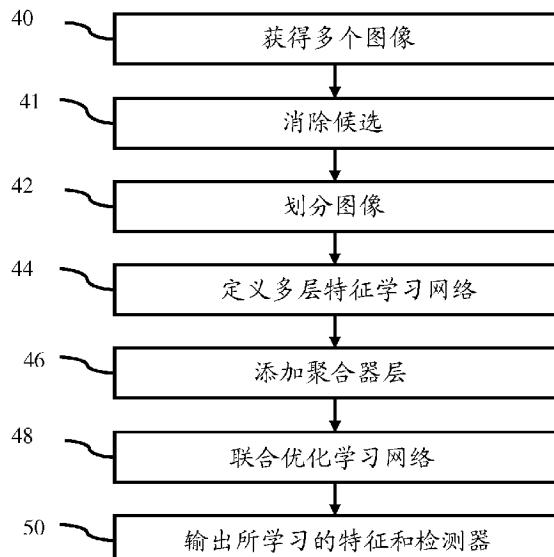
权利要求书2页 说明书11页 附图3页

(54)发明名称

用于对象检测的多层聚合

(57)摘要

对象检测(58)使用深的或多层的网络(72-80)以学习用于检测(58)图像中的对象的特征。聚合(46)来自不同层的多个特征,以训练(48)用于所述对象的分类器。除了来自不同层的特征聚合之外或者作为来自不同层的特征聚合的替换,初始层(72)可以具有用于图像(70)的不同区域的分离的所学习的节点,以减少自由参数的数量。学习对象检测(58)或应用所学习的对象检测器。



1. 一种用于对象检测的方法,所述方法包括:
 获得(40)对象的图像;
 定义(44)多层特征学习网络的多个顺序特征层;
 提供(46)从所述多层特征学习网络的多个层接收特征的聚合器层;
 联合地并且通过处理器使用所述对象的图像来优化(48)所述多层特征学习网络和所述聚合器层;以及
 通过所述处理器输出(50)由被优化的多层特征学习网络表示的所学习的特征的集合以及通过被优化的聚合器层利用所生成的特征的检测器,所学习的特征的集合用于区分所述对象,并且所述检测器用于对所述对象进行分类。
2. 如权利要求1所述的方法,其中,获得(40)所述图像包括:获得(40)具有包括器官或医学异常的对象的医学图像。
3. 如权利要求1所述的方法,其中,获得(40)所述图像包括:获得(40)具有包括人的对象的安全图像。
4. 如权利要求1所述的方法,其中,进行定义(44)包括:将所述多层特征学习网络定义(44)为深架构。
5. 如权利要求1所述的方法,其中,进行定义(44)包括:将所述顺序特征层定义(44)为自动编码器层。
6. 如权利要求1所述的方法,其中,进行定义(44)包括:将所述顺序特征层定义(44)为受限制的玻尔兹曼机层。
7. 如权利要求1所述的方法,其中,进行定义(44)包括:将所述多层特征学习网络定义(44)为具有所述顺序特征层中的至少四个,来自所述顺序特征层中的第一层的输出是馈送到所述顺序特征层中的第二层的第一特征,来自所述顺序特征层中的第二层的输出是馈送到所述顺序特征层中的第三层的第二特征,所述顺序特征层中的第三层的输出是馈送到所述顺序特征层中的第四层的第三特征,并且所述顺序特征层中的第四层、第三层以及第二层的输出被馈送到所述聚合器层。
8. 如权利要求1所述的方法,其中,进行优化(48)包括:机器学习。
9. 如权利要求1所述的方法,其中,联合地进行优化(48)包括:使用从所述聚合器层到所述特征层中的多个层的向后投射。
10. 如权利要求1所述的方法,其中,进行优化(48)包括:学习具有概率提升树或支撑矢量机的所述聚合器层的映射函数。
11. 如权利要求1所述的方法,其中,输出(50)所述检测器包括:基于所述多个特征层的特征而输出(50)所述检测器作为多个属性的分类器。
12. 如权利要求1所述的方法,其中,输出(50)所学习的特征包括:基于所述对象的重构而输出(50)分离的特征层的特征作为不同的抽象水平。
13. 如权利要求1所述的方法进一步包括:将所述图像中的每一个划分(42)为子区域,其中,进行定义(44)包括:将所述顺序特征层中的第一层定义(44)到用于相应的子区域的分离的局部特征学习器中,并且将所述顺序特征层中的第二层定义(44)为用于来自所述分离的局部特征学习器的输出的图像的整体的全局特征学习器。
14. 如权利要求1所述的方法进一步包括:消除(41)针对具有预先编程的特征的所述对

象的候选，并且其中，进行优化(48)包括：在所述消除(41)之后使用其余的候选进行优化(48)。

15. 一种非暂态计算机可读存储介质，其中已经存储表示由用于对象检测的编程处理器可执行的指令的数据，所述存储介质包括用于如下的指令：

接收(54)对象的图像；

通过所述处理器检测(58)具有从深网络的隐藏层生成的隐藏特征的所述图像中的所述对象，从所述隐藏层学习的所述隐藏特征包括不同的抽象层，其中，所述检测(58)使用来自不同的层输入的所述隐藏特征的通过机器进行的在机器学习分类器中的聚合作为特征矢量；以及

输出(60)所述对象的检测。

16. 如权利要求15所述的非暂态计算机可读存储介质，其中，接收(54)所述图像包括：接收(54)所述图像作为可视信息，并且其中，进行检测(58)包括：利用包括多层网络的节点输出的深网络的隐藏特征进行检测(58)，其中，所述节点是被训练的自动编码器或受限制的玻尔兹曼机块。

17. 如权利要求15所述的非暂态计算机可读存储介质，其中，进行检测(58)包括：检测(58)所述隐藏特征作为第一层中的第一特征，在所述第一特征作为输入的情况下第二层中的第二属性、以及在所述第二属性作为输入的情况下第三层中的第三属性，所述机器学习分类器的所述特征矢量至少是所述第二属性和第三属性。

18. 如权利要求15所述的非暂态计算机可读存储介质，其中，进行检测(58)包括：在所述隐藏特征的第一集合用于所述图像的相应的子部分的情况下并且在所述隐藏特征的第二集合用于具有来自所有所述第一集合的输入的所述图像的情况下进行检测(58)。

19. 如权利要求15所述的非暂态计算机可读存储介质，其中，进行检测(58)进一步包括：在利用所述深网络的所述隐藏特征进行检测(58)之前，限制具有预先编程的特征的候选。

20. 一种用于对象检测的方法，所述方法包括：

将对象的图像划分(42)为局部部分；

通过机器训练(48)特征节点的第一块，以重构所述图像的所述局部部分；

通过所述机器训练(48)第二属性节点的第二块，以重构所述特征节点；

通过所述机器训练(48)第三属性节点的第三块，以重构所述第二属性节点；以及

通过所述机器训练(48)特征聚合器，以对所述对象进行分类，所述特征聚合器是利用作为输入的所述第二和第三属性节点来训练的。

用于对象检测的多层聚合

背景技术

- [0001] 本实施例涉及对象检测以及对象检测的机器学习。
- [0002] 对于机器学习的对象检测而言,使用来自图像的输入特征以训练并且应用检测器。特征的质量对于很多图像分析任务的性能是关键的。科学家已经利用对手头的数据和任务的深入理解而提出了各种特征。例如,Haar特征归因于它们的廉价计算而被使用在器官检测和分段中。局部二元模式(LBP)特征对于表示形状或纹理而言是良好的,并且适合于人类重新标识。对于所有任务或所有类型的数据而言没有特征是优化的。通常的是要求相当大量的经验以针对特定的应用选择良好的特征。
- [0003] 感测技术的演进或者现有的感测技术上的变化偶尔使新的特征的设计成为必须。该处理通常是有挑战性的,因为数据生成的潜在物理学可能不容易理解。字典学习和稀疏编码算法已经被用于学习特征。例如,字典学习和稀疏编码算法当被应用于自然图像时学习类似Gabor的模式。然而,归因于字典原子的高相干性,该方法可能产生不稳定的特征。换言之,如果存在看起来相似的多个字典原子,则稀疏码可能在任何的这些原子之间跳转。虽然这对于图像重构或图像降噪应用而言并不是问题,但是特征的不稳定性对于学习良好的分类器或检测器而言可能提出严重挑战。
- [0004] 用以利用机器学习特征的另一方法使用自动编码器(AE)或受限制的玻尔兹曼机(RBM)。由于解码步骤仅牵涉矩阵乘法,因此产生稳定的特征。此外,AE或RBM通常是被堆栈的,以创建多层的深网络。这种层级结构可以在稍后的层处捕获更抽象的特征。来自深网络的最后的层的输出被用作为所学习的特征。虽然可以使用有分辨力的训练来精细调谐网络,但是输出可能不会捕获所有有关的信息。这可能是由缺少所标注的样本或受困于不良的局部最小值的向后传播优化而引起的。此外,深网络的完全连接的结构使得难以在大图像或体积的情况下进行学习。例如,当从 100×100 像素的彩色图像学习1000个特征时,网络将具有大约30兆个自由参数。

发明内容

- [0005] 提供了用于对象检测的系统、方法以及计算机可读介质。所述对象检测使用深的或多层的网络以针对检测图像中的对象而学习特征。来自不同层的多个特征被聚合并且被用于训练用于对象的分类器。除了来自不同层的特征聚合以外或者作为对于来自不同层的特征聚合的替换,初始层可以具有用于图像的不同的区域的分离的节点,以减少自由参数的数量。学习对象检测,或者应用所学习的对象检测器。
- [0006] 在第一方面中,提供了用于对象检测的方法。获得对象的图像。定义多层特征学习网络的多个顺序特征层。提供从网络的多个层接收特征的聚合器层。处理器使用对象的图像来联合地优化所述多层特征学习网络和所述聚合器层。处理器输出由被优化的多层特征学习网络表示的所学习的特征的集合以及利用通过所述聚合器层而生成的特征以检测对象的检测器。所学习的特征的集合用于区分所述对象,并且所述检测器用于对所述对象进行分类。

[0007] 在第二方面中,一种非暂态计算机可读存储介质已经在其中存储有表示由被编程的处理器可执行以用于对象检测的指令的数据。所述存储介质包括用于如下的指令:接收对象的图像;通过所述处理器检测具有从深网络的隐藏层生成的特征的图像中的对象,隐藏特征是从所聚合的不同的抽象层学习的,并且被由机器从不同的层输入馈送到所述隐藏特征的机器学习分类器中作为特征矢量;以及输出对象的检测。

[0008] 在第三方面中,提供了用于对象检测的方法。对象的图像被划分为局部部分。机器训练特征节点的第一块以重构图像的局部部分。机器训练第二属性节点的第二块以重构特征节点。机器还训练第三属性节点的第三块以重构第二属性节点。机器训练特征聚合器以对对象进行分类。利用第二和第三属性节点作为输入来训练特征聚合器。

[0009] 可以单独地或组合地使用以上所描述的各方面中的任何一个或多个。根据要被有关于随附附图来阅读的优选实施例的以下详细描述,这些和其它方面、特征和优点将变得明显。本发明由随后的权利要求限定,并且该部分中的内容不应当被看作为对那些权利要求的限制。本发明的进一步的方面和优点被在下面结合优选实施例来进行讨论,并且可以稍后被独立地或组合地要求保护。

附图说明

[0010] 组件和各图未必成比例,相反重点被放在图解实施例的原理。此外,在各图中,相同的参考标号贯穿不同的视图指定对应的部分。

[0011] 图1是用于对象检测训练的方法的一个实施例的流程图示图;

图2示出用于使用机器学习的对象检测的重新标识的示例图像;

图3图解具有用于图像的局部部分的分离的所学习的特征和来自多个层的特征聚合器的深网络的示例配置;

图4是用于对象检测应用的方法的一个实施例的流程图示图;

图5是示出用于使用不同的深网络的重新标识的累计匹配特性中的错误率的图线;以及

图6是用于对象检测的系统的一个实施例的框图。

具体实施方式

[0012] 一个或多个特征的质量对于很多图像分析任务是重要的。可以使用机器学习从原始数据构造有用的特征。与人类相比,机器的涉入可以更好地区分或标识有用的特征。给定用于图像的大量可能的特征以及各种各样的图像源,机器学习方法比人工编程更鲁棒。对特征进行区分的机器学习可以被应用于依赖于图像分析(诸如人重新标识或器官检测)的各种产品。通过利用机器来学习区分特征并且训练分类器来提供对象检测(无论是器官、人重新标识还是其它对象)。

[0013] 提供用于从原始图像数据构造特征的深网络框架。并非仅使用预先编程的特征(诸如所提取的Haar小波、色调饱和度值(HSV)直方图或局部二元模式(LBP)),深网络框架用于学习用于检测(诸如用于识别或重新标识的检测)的特征。例如,在检测中,在第一层中学习用于对象的各部分检测的局部特征。局部特征可以被用于快速地摆脱并非关注对象或对象特性的大量容易的候选。可以使用在深网络的稍后的层中学习的更抽象的特征来剪除

更难的候选。利用具有任何数量的层或深度的深网络来执行学习处理。最后,来自多于一个的层(诸如两个或更多个中间层)的所有或一些特征被聚合并且被直接馈送到分类器以用于最终预测。作为机器学习部分,分类器被与深网络的其它层一起联合地优化。与利用预先编程的特征进行的机器训练相比和/或与在没有聚合或没有局部到全局层的情况下利用深网络来学习特征相比,这种聚合、局部到全局网络结构或这两者可以提供更高效和/或更不易于出错的检测。

[0014] 图1和图4示出用于对象检测的方法。对象的检测可以是对象的属性或区分特性的检测。可以检测部分或整个对象。用于对象检测的方法可以是用以学习如何检测对象的方法,或可以是用于检测对象的方法。图1针对对象检测器的机器训练。图4针对机器学习的对象检测器的应用。在这两种情况下,机器(诸如处理器、计算机或服务器)实现一些或所有动作。在一个实施例中,图6的系统实现方法。用户可以选择用于利用处理器进行的对象检测器的应用的图像文件,或选择用以从中学习特征的图像和利用处理器的分类器。使用机器允许处理大容量(例如很多像素的图像和/或很多图像)的信息,由人类处置该大容量的信息可能是非高效的、由人类在所需要的时间帧中处置该大容量的信息可能是不现实的、或者该大容量的信息归因于细微性和/或定时可能甚至不可能由人类来进行处理。

[0015] 以所示出的顺序提供了方法,但是可以提供其它顺序。对于图1而言,动作44和46可以被执行为定义具有用于分类器的聚合器层的深网络的一个动作。

[0016] 附加地,可以提供不同的或更少的动作。例如,不提供图1的动作41和/或42。作为另一示例,不提供图4的动作56。在这两个示例中,在没有针对图像的不同部分的分离的特征学习的情况下使用从不同的层直接到分类器中的特征聚合。还在其它示例中,提供用于捕获图像的动作和/或使用所检测的信息的动作。

[0017] 图1示出用于在通过处理器学习区分对象或对象特性的一个或多个特征中的对象检测的方法。特征可以由处理器使用以对对象或图像进行分类和/或由处理器使用以训练分类器。

[0018] 在动作40中,获得对象的一个或多个图像。通过数据传送、捕获和/或来自存储器的加载来获得图像。获得同一对象的任何数量的图片(诸如对象的一个、两个、几十个或几百个图像)。图像是在相同传感器或不同传感器的情况下对象。图像中可以包括一个立体图中的对象或来自很多立体图的对象。

[0019] 图像可以是任何对象(诸如器官、人或建筑)或多个对象(例如人群或多个器官)。在正性匹配基本事实(groud truth)图像中使用同一对象。替换地,在其中一般而言特征被训练以区分属性或人的正性匹配基本事实图像中使用很多不同的对象。

[0020] 使用任何一个或多个传感器来捕获图像。例如,使用x射线、计算机层析、荧光透视法、血管造影术、磁共振、超声、正电子发射层析或单光子发射计算层析来捕获器官的图像。可以获得在相同或不同的设置(例如视场)中使用相同或不同的成像形态(即传感器或传感器的类型)的相同或不同的患者的多个图像。医学图像中的关注对象可以是器官、囊肿、瘤、钙化其它异常。

[0021] 在一个实施例中,图像是安全或区域监控图像,其中对象是个人或人群。从摄像机或其它视频设备获得图像。可以以各种方式(诸如,经有线的或无线的网络的传送、从存储器加载、直接馈送、或对图像的其它存取)获得图像。

[0022] 图像在不同时间来自不同的摄像机和/或相同的摄像机。所获得的图像表示相应的摄像机的视场中的一个或多个人。随着人典型地沿着普通行进路径行进通过视频监测区域,不同的摄像机或相同的摄像机在不同的时间捕获同一人。

[0023] 各图像一般地来自相同的时段(诸如,在彼此的几分钟或几小时内)。时段覆盖其间人可能穿着相同衣物和/或其间人处于被监测区域内的期间。可以使用更长或更短的时段,诸如在稍后的数小时、数天或数星期重新标识已经返回被监测区域的同一人。

[0024] 图像可以是或可以不是同一人。例如,机场视频监测系统可以捕获几百或几千个不同的人的图像。一个图像可以不包括人、包括一个或很多人。很多人中的一个可以是在不同图像中所捕获的同一人,或者图像中的很多人中可以没有任何人是同一人。对于训练而言,具有同一人的图像以及不具有同一人的图像以及给定图像中的人中的哪些属于同一人是已知的,并且被用作为基本事实。

[0025] 图像可以是区域或者可以被剪修。对于重新标识或匹配而言,可以使用进行配合以仅包围任何图像中的一人的矩形框。每个这样的框可以被用作为图像,从而具有很多人的图片可以被用作为很多图像。区域的图像中的任何人被剪修从而所得到的人图像至少主要表示仅一个人。可以使用轮廓剪修,诸如其中检测并且剪修人或特定人的一般边界。可以使用对区域的图像和/或对个人的图像的其它图像处理,诸如低通、高通、带通、定向或其它滤波。

[0026] 图2示出要被使用在用于重新标识的训练中的示例图像。左边的两行图像示出要在其它图像中被重新标识的两个人。右边的图像集合示出用于重新标识的六个可能的候选。对于训练而言,已知的是,候选的第一图像是匹配,而其它五个不是。可以使用更大数量的匹配和/或非匹配。

[0027] 人重新标识的任务是确定给定主体是否已经出现在摄像机的网络上,并且如果他或她出现则检索该主体的有关图像。归因于跨各摄像机的视点、照度以及分辨率的改变,该问题是困难的。在先前的方法中,人重新标识系统经常包含两个步骤。第一步骤是特征提取(诸如,提取HSV直方图和LBP特征)。第二步骤学习分级函数(诸如使用基本事实以及HSV和LBP特征作为输入来训练分类器)。

[0028] 来自深网络的所学习的特征可以与(例如预先编程的或人工地设计的)其它手动设计的特征组合,以改进性能。例如,采用3D Haar特征以在器官检测任务中快速地消除容易的候选。从深网络学习的特征可以被用在随后阶段中,以剪除不能通过Haar特征检测的困难候选。这种混合方法可以改进检测算法的速度和准确度这两者。

[0029] 与传统方法相比,使用深网络可能是相对处理密集的。为了减少处理,对于动作42-48而言,传统方法可以与以下所使用的深网络方法组合。在动作41中,利用预先编程的特征来消除用于对象的匹配的候选。例如,HSV直方图和/或LBP以及对应的分类器用于排除并非是匹配的一个或多个图像。在图2的顶部行示例中,可以利用预先编程的特征使用机器训练的分类器将顶部行的第三、第五和第六图像确定为并非是匹配。在图2的底部行示例中,可以利用预先编程的特征使用机器训练的分类器将底部行的第三图像确定为并非是匹配。在其它示例中,候选全都处于同一图像中。在医学环境中,Haar小波特征可以被用于标识图像的并非为对象的区域。

[0030] 作为消除的结果,深网络具有更少的候选来应用。在训练的情形中,传统方法可以

被用于减少训练数据或图像。通过另一方法使用将通过的或未被消除的图像来训练深网络。动作48的优化在消除之后使用其余候选。

[0031] 消除可以被构建到深网络中。来自早前的层(诸如初始层)的特征被用于剪除候选,从而针对这些候选的处理不对稍后的或更深的层造成负担。更早前的层的特征对于用于在检测问题中快速剪除容易的候选的计算可能是更快的。

[0032] 在动作42中,图像中的每一个被划分为子区域。可以使用任何大小、形状和/或数量的子区域。给定的图像的各子区域是相同的大小和形状的,或可以具有不同的大小和形状。例如,图像的中心(例如在医学上的)或其它部分(例如人的剪修图像的上部中心)可能更有可能包括对象或关注的对象的部分,从而更小或更大的子区域形成在图像的该部分处。各子区域是相互排斥的,但是可以提供某种重叠。

[0033] 图3示出划分图像70的一个示例。给定的图像被划分为六个矩形局部部分。局部用于指示小于全部或并非全局的。局部部分要被分离地处置。针对每个局部部分独立于来自其它局部部分的信息而学习至少一些特征。在替换的实施例中,不划分图像。相反地,图像被作为整体来使用,或者从全局图像学习特征。

[0034] 在动作44中,定义深网络。定义是通过对学习的配置或编程来进行的。层的数量、学习的类型以及网络的其它特性由编程者或用户控制。在其它实施例中,在学习期间由机器定义并且选择一个或多个方面(例如节点的数量、层的数量或学习的类型)。在图3的示例中,深架构具有五个层,不包括聚合器。可以提供任何数量的层(诸如两个、三个、四个或更多)。

[0035] 多层特征学习网络被定义为多个顺序特征层。顺序用于指示作为对下一层的输入的可视图像的一般流动,其中,来自该下一层的信息被馈送到下一层,并且依此类推直到输出。例如,在图3中,来自第一顺序特征层72的输出是馈送到第二顺序特征层74的特征,来自第二顺序特征层74的输出是馈送到第三顺序特征层76的第二特征,第三顺序特征层76的输出是馈送到第四顺序特征层78的第三特征,并且第四顺序特征层78的输出是馈送到第五顺序特征层80的第四特征。该处理继续,直到网络到达想要的深度。所定义的层的数量可以基于试错。

[0036] 各层可以仅前向馈送,或可以是双向的,包括对先前的层的某种反馈。每个层的节点可以与先前的层或随后的层的所有节点或仅节点的子集连接。

[0037] 在层72-80内,提供任何数量的节点。例如,提供100个节点。可以针对不同的层72-80提供不同数量的节点(诸如层72的块中的每一个中的10个节点、用于层74和76的100个节点、以及用于层78和80的更少的节点)。稍后的或随后的层可以具有更多、更少或相同数量的节点。一般而言,随后的层具有更高抽象。例如,第一层提供来自图像的特征,诸如一个节点或特征是图像中所发现的线条。下一层组合线条,从而节点中的一个是角。下一层可以组合来自先前的层的特征(诸如角和线条的长度),从而节点提供形状或构建指示。

[0038] 在重新标识示例中,替代直接从低级特征寻找分级函数,将低级特征用于预测称为“属性”的更高级的或更抽象的特征。属性的示例是头发颜色、衬衫颜色、性别、裤子、裙子、包或其它。所预测的属性标注被用于检索主体的有关实例。例如,“寻找具有金色头发和深色衬衫的人”。属性可以导致更好的性能和如下的更直观的检索系统:人类可以对该系统提供直接的输入或询问。作为特征的属性是抽象的,并且其可以并不清楚什么低级特征(例

如HSV色彩直方图或LBP)对于属性(如性别)是有关的。

[0039] 定义深架构以学习在不同抽象水平上的特征,而不是对特征预先编程或尝试使特征与属性相关。学习特征以重构更低级的特征。例如,学习用于重构图像的特征。对于下一层而言,学习用于重构先前的层的特征的特征,提供更高抽象。

[0040] 深架构包括卷积神经网络(CNN)或深信念网络(DBN),但可以使用其它深网络。CNN学习前馈映射函数,而DBN学习数据的生成模型。此外,CNN针对所有局部区域使用共享权重,而DBN是全连接网络(即,针对图像的所有区域具有不同的权重)。CNN的训练是通过向后传播而是完全区分的。另一方面,DBN采用逐层的无监测训练(例如,预先训练),后随有具有向后传播的区分细化(如果必要的话)。

[0041] 层的每个节点表示特征。提供不同的节点以用于学习不同的特征。可以使用任何数量的节点。在图3的示例中,每个框或层72-80一般表示多个节点。

[0042] 机器使用任何构建块来学习节点的特征。例如,使用自动编码器(AE)或受限制的玻尔兹曼机(RBM)。图3示出针对每个层72-80使用AE。AE线性地变换数据,并且然后应用非线性校正(如S形函数)。AE的目标函数是输入图像与使用所学习的特征的所重构的图像之间的期望均方差。可以使用随机梯度下降或其它方法来训练AE,以通过机器学习导致最佳重构的特征。

[0043] RBM的目标函数是能量函数。精确计算与RBM关联的似然率项是棘手的。因此,近似算法(诸如基于k步Gibb采样的对比散度或其它)被用于训练RBM以从特征重构图像。

[0044] 对于高维输入数据而言AE或RBM的训练倾向于过度拟合。采用稀疏性或降噪技术(例如稀疏降噪AE(SDAE))以约束参数的自由度,并且强制学习数据内的关注结构。添加噪声以训练图像并且要求网络重构无噪声图像可以防止过度拟合。在隐藏层内强制稀疏性(即,仅同时地激活隐藏层中的少数量的单元)也可以调整网络。

[0045] 图3示出多层次特征学习和聚合的网络架构的一个示例定义。在该示例中,网络的框架是用于从原始输入图像自动化地构造有用的特征的深架构。网络使用局部到全局方法,其中,初始层72具有用于独立地学习用于局部区域的特征的分离的块,并且稍后的层74-80全局地对针对整个图像70的特征进行操作。第一顺序特征层72被定义为用于图像70的相应的子区域的分离的局部特征学习器(例如AE的块)。提供层72中的六个这样的子区域以及对应的六个分离地学习的或训练的节点或特征的集合。例如,图像被划分为多个 16×16 像素片块。空间划分有助于在利用大图像的学习中减少计算负担。例如,与全连接网络(即,第一层是全局的)相比,自由变量的数量减少至少于十分之一。学习具有m个隐藏单元或节点(例如m=100)的稀疏降噪自动编码器(SDAE)(即具有稀疏性和降噪正规化子的自动编码器)得以学习,以良好地重构每个区域。针对不同的局部区域学习不同滤波器。可以使用其它数量的子区域和隐藏节点。

[0046] 可以提供使用局部划分的任何数量的层。图3示出具有局部特征学习的仅一个层72。其它层(诸如第二层74)被定义为全局特征学习器,以用于从分离的局部特征学习器的输出学习用于各图像的整体的特征。例如,来自层72的所有局部SDAE的输出被聚合,并且被馈送到第二层74中的SDAE。来自第二层74的输出被馈送到第三层76中的另一SDAE。在替换的实施例中,不使用局部到全局方法。

[0047] 在图1的动作46中,提供聚合器层82。该聚合器层被定义为深网络的部分,或者在

对网络进行定义当中被分离地处理。聚合器层82使用特征以进行分类或检测而不是提供重构以学习特征。聚合器层82是终止特征学习网络的分类器。例如，聚合器层82是概率提升树、支撑矢量机或其它机器学习分类器。其它分类器可以包括单类或二元分类器、不同分类器的集合、级联分类器、层级分类器、多类分类器、基于模型的分类器、基于机器学习的分类器，或者可以使用它们的组合。多类分类器包括CART、K最近邻、神经网络(例如多层次感知)、混合模型或其它。可以使用纠错输出码(ECOC)。

[0048] 聚合器层82从多个特征层的节点接收特征。聚合器层82的输入矢量包括直接来自多个不同层74-80的所有或一些特征以及对应的抽象水平，而不是利用最抽象层80的特征进行工作。例如，来自最后两个、三个或四个层的特征被聚合。在图3的示例中，聚合器层82从层74-80接收所有特征或节点输出。在替换的实施例中，来自更少或更多的层72-80的特征被聚合。在其它替换的实施例中，使用比来自给定层的所有特征或节点输出更少的特征或节点输出，诸如其中特征并非是类或检测的确定而是用于确定随后的层的特征。

[0049] 在图1的动作48中，使用对象的图像来对多层次特征学习网络和聚合器层一起进行训练。执行机器学习以使用所定义的深架构来训练各个层72-82。学习确定性的或允许重构输入的特征。类似地，学习提供由聚合器层82实现的检测器的想要的结果或分类的特征。还学习分类器。

[0050] 为了进行训练以一起工作，联合地优化多层次特征学习网络和聚合器层的分类器。与用于分类的基本事实以及针对用于特征学习网络的重构的错误有关的结果被向后投射，以对哪些特征工作得最佳进行训练。使用从聚合器层82到特征层74-80中的多个层的向后投射以及各顺序层之间的向后投射(例如，80到78、78到76、76到74、以及74到72)造成与用于输入图像70的基本事实有关的所学习的特征或节点和分类器的细化。在训练中，网络不仅将梯度从最后的层传播到先前的层，而且还传播到与聚合器层82连接的所有中间层，联合地优化整个网络。联合优化造成功能和特征学习网络被一起训练以最佳地检测对象(诸如，检测对象的一个或多个特性)。

[0051] 还可能的是连同学习分类器(诸如，连同学习属性分类器)一起使用向后传播来精细调谐网络。例如，重构错误被用于向后投射，以用于训练用以在层72-80中进行检测的特征。然后利用或不利用其它层72-80的特征的细化来训练聚合器层82。所学习的特征被用于训练聚合器层82的一个或多个分类器。

[0052] 在优化期间，学习不同的区分特征。在图3的示例中，机器训练特征节点的块以重构图像的局部部分。机器还训练属性节点的块以重构特征节点，并且训练其它属性节点的块，以重构先前的属性节点。机器还诸如通过利用概率提升树或支撑矢量机进行训练来学习聚合器层的映射函数。特征聚合器被训练以使用来自多个其它层的特征作为输入来对对象进行分类。映射函数的训练与用于检测的所有输入特征或输入特征的子集有关。例如，聚合器层学习映射函数，以从来自多层次特征学习网络的多个层的节点的特征的任何组合预测21个属性。学习可以仅指示来自要被用于分类的不同节点的特征的子集。

[0053] 在动作50中，输出所学习的特征和检测器。所学习的特征由被优化的多层次特征学习网络表示。单独的特征的集合形成用于将对象与其它对象进行区分的特征或特征集合。基于从图像重构对象而提供特征作为不同的抽象水平上的特征层的节点。节点定义被训练以提取特征的处理。

[0054] 检测器是用于对对象进行分类的分类器。使用输出特征，检测器检测对象的特性或属性。例如，在人重新标识实施例中，输出检测器提供用于21个属性的标注。属性是用于检测匹配的特征。标注可以是二元的（例如，对于属性而言，“是”或“否”）或被分级的（诸如1-10的大小）。替换地，检测器对来自一个图像的输入人是否为另一图像中的同一人进行分类。检测器可以使使用来自早前的层的特征，以在检测处理期间快速地拒绝不可能的候选。

[0055] 一旦被训练，就输出矩阵。矩阵表示用于特征和检测器的所学习的算法。替换地，分离的矩阵被用于任何节点、层、网络和/或检测器。

[0056] 当新的传感器被添加到系统（诸如，摄像机被添加到安全或居住监控系统）时，用以使用的特征以及检测器被训练，以用于与该摄像机一起使用。当开发了新的类型的传感器时，可以从来自传感器数据的该新的类型的传感器训练特征和检测。在其它实施例中，针对任何给定的匹配或检测情况执行训练。例如，在重新标识中，要定位在其它图像中的一个人的一个或多个图像被用于学习特征并且训练检测器。其它图像可以于是被用于确定匹配与否。随着人被定位，针对该人来具体地训练特征和检测器。

[0057] 一旦学习了特征并且训练了检测器，就可以应用特征和检测器。定义一个或多个特征的矩阵被用于从输入图像进行提取。检测器使用来自图像的所提取的特征，以将输入图像分类为表示对象或不表示对象。输出可以是二元的“是/否”或者为是同一人的概率。输出可以用于不同的属性而非匹配或不匹配。可以提供诸如与替换的值关联的概率或其它信息。

[0058] 图4是对象检测应用的一个实施例的流程图。相同的处理器或不同的处理器应用所学习的特征和检测器。例如，一个或多个矩阵被传送到另一处理器以用于该处理器进行应用。

[0059] 在动作54中，处理器接收对象的一个或多个图像。图像可以包括相同的对象或不同的对象。应用用来确定对象或对象的特性是否相同。从传感器接收图像作为可视信息。对于其它类型的传感器而言，信息可以是数据而非图像。

[0060] 在动作56中，在利用深网络的隐藏特征进行检测之前，处理器利用预先编程的特征来限制候选。例如，处理器使用Haar、HSV色彩直方图、LBP、梯度和/或其它信息来应用一个或多个分类器以确定每个图像是否为对象，或确定图像的不同部分是否为对象。然后可以仅对于其余的候选（例如图像和/或图像的部分）应用所学习的多层网络和分类器。

[0061] 在动作58中，使用深网络的隐藏特征来检测输入图像或图像的部分是否表示对象。例如，所训练的AE或RBM块被应用于适当的输入，以提取对应的特征。隐藏特征是从多层网络中的不同的抽象层学习的特征节点。使用所学习的变换、非线性修正、能量函数或其它函数，从图像提取一个或多个输入图像的特征。可以使用层结构从那些所提取的特征提取其它更抽象的特征。取决于层的数量，从先前的层的特征提取其它特征。特征是表示相似性的量（诸如是对象中的特征的概率）的值（诸如二进制值、整数值或分数值）。

[0062] 在一个实施例中，检测包括针对图像的不同子部分使用分离的隐藏特征。提取局部或子区域特定特征。可以提取使用所提取的局部特征的其它特征（例如全局特征）。

[0063] 检测使用来自不同的层的所提取的特征作为对聚合器的机器学习分类器的特征矢量或输入。来自不同层的隐藏特征被聚合为对机器学习分类器的输入特征矢量。分类器然后输出检测（诸如图像是否为匹配、匹配的概率（例如人具有长袖衬衫的88%概率）、或者

对象或对象的特性的其它指示符)。

[0064] 在动作60中,输出检测。输出结果或所检测的信息。例如,输出是否存在匹配。作为另一示例,输出针对一个或多个属性的匹配的概率。可以针对对象或对象的部分输出任何信息或分类。例如,输出针对很多(例如10个或更多个)属性的匹配的概率。

[0065] 输出是对于图像的。例如,同时显示基本事实图像以及检测所应用于的图像。检测的结果指示是否存在匹配或其它检测。在图2的示例中,任何匹配的图像被突出显示(诸如被围绕图像放置的框)。在其它示例中,以不同的方式突出显示匹配的属性和/或突出显示非匹配的属性。

[0066] 在用于重新标识的训练和应用的一个示例中,与使用HSV色彩直方图和LBP的机器学习方法相比,使用来自分类器中的不同层的特征的局部到全局特征学习和聚合改进属性分类。VIPER数据集具有632个主体,从不同的视点来看,主体的每一个具有两个图像。利用具有第一层中的32个局部区域、来自3个最后的层的5个SDAE块以及聚合的L个层(例如,L=5,网络大小18432–4800–3000–500–200–42)来定义网络。使用来自深网络的所学习的特征,级别-1检索结果内的34.2%的准确度与使用级别提升方法(该方法使用HSV直方图和LBP)的当前现有技术状况的15.2%比较。与现有技术状况的40.6%和53.3%相比,级别-5和级别-10结果分别是52.1%和63.3%。

[0067] 图5示出针对不同方法的分级性能的比较,其中SAE属性和CNN属性对应于如下的多层特征聚合网络的性能:所述网络的构建块是SAE和CNN,所述网络针对人标识任务使用添加有中间的所学习特征的属性分类输出。LBP-HSV-EFM是现有技术方法的状况,现有技术方法使用具有局部二元模式、色彩直方图以及显式特征映射技术的组合的级别提升算法。PCA对应于与支持矢量分级方法组合来使用主分量分析以产生特征。SAE是在对于人重新标识而言不使用属性分类输出的情况下针对多层特征聚合网络的结果。

[0068] 可以仅使用最后的层的输出作为用于分类器的输入特征来执行同一人重新标识实验。与多层特征聚合相比,针对级别-1、级别-5以及级别-10准确度分别下降到27.4%、49.4%以及58.2%。中间层的特征聚合可以一致地改进性能。

[0069] 在上面的一些示例中使用SDAE。相同的网络定义可以被用于从L个层(例如,L=5,网络配置是 5×5 滤波器 $\rightarrow 2 \times 2$ 池化 $\rightarrow 5 \times 5$ 滤波器 $\rightarrow 3 \times 3$ 池化 \rightarrow 21个输出)的CNN网络构造特征。来自最后3个层的特征被聚合。针对级别-1、级别-5和级别-10结果准确度可以是20.9%、46.8%以及61.8%。

[0070] 在另一示例中,对比于HSV和LBP特征而使用所学习的特征的属性分类执行得更好。例如,在关于三个属性“深色衬衫”、“深色底部”以及“男性”的级别提升方法的情况下,与HSV直方图和LBP特征的20.5%、32.7%和46%相比,针对利用深网络中的层聚合的局部到全局的错误率分别是16%、27%和39.5%。

[0071] 图6示出用于人重新标识中的对象检测的系统。系统是主机计算机、控制站、工作站、服务器或用于人重新标识的其它布置。系统包括显示器14、存储器16以及处理器18。可以提供附加的、不同的或更少的组件。系统用于训练,诸如使用来自监测系统11的图像作为基本事实。替换地,系统用于所学习的特征和分类器的应用。在其它实施例中,监测系统11和摄像机10被医学成像器或其它传感器替代。

[0072] 显示器14、处理器18和存储器16可以是用于对来自视频监测系统11的摄像机10的

图像进行图像处理的计算机、服务器或其它系统的部分。可以使用用于系统11的工作站或控制站。替换地，使用并非为视频监测系统11的部分的分离的或远程的设备。替代地，远程地执行重新标识。

[0073] 在一个实施例中，处理器18和存储器16是管控用于由作为客户机的监测系统11使用的重新标识功能的服务器的部分。客户机和服务器通过网络(诸如内联网或互联网)互连。客户机可以是视频监测系统11的计算机，并且服务器可以由视频监测系统11的制造商、提供商、主机或创建者提供。

[0074] 视频监测系统11包括多个摄像机10。摄像机被遍布于区域(诸如建筑、建筑的楼层、仓库、校园、综合楼、室外区域、表演场、机场、交通区域、枢纽、城市或隧道系统)而分布。摄像机10被安装并且连接到通信网络。每个摄像机10捕获区域的场景。视频摄像机10可以遵循开放网络视频接口论坛(ONVIF)规范。摄像机视频流被耦合到视频处理单元，视频处理单元被耦合到视频存储体以及视频内容分析(VCA)模块。VCA模块将加时间戳的视频数据和元数据输出到它们的相应的存储体。VCA模块根据算法来生成并且输出元数据。视频和元数据存储体可以被实现为数字视频记录器(DVR)或网络视频记录器(NVR)，以用于存储并且回放。对视频数据进行编码并且存储。可以与视频一起存储基于流的元数据。可以提供用于视频监测的其它布置(诸如规则地捕获静止图像而非视频)。

[0075] 显示器14是CRT、LCD、投影仪、等离子体、打印机、智能电话或其它现在已知的或稍后开发的显示设备，以用于显示图像、属性、重新标识和/或对象检测信息。例如，显示器14显示两个图像、关于用于图像的摄像机位置的信息、以及同一人是否处于这两个图像中的指示。显示器14协助确认匹配或重新标识。可以替代地或附加地显示属性信息。在训练环境中，显示可以是网络定义、特征信息、向后投射信息、联合优化信息或其它训练信息。

[0076] 指令、网络定义、特征、分类器、矩阵、输出和/或其它信息存储在非暂态计算机可读存储器(诸如存储器16)中。存储器16是外部存储设备、RAM、ROM、数据库和/或本地存储器(例如固态驱动器或硬驱动器)。相同或不同的非暂态计算机可读介质可以被用于指令和其他数据。可以使用由处理器18管理的并且驻留在存储器(诸如硬盘、RAM或可拆卸介质)上的数据库管理系统(DBMS)来实现存储器16。替换地，存储器16处于处理器18内部(诸如缓存)。

[0077] 在非暂态计算机可读存储介质或存储器(诸如缓存、缓冲器、RAM、可拆卸介质、硬驱动器或其它计算机可读存储介质(诸如存储器16))上提供用于实现此所讨论的训练或应用处理、方法和/或技术中的对象检测的指令。计算机可读存储介质包括各种类型的易失性和非易失性存储介质。响应于在计算机可读存储介质之中或之上存储的一个或多个指令集而执行各图中所图解或在此所描述的功能、动作或任务。功能、动作或任务独立于特定类型的指令集、存储介质、处理器或处理策略，并且可以由单独地或组合地操作的软件、硬件、集成电路、固件、和微代码等执行。

[0078] 在一个实施例中，指令被存储在可拆卸介质设备上以用于由本地或远程系统读取。在其它实施例中，指令被存储在远程位置中以用于通过计算机网络传送。在又一实施例中，指令被存储在给定的计算机、CPU、GPU或系统内。因为随附各图中所描绘的构成系统组件和方法步骤中的一些可以被实现在软件中，所以各系统组件(或处理步骤)之间的实际连接可以取决于对本实施例进行编程的方式而不同。

[0079] 程序可以被加载到包括任何合适的架构的处理器18中并且由该处理器18执行。类

似地,处理策略可以包括多处理、多任务和并行处理等。在具有硬件(诸如一个或多个中央处理单元(CPU)、随机存取存储器(RAM)和(多个)输入/输出(I/O)接口)的计算机平台上实现处理器18。计算机平台还包括操作系统和微指令代码。在此所描述的各种处理和功能可以是经由操作系统执行的微指令代码的一部分或程序的一部分(或其组合)。替换地,处理器18是网络中的一个或多个处理器。

[0080] 处理器18被配置为获得图像。图像可以是视场或从视场剪切出的人。处理器18可以执行剪切。图像的区域或划分可以由处理器18描画。

[0081] 处理器18被配置为学习特征或提取所学习的特征。例如,多层次特征学习网络被定义并且被用于学习特征以重新标识人。在不同的层处学习特征。在一个实施例中,学习局部特征和全局特征。可以通过应用所学习的处理来提取所学习的特征。

[0082] 处理器18被配置为基于所学习的特征而进行分类。输入特征以进行学习,以对输入图像进行分类。使用基本事实和联合优化,利用对特征的学习来学习分类器。对于应用而言,处理器18基于所输入的所学习的特征来确定机器学习分类器的输出。

[0083] 可以一起地或分离地使用在此所描述的各种改进。虽然已经参照随附附图在此描述了本发明的说明性的实施例,但是要理解的是,本发明不限制于这些精确实施例,并且在不脱离本发明的范围或精神的情况下,本领域技术人员可以在其中实行各种其它改变和修改。

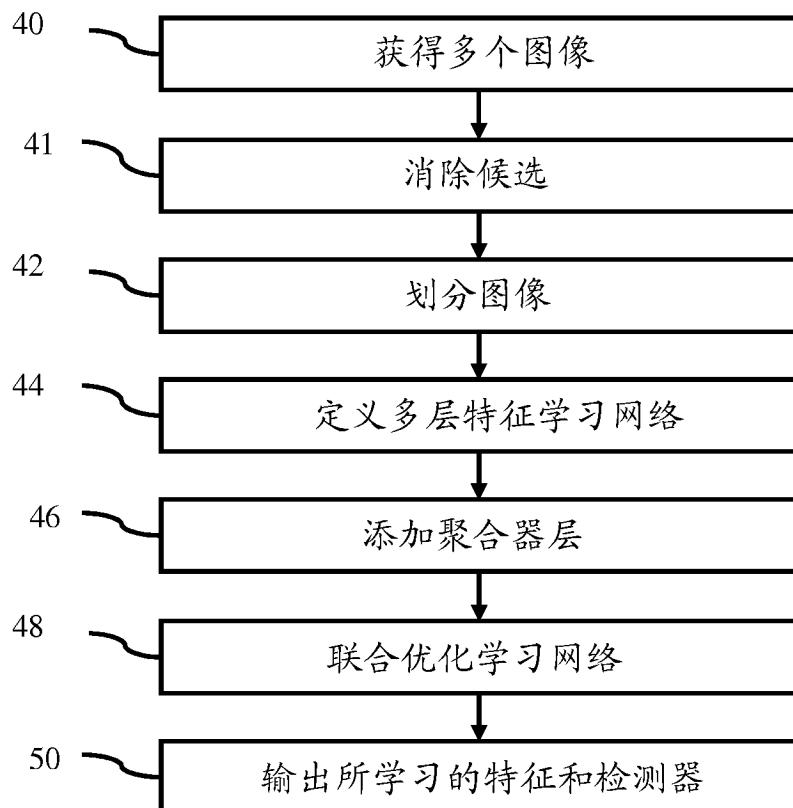


图 1



图 2

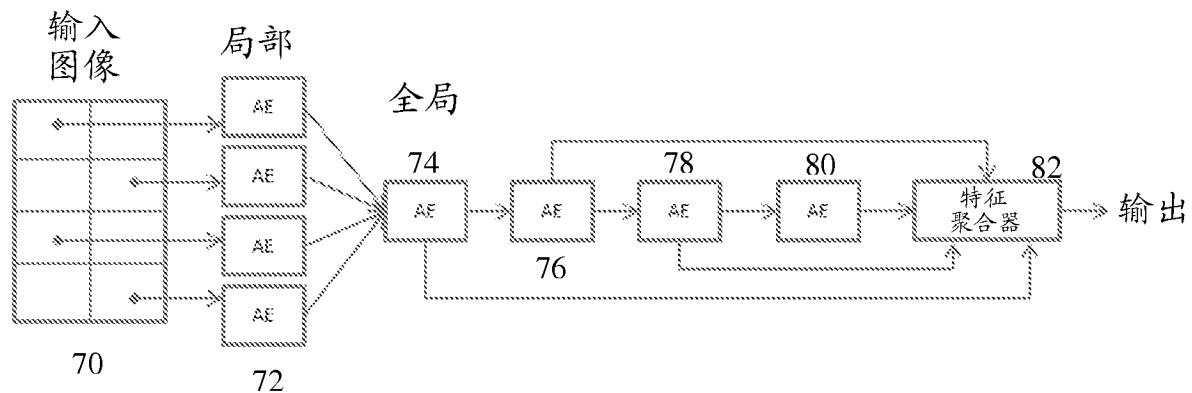


图 3

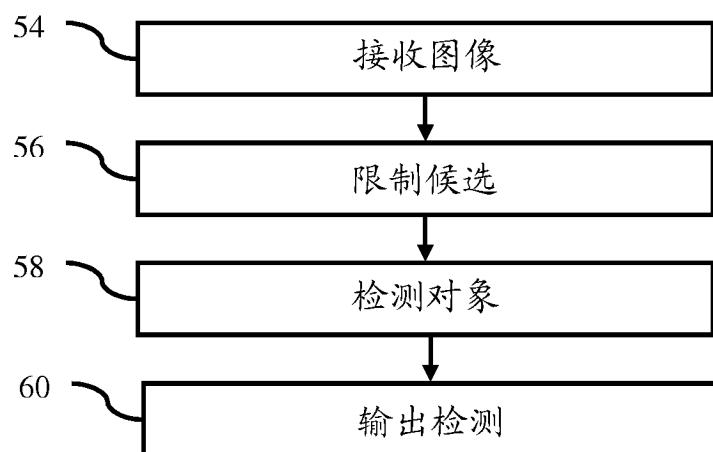


图 4

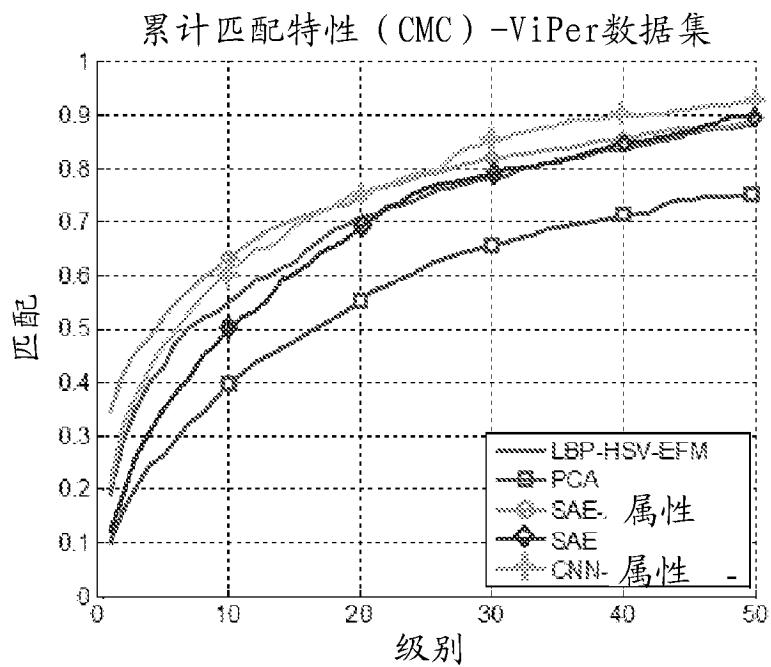


图 5

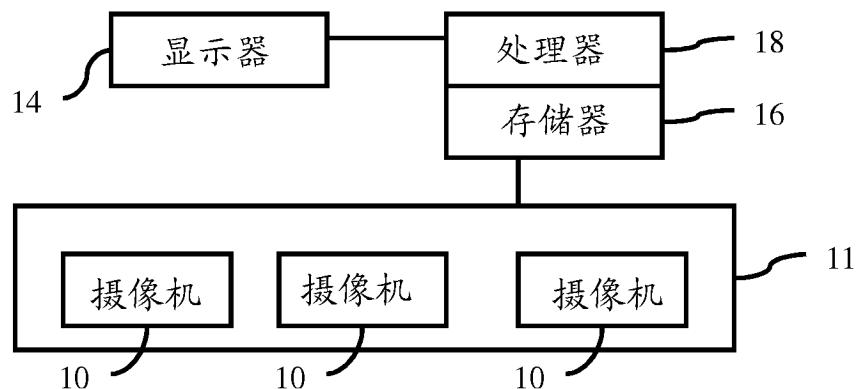


图 6