



(12) 发明专利

(10) 授权公告号 CN 110582777 B

(45) 授权公告日 2023. 08. 01

(21) 申请号 201880022737.X

(22) 申请日 2018.04.10

(65) 同一申请的已公布的文献号
申请公布号 CN 110582777 A

(43) 申请公布日 2019.12.17

(30) 优先权数据
62/502,461 2017.05.05 US

(85) PCT国际申请进入国家阶段日
2019.09.29

(86) PCT国际申请的申请数据
PCT/US2018/026951 2018.04.10

(87) PCT国际申请的公布数据
W02019/018022 EN 2019.01.24

(73) 专利权人 赫尔实验室有限公司

地址 美国加利福尼亚州

(72) 发明人 S·科洛瑞 S·R·劳 金劲男

(74) 专利代理机构 北京三友知识产权代理有限公司 11127

专利代理师 王万影 王小东

(51) Int.Cl.

G06V 20/00 (2022.01)

G06V 10/764 (2022.01)

G06V 20/70 (2022.01)

G06V 10/82 (2022.01)

审查员 杨斌

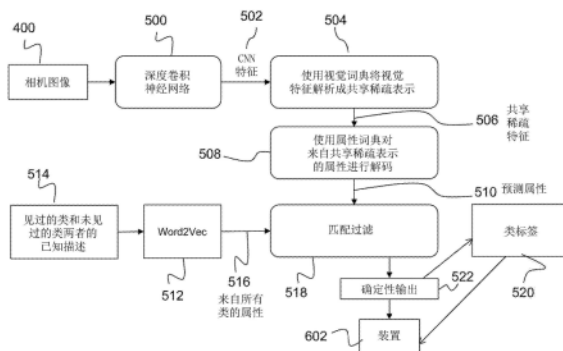
权利要求书2页 说明书9页 附图6页

(54) 发明名称

借助于联合稀疏表示的零样本机器视觉系统

(57) 摘要

描述了这样一种系统,即,所述系统可以识别该系统以前从未见过的新颖对象。所述系统使用训练图像集来学习将来自己知图像的视觉特征映射至语义属性的模型。使用所学到的模型将未见过的输入图像的视觉特征映射至语义属性。利用类标签将所述未见过的输入图像分类为属于图像类。基于所述类标签来控制装置。



1. 一种用于对象识别的系统,所述系统包括:

一个或多个处理器以及非暂时性计算机可读介质,在所述非暂时性计算机可读介质上编码有可执行指令,以使当执行所述可执行指令时,所述一个或多个处理器执行以下操作:

使用训练图像集来学习将来自己知图像的视觉特征映射至语义属性的模型,其中,所学到的模型是提供视觉特征的稀疏表示和语义属性的稀疏表示的联合稀疏表示的联合稀疏词典模型,其中,所述联合稀疏词典模型强制所述视觉特征的稀疏表示和所述语义属性的稀疏表示是相同的;

使用所学到的模型将未见过的输入图像的视觉特征映射至语义属性;

利用类标签将所述未见过的输入图像分类为属于图像类;以及

基于所述类标签来控制装置。

2. 根据权利要求1所述的系统,其中,所述装置是运载工具组件,并且其中,所述一个或多个处理器还执行控制所述运载工具组件以使得执行运载工具操作的操作,其中,所述运载工具操作是制动和转弯中的至少一者。

3. 根据权利要求1所述的系统,其中,所述一个或多个处理器还执行生成训练图像集的操作,所述训练图像集包括来自多个图像类的对象的图像,其中,所述训练图像集中的各个图像已经注释有类标签和语义属性。

4. 根据权利要求1所述的系统,其中,卷积神经网络CNN从已知图像中提取视觉特征。

5. 根据权利要求1所述的系统,其中,所述一个或多个处理器还执行以下操作:

从所述未见过的输入图像中提取视觉特征;

针对各个提取图像特征,确定所述未见过的输入图像的稀疏表示;并且

使用所述未见过的输入图像的所述稀疏表示来确定至少一个图像属性;以及

通过将所述至少一个图像属性映射至标签空间来确定所述未见过的输入图像的所述类标签。

6. 根据权利要求5所述的系统,其中,求解Lasso问题,以确定所述未见过的输入图像的所述稀疏表示。

7. 根据权利要求1所述的系统,其中,所述系统是零样本学习机器视觉系统。

8. 一种用于对象识别的计算机实现的方法,所述方法包括以下动作:

使一个或多个处理器执行被编码在非暂时性计算机可读介质上的指令,使得在执行所述指令时,所述一个或多个处理器执行以下操作:

使用训练图像集来学习将来自己知图像的视觉特征映射至语义属性的模型,其中,所学到的模型是提供视觉特征的稀疏表示和语义属性的稀疏表示的联合稀疏表示的联合稀疏词典模型,其中,所述联合稀疏词典模型强制所述视觉特征的稀疏表示和所述语义属性的稀疏表示是相同的;

使用所学到的模型将未见过的输入图像的视觉特征映射至语义属性;

利用类标签将所述未见过的输入图像分类为属于图像类;以及

基于所述类标签来控制装置。

9. 根据权利要求8所述的方法,其中,所述一个或多个处理器还执行生成训练图像集的操作,所述训练图像集包括来自多个图像类的对象的图像,其中,所述训练图像集中的各

个图像已经注释有类标签和语义属性。

10. 根据权利要求8所述的方法,其中,卷积神经网络CNN从已知图像中提取视觉特征。

11. 根据权利要求8所述的方法,其中,所述一个或更多个处理器还执行以下操作:

从所述未见过的输入图像中提取视觉特征;

针对各个提取图像特征,确定所述未见过的输入图像的稀疏表示;并且

使用所述未见过的输入图像的所述稀疏表示来确定至少一个图像属性;以及

通过将所述至少一个图像属性映射至标签空间来确定所述未见过的输入图像的所述类标签。

12. 根据权利要求11所述的方法,其中,求解Lasso问题,以确定所述未见过的输入图像的所述稀疏表示。

13. 一种用于对象识别的非暂时性计算机可读介质,所述非暂时性计算机可读介质包括计算机可读指令,所述计算机可读指令能够由具有一个或更多个处理器的计算机执行,以使所述一个或更多个处理器执行以下操作:

使用训练图像集来学习将来自已知图像的视觉特征映射至语义属性的模型,其中,所学到的模型是提供视觉特征的稀疏表示和语义属性的稀疏表示的联合稀疏表示的联合稀疏词典模型,其中,所述联合稀疏词典模型强制所述视觉特征的稀疏表示和所述语义属性的稀疏表示是相同的;

使用所学到的模型将未见过的输入图像的视觉特征映射至语义属性;

利用类标签将所述未见过的输入图像分类为属于图像类;以及

基于所述类标签来控制装置。

14. 根据权利要求13所述的非暂时性计算机可读介质,其中,所述装置是运载工具组件,并且其中,所述一个或更多个处理器还执行控制所述运载工具组件以使得执行运载工具操作的操作,其中,所述运载工具操作是制动和转弯中的至少一者。

15. 根据权利要求13所述的非暂时性计算机可读介质,其中,卷积神经网络CNN从已知图像中提取视觉特征。

16. 根据权利要求13所述的非暂时性计算机可读介质,所述非暂时性计算机可读介质还包括使所述一个或更多个处理器进一步执行以下操作的指令:

从所述未见过的输入图像中提取视觉特征;

针对各个提取图像特征,确定所述未见过的输入图像的稀疏表示;并且

使用所述未见过的输入图像的所述稀疏表示来确定至少一个图像属性;以及

通过将所述至少一个图像属性映射至标签空间来确定所述未见过的输入图像的所述类标签。

17. 根据权利要求16所述的非暂时性计算机可读介质,其中,求解Lasso问题,以确定所述未见过的输入图像的所述稀疏表示。

借助于联合稀疏表示的零样本机器视觉系统

[0001] 相关申请的交叉引用

[0002] 本申请是2017年5月5日在美国提交的题为“Zero Shot Machine Vision System via Joint Sparse Representations”的美国临时申请No.62/502,461的非临时专利申请,该美国临时申请的全部内容通过引用并入于此。

技术领域

[0003] 本发明涉及用于识别新颖对象的系统,并且更具体地,涉及通过使用联合稀疏表示对视觉特征与语义属性之间的关系进行建模来识别新颖对象的系统。

背景技术

[0004] 尽管没有接收到任务的任何训练示例,零样本学习(Zero-shot learning)也能够解决该任务。Akata等人(参见所并入的参考文献的列表中的参考文献No.1)和Romera等人(参见参考文献No.2)描述了零样本机器视觉方法。在参考文献No.1中,作者提出了一种将图像特征和语义属性嵌入在公共空间中的模型(即,潜在嵌入(latent embedding)),其中,借助于双线性函数来测量它们之间的兼容性。

[0005] Romera等人(参见参考文献No.2)提出了一种涉及图像特征、属性以及类的通用线性框架。参考文献No.2呈现的方法利用使作者能够推动针对问题的简单封闭形式解的正则化项原理选择。上述方法的主要缺点包括:1)通过假设数据特征与语义属性之间的线性关系来过度简化问题,以及2)对需要针对各个应用进行调整的特设正则化项的敏感性。

[0006] 因此,仍然需要这样的系统,即,该系统通过利用对对象与对象的语义属性之间的关系进行编码的数学上严密的模型来改进机器视觉系统的零样本能力。

发明内容

[0007] 本发明涉及用于识别新颖对象的系统,并且更具体地,涉及通过使用联合稀疏表示对视觉特征与语义属性之间的关系进行建模来识别新颖对象的系统。所述系统包括一个或多个处理器以及非暂时性计算机可读介质,在所述非暂时性计算机可读介质上编码有可执行指令,以使当执行所述可执行指令时,所述一个或多个处理器执行多个操作。使用训练图像集来学习将来自已知图像的视觉特征映射至语义属性的模型。使用所学到的模型将未见过的输入图像的视觉特征映射至语义属性。利用类标签将所述未见过的输入图像分类为属于图像类。基于所述类标签来控制装置。

[0008] 在另一方面,所述装置是运载工具组件,并且控制所述运载工具组件以使得执行运载工具操作,其中,所述运载工具操作是制动和转弯中的至少一者。

[0009] 在另一方面,所述系统生成训练图像集,所述训练图像集包括来自多个图像类的对象的图像,其中,所述训练图像集中的各个图像已经注释有类标签和语义属性。

[0010] 在另一方面,卷积神经网络(CNN)从已知图像中提取视觉特征。

[0011] 在另一方面,所学到的模型是提供视觉特征和语义属性的联合表示的联合稀疏词

典模型。

[0012] 在另一方面,所述系统从所述未见过的输入图像中提取视觉特征。针对各个提取图像特征,所述系统确定所述未见过的输入图像的稀疏表示。所述系统使用所述未见过的输入图像的所述稀疏表示来确定至少一个图像属性。通过将所述至少一个图像属性映射至标签空间来确定所述未见过的输入图像的所述类标签。

[0013] 在另一方面,求解Lasso问题,以确定所述未见过的输入图像的所述稀疏表示。

[0014] 在另一方面,所述系统是零样本学习机器视觉系统。

[0015] 最后,本发明还包括计算机程序产品和计算机实现的方法。所述计算机程序产品包括存储在非暂时性计算机可读介质上的计算机可读指令,所述计算机可读指令能够由具有一个或多个处理器的计算机执行,使得在执行所述指令时,所述一个或多个处理器执行本文列出的操作。另选地,所述计算机实现的方法包括使计算机执行这种指令并且执行所得操作的动作。

附图说明

[0016] 结合对附图的参考,本发明的目的、特征以及优点将从本发明的各个方面的以下详细描述变得显而易见,其中:

[0017] 图1是对根据本公开的一些实施方式的用于识别新颖对象的系统的组件进行描绘的框图;

[0018] 图2是根据本公开的一些实施方式的计算机程序产品的例示图;

[0019] 图3是根据本公开的一些实施方式的零样本机器视觉系统的训练阶段的例示图;

[0020] 图4是根据本公开的一些实施方式的、针对先前未见过的输入图像的零样本机器视觉的例示图;

[0021] 图5是根据本公开的一些实施方式的、用于对属于未见过的图像类的输入数据进行分类的系统的处理流程的例示图;以及

[0022] 图6是例示根据本公开的一些实施方式的、使用未见过的实例的标签来控制装置流程图。

具体实施方式

[0023] 本发明涉及用于识别新颖对象的系统,并且更具体地,涉及通过使用联合稀疏表示对视觉特征与语义属性之间的关系进行建模来识别新颖对象的系统。呈现以下描述以使本领域普通技术人员能够作出和使用本发明并将其结合到特定应用的背景中。多种修改以及不同应用中的多种用途对于本领域技术人员来说是显而易见的,并且这里限定的总体构思可以应用于广泛方面。因此,本发明不应限于所呈现的方面,而是涵盖与本文所公开的构思和新颖特征相一致的最广范围。

[0024] 在下面的详细说明中,阐述了许多具体细节,以使得能够更加彻底地理解本发明。然而,本领域技术人员应当明白,本发明可以在不限于这些具体细节的情况下来实施。在其它情况下,公知结构和装置按框图形式示出而不被详细示出,以免模糊本发明。

[0025] 读者应留意与本说明书同时提交的所有文件和文档,这些文件和文档与本说明书一起公开以供公众查阅,所有这些文件和文档的内容通过引用并入于此。本说明书(包括任

何所附权利要求、摘要以及附图)中公开的所有特征可以由用于相同、等同或相似目的的替代特征来代替,除非另有明确说明。因此,除非另有明确说明,否则所公开的各个特征仅仅是典型系列的等同或相似特征的一个示例。

[0026] 此外,权利要求中的未明确陈述用于执行指定功能的“装置”或用于执行特定功能的“步骤”的任何要素不被解释为在35U.S.C.第112节第6款中指定的“装置”或“步骤”条款。具体地,在本文的权利要求中使用“…的步骤”或“…的动作”不旨在援引35U.S.C.第112节第6款的规定。

[0027] 在详细描述本发明之前,首先提供了引用参考文献的列表。接下来,提供了本发明各个主要方面的说明。最后,提供本发明的各个实施方式的具体细节,以给出具体方面的理解。

[0028] (1)所并入的参考文献的列表

[0029] 在本申请中引用和并入以下参考文献。为了清楚和方便起见,这些参考文献在此被列为读者的中心资源。下列参考文献通过引用并入于此,就像在此完全陈述一样。这些参考文献通过参照如下对应文献参考号而在本申请中加以引用:

[0030] 1Akata,Zeynep,Florent Perronnin,Zaid Harchaoui,and Cordelia Schmid."Label-embedding for attribute-based classification" In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition,pp.819-826.2013.

[0031] 2.Romera-Paredes,Bernardino,and P.H.S.Torr"An embarrassingly simple approach to zero-shot learning." In Proceedings of The 32nd International Conference on Machine Learning,pp.2152-2161.2015.

[0032] 3Simonyan,Karen,and Andrew Zisserman."Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556.2014.

[0033] 4Mikolov,T.,Sutskever,I.,Chen,K.,Corrado,G.S.and Dean,J.,20."Distributed representations of words and phrases and their compositionality." In Advances in Neural Information Processing Systems, pp.3111-3119.2013.

[0034] 5.Tibshirani,Ryan J."The lasso problem and uniqueness." Electronic Journal of Statistics.7,pp.1456-1490.2013.

[0035] 6.Huang,Yongwei,and Daniel P.Palomar."Randomized algorithms for optimal solutions of double-sided QCQP with applications in signal processing." IEEE Transactions on Signal Processing,62,no.5,pp.1093-1108.2014.

[0036] 7.Yang,J.,Wright,J.,Huang,T.S.and Ma,Y.,2010."Image super-resolution via sparse representation." IEEE Transactions on Image Processing.19(11),pp.2861-2873.2010.

[0037] 8.Welinder,Peter,Steve Branson,Takeshi Mita,Catherine Wah,Florian Schroff,Serge Belongie,and Pietro Perona."Caltech-UCSD birds200."2010.

[0038] (2)主要方面

[0039] 本发明的各个实施方式包括三个“主要”方面。第一方面是用于识别新颖对象的系

统。该系统通常采用计算机系统操作软件的形式或采用“硬编码”指令集的形式。该系统可以结合到提供不同功能的各种各样的装置中。第二主要方面是使用数据处理系统(计算机)运行的通常采用软件形式的方法。第三主要方面是计算机程序产品。所述计算机程序产品通常表示存储在诸如光学存储装置(例如,光盘(CD)或数字通用盘(DVD))或磁存储装置(诸如,软盘或磁带)的非暂时性计算机可读介质上的计算机可读指令。计算机可读介质的其它非限制性示例包括硬盘、只读存储器(ROM)以及闪存型存储器。这些方面将在下文进行更详细的说明。

[0040] 图1提供了对本发明的系统(即,计算机系统100)的示例进行描绘的框图。计算机系统100被配置成执行与程序或算法相关联的计算、处理、操作和/或功能。在一个方面中,本文讨论的某些处理和步骤被实现为存在于计算机可读存储器单元内并由计算机系统100的一个或更多个处理器执行的一系列指令(例如,软件程序)。在执行时,这些指令使计算机系统100执行特定动作并呈现特定行为,诸如本文所描述的。

[0041] 计算机系统100可以包括被配置成传送信息的地址/数据总线102。另外,一个或更多个数据处理单元(诸如处理器104(或多个处理器))与地址/数据总线102联接。处理器104被配置成处理信息和指令。在一个方面中,处理器104是微处理器。另选地,处理器104可以是不同类型的处理器,诸如并行处理器、专用集成电路(ASIC)、可编程逻辑阵列(PLA)、复杂可编程逻辑器件(CPLD)或现场可编程门阵列(FPGA)。

[0042] 计算机系统100被配置成利用一个或更多个数据存储单元。计算机系统100可以包括与地址/数据总线102联接的易失性存储器单元106(例如,随机存取存储器(“RAM”)、静态RAM、动态RAM等),其中,易失性存储器单元106被配置成存储用于处理器104的信息和指令。计算机系统100还可以包括与地址/数据总线102联接的非易失性存储器单元108(例如,只读存储器(“ROM”)、可编程ROM(“PROM”)、可擦除可编程ROM(“EPROM”)、电可擦除可编程ROM(“EEPROM”)、闪存等),其中,非易失性存储器单元108被配置成存储用于处理器104的静态信息和指令。另选地,计算机系统100可以执行诸如在“云”计算中从在线数据存储单元取回的指令。在一个方面中,计算机系统100还可以包括与地址/数据总线102联接的一个或更多个接口(诸如接口110)。所述一个或更多个接口被配置成使得计算机系统100能够与其它电子装置和计算机系统对接。由所述一个或更多个接口实现的通信接口可以包括有线通信技术(例如,串行电缆、调制解调器、网络适配器等)和/或无线通信技术(例如,无线调制解调器、无线网络适配器等)。

[0043] 在一个方面中,计算机系统100可以包括与地址/数据总线102联接的输入装置112,其中,输入装置112被配置成将信息和命令选择传送至处理器100。根据一个方面,输入装置112是可以包括字母数字键和/或功能键的字母数字输入装置(诸如键盘)。另选地,输入装置112可以是除字母数字输入装置之外的输入装置。在一个方面中,计算机系统100可以包括与地址/数据总线102联接的光标控制装置114,其中,光标控制装置114被配置成将用户输入信息和/或命令选择传送至处理器100。在一个方面中,光标控制装置114利用诸如鼠标、轨迹球、触控板、光学跟踪装置或触摸屏的装置来实现。尽管如此,但在一个方面中,诸如响应于使用与输入装置112相关联的特殊键和键序列命令,光标控制装置114借助于来自输入装置112的输入被引导和/或激活。在另选方面中,光标控制装置114被配置成由语音命令来引导或指导。

[0044] 在一个方面中,计算机系统100还可以包括与地址/数据总线102联接的一个或更多个可选计算机可用数据存储装置(诸如存储装置116)。存储装置116被配置成存储信息和/或计算机可执行指令。在一个方面中,存储装置116是诸如磁或光盘驱动器(例如,硬盘驱动器(“HDD”)、软盘、光盘只读存储器(“CD-ROM”)、数字通用盘(“DVD”))的存储装置。依据一个方面,显示装置118与地址/数据总线102联接,其中,显示装置118被配置成显示视频和/或图形。在一个方面中,显示装置118可以包括阴极射线管(“CRT”)、液晶显示器(“LCD”)、场发射显示器(“FED”)、等离子体显示器或适于显示视频和/或图形图像以及用户可识别的字母数字字符的任何其它显示装置。

[0045] 本文所提出的计算机系统100是根据一个方面的示例计算环境。然而,计算机系统100的非限制性示例并不严格限于是计算机系统。例如,一个方面规定了计算机系统100表示可以根据本文所述各个方面使用的一种数据处理分析。此外,还可以实现其它计算系统。实际上,本技术的精神和范围不限于任何单个数据处理环境。因此,在一个方面中,使用通过计算机执行的计算机可执行指令(诸如程序模块)来控制或实现本技术的各个方面的一个或更多个操作。在一个实现中,这样的程序模块包括被配置成执行特定任务或实现特定抽象数据类型的例程、程序、对象、组件和/或数据结构。另外,一个方面规定了通过利用一个或更多个分布式计算环境来实现本技术的一个或更多个方面,诸如,在分布式计算环境中,由通过通信网络链接的远程处理装置执行任务,或者诸如,在分布式计算环境中,各种程序模块位于包括存储器-存储装置的本地和远程计算机存储介质中。

[0046] 图2描绘了实施本发明的计算机程序产品(即,存储装置)的例示图。计算机程序产品被描绘为软盘200或诸如CD或DVD的光盘202。然而,如先前提到的,计算机程序产品通常表示存储在任何兼容的非暂时性计算机可读介质上的计算机可读指令。如关于本发明所使用的术语“指令”通常指示要在计算机上执行的一组操作,并且可以表示整个程序的片段或者单个可分离的软件模块。“指令”的非限制性示例包括计算机程序代码(源或目标代码)和“硬编码”电子器件(即,编码到计算机芯片中的计算机操作)。“指令”被存储在任何非暂时性计算机可读介质上,诸如存储在计算机的存储器中或软盘、CD-ROM以及闪存驱动器上。在任一种情况下,这些指令被编码在非暂时性计算机可读介质上。

[0047] (3)各个实施方式的具体细节

[0048] 描述了一种“零样本学习”(ZSL)机器视觉系统,所述系统可以识别该系统以前从未见过的对象。在训练阶段,根据本公开的实施方式的系统将包含来自许多感兴趣类的对象的训练图像集作为输入,其中,各个对象图像注释有类标签(例如,“熊”、“汽车”、“房子”)以及若干语义属性(例如,“可以飞”、“有轮子”、“绿色”)两者。

[0049] 然后,本发明使用训练图像集和注释属性来训练将低级图像特征(诸如,边缘、拐点、梯度)映射至语义属性的模型。然后,在测试阶段,系统将不在该训练集中的对象的图像作为输入,并且使用所学习的模型来将测试图像特征映射至语义属性,使得在给出测试对象的语义描述的情况下,尽管测试对象不存在于训练图像集中,仍然可以正确地对测试图像进行分类。本发明使用关于视觉特征词典和语义属性词典的联合稀疏表示来唯一地对视觉特征与语义属性之间的关系进行建模。

[0050] 本文所述的ZSL机器视觉系统结合了联合稀疏词典学习,以对对象的视觉特征与该对象的语义属性之间的关系进行建模。ZSL系统背后的主要假设是:通过从其它信息源

(例如,新类的词描述)或不同传感器模态获得的语义辅助信息来连接训练类和测试类。根据本公开的实施方案的系统通过利用对对象与该对象的语义属性之间的关系进行编码的数学上严密的模型来改进机器视觉系统的零样本能力。这种联合稀疏词典模型显著提高了机器视觉系统的针对不是训练集的一部分的新颖类的识别率。

[0051] 所述系统可以被用于识别相机图像中的新颖对象或场景。相机可以是电的、光学的、红外(IR)的、短波IR的等,以使输出是具有一个或更多个类似颜色通道的强度图像。在ZSL问题中,希望识别图像特征与属性之间的关系。本文所描述的联合词典学习方法自然适合该目的。简单地说,将特征空间和属性空间建模为由低维线性空间的并集表征的非线性空间。这两个非线性空间被约束成具有同源分量(homologous component)(因此得名联合词典),并且这两个非线性空间被建模成为图像特征及其对应属性提供相同表示。

[0052] (3.1) 训练阶段

[0053] 图3描绘了根据本公开的实施方案的零样本机器视觉系统的训练阶段。从深度卷积神经网络中提取图像特征(要素300),同时使用word2vec(其在参考文献No.4中进行了描述)从类的言语描述中提取属性(要素302)。联合词典学习方法(要素304)强制特征的稀疏表示(要素306)及该特征的对应属性的稀疏表示(要素308)是相同的。

[0054] 在训练阶段,存在已知图像集(即,图像空间310)及其对应语义属性(即,标签空间312),以及未见过的图像的语义属性。首先,利用经预训练的卷积神经网络(CNN)(参见参考文献No.3)从已知图像(要素310)中提取特征(要素300)。现有技术已经表明,在从相机图像中提取特征方面,CNN是非常高效的。使 $X = [x_1, \dots, x_N] \in \mathbb{R}^{P \times N}$ 表示从数据集中的图像中提取的特征(要素300),并且使 $Z = [z_1, \dots, z_N] \in \mathbb{R}^{Q \times N}$ 表示所述图像的对应属性(要素302)。术语“属性”是按其最广泛的意义使用的,其涵盖了图像的文字嵌入或任何其它语义信息。此外,使 $Z' = [z'_1, \dots, z'_M] \in \mathbb{R}^{Q \times M}$ 是未见过的图像的属性。

[0055] 将第*i*个图像的标签表示为 $y_i \in \mathbb{R}^K$,其中,相机图像可以潜在地具有*K*个类的多个成员资格(membership)。从属性空间(要素314)到标签空间(要素312)的映射通常被视为线性的($y_i = Wz_i$)并且是已知的。为了进一步阐述这个问题,考虑ZSL中的常见场景,其中,在*X*中包含马和老虎的图像,但*X*不包含斑马的图像。另一方面,语义属性包含所有见过(*Z*)和没见过(*Z'*)的图像(包括斑马)的信息。在这种情况下,语义属性可以是动物的言语定义(例如,“斑马是具有黑色虎状条纹的白色马状动物”)。可以看出,通过从见过的图像(要素310)中学习图像特征(要素300)与属性(要素302) (“马状”和“具有条纹”)之间的关系,应当能够将未见过的斑马图像指派到该斑马图像的对应属性。

[0056] 在本文所描述的发明中,利用*X*(要素300)和 $[Z, Z']$ (要素302)的对应基元(atom)来学习(要素304)两个词典,所述两个词典为 x_i 和 z_i 提供相同的稀疏表示 a_i 。然后,将该构思用于第*j*个未见过的图像,该图像的稀疏表示 b_j 应当接近 z'_j 的稀疏表示。更准确地说,在训练阶段,所述问题可以用公式表示为,

[0057] (1)
$$D_x^*, D_z^*, A^*, B^* = \underset{D_x, D_z, A, B}{\operatorname{argmin}} \|X - D_x A\|_F^2 + \|Z - D_z A\|_F^2 + \|Z' - D_z B\|_F^2 + \lambda(\|A\|_1 + \|B\|_1)$$

[0058] 其中, $\|D_x^{[i]}\|_2, \|D_z^{[i]}\|_2 \leq 1, \quad \forall i \in \{1, \dots, L\}$

[0059] 其中, $D_x \in \mathbb{R}^{P \times L}$ 是图像特征词典, $D_z \in \mathbb{R}^{Q \times L}$ 是属性词典, $A \in \mathbb{R}^{L \times N}$ 是*X*和*Z*的联合稀疏表

示, $B \in \mathbb{R}^{L \times M}$ 是 Z' 的稀疏表示, $D_x^{[i]}$ 和 $D_z^{[i]}$ 是对应矩阵的第 i 列, 并且 λ 是稀疏正则化参数。注意, 在本文所描述的系统中, λ 和 L 是可调参数。式 (1) 在 (D_z, D_x, A, B) 中不是凸的; 在剩余部分中, 该式对各个独立参数而言是凸的。因此, 设计了迭代方案, 以固定其它变量一次求解一个变量, 直到达到收敛为止。所述算法的细节在下文给出。

[0060] (3.1.1) 算法

[0061] 从将 D_x 和 D_z 初始化为具有正态分布的随机矩阵开始。然后, 通过迭代以下步骤来训练变量 (即, 词典 (要素304) 和稀疏表示 (要素306和要素308)):

[0062] 1. 根据下式, 借助于后续Lasso问题 (参见论述Lasso问题的参考文献No.5) 来更新 A :

$$[0063] \quad \operatorname{argmin}_A \frac{1}{2} \left\| \begin{bmatrix} X \\ Z \end{bmatrix} - \begin{bmatrix} D_x \\ D_z \end{bmatrix} A \right\|_F^2 + \lambda \|A\|_1$$

[0064] 2. 根据下式, 借助于后续Lasso问题来更新 B :

$$[0065] \quad \operatorname{argmin}_B \frac{1}{2} \|Z' - D_z B\|_F^2 + \lambda \|B\|_1$$

[0066] 3. 根据下式, 借助于后续二次约束二次规划 (QCQP) 问题 (参见论述QCQP问题的参考文献No.6) 来更新 D_x :

$$[0067] \quad \operatorname{argmin}_{D_x} \frac{1}{2} \|X - D_x A\|_F^2, \quad \text{其中, } \|D_x^{[i]}\|_2 \leq 1$$

[0068] 4. 根据下式, 借助于后续QCQP问题来更新 D_z :

$$[0069] \quad \operatorname{argmin}_{D_z} \frac{1}{2} \|[Z, Z'] - D_z [A, B]\|_F^2, \quad \text{其中, } \|D_z^{[i]}\|_2 \leq 1$$

[0070] 重复上述步骤直到达到收敛为止。在上述算法收敛之后完成训练阶段。

[0071] (3.2) 测试阶段

[0072] 如图4描绘的, 对于先前未见过的输入图像 (要素400), 首先使用深度卷积神经网络提取图像特征 (要素402) 并求解Lasso问题, 以找到图像的稀疏表示 (要素404)。考虑到所述稀疏表示在特征与属性之间共享, 估计 (要素406) 图像属性, 并且找到属性空间 (要素408) 中的最接近的属性。

[0073] 在测试阶段, 对于从未见过的图像 (要素400) 中提取的特征 $x \in \mathbb{R}^p$, 通过求解以下Lasso问题找到该特征的稀疏表示 (要素404):

$$[0074] \quad b^* = \operatorname{argmin}_b \|x - D_x b\|_2^2 + \lambda \|b\|_1.$$

[0075] 然后, 按照此前未见过该输入图像的常见假设, 将对应属性重构如下:

$$[0076] \quad z^* = \operatorname{argmin}_{z \in Z'} \|z - D_z b^*\|_2^2.$$

[0077] 当上述假设不成立时, 这意味着不知道此前是否已经见过所述对象, 使用上式的以下替代:

$$[0078] \quad z^* = \operatorname{argmin}_{z \in \{Z, Z'\}} \|z - D_z b^*\|_2^2.$$

[0079] 最后, 通过将找到的属性 (要素410) 映射至标签空间 $= Wz^*$ (要素412) 来找到未见

过的图像(要素400)的标签。

[0080] 图5是根据本公开的一些实施方式的、用于对属于未见过的图像类的输入数据进行分类的系统的处理流程的例示图。对于先前未见过的输入图像(要素400),使用深度卷积神经网络(要素500)提取图像特征,从而得到CNN特征502。使用视觉词典将视觉特征(CNN特征502)解析成(要素504)共享稀疏表示,从而得到共享稀疏特征(要素506)。使用属性词典对来自共享稀疏表示的属性进行解码(要素508),从而得到预测属性(要素510)。虽然CNN特征502是从深度卷积神经网络(要素500)中提取的,但使用word2vec(要素512)(在参考文献No.4中描述的)从类的言语描述中提取见过的类和未见过的类两者的已知描述(要素514),从而得到来自所有类的属性(要素516)。联合词典学习方法通过匹配过滤(要素518)来强制特征的稀疏表示(来自要素508)及所述特征的对应属性(要素516)的稀疏表示是相同的。

[0081] 匹配过滤模块(要素518)计算预测/解码属性(要素508)与来自所有类的属性(要素516)之间的成对欧几里德距离(Euclidean distance)。匹配过滤(要素518)输出对应于与预测/解码属性间隔最小距离的类标签(要素520)。另外,匹配过滤模块生成确定性输出(要素522),该确定性输出是根据成对的欧几里德距离计算出的概率矢量,以供控制器做出决定。匹配过滤(要素518)的不确定性输出(要素522)是包含用于识别类标签(要素520)的信息的富(rich)表示。

[0082] 然后,确定性输出(要素522)标识预测/解码属性(要素508)与来自所有类的属性(要素516)之间的匹配程度,其中,1是与类完美匹配,并且0是完全不匹配。中间值表示匹配程度。通过匹配过滤(要素518),可以将匹配程度与用户限定阈值进行比较。如果超过阈值,则向诸如机动运载工具组件的装置(要素602)发出命令(例如,制动、转弯)。在该示例中,如果未超出阈值,则使机动运载工具以安全且受控的方式停止。

[0083] 如上所述,通过将图像属性映射至标签空间来确定未见过的输入图像(要素400)的类标签(要素520)。类标签(要素520)同样可以用于控制装置(要素602),下文将对此进行更详细的描述。

[0084] (3.3) 实验研究

[0085] 为了测试根据本公开的实施方式的系统,使用Caltech-UCSD Birds 200 (CUB 200)数据集,该数据集是用于零样本学习的普遍数据集。该数据集包含属于200个类别的、超过6,000个的鸟类图像。该数据集提供312种鸟类特定属性。所述属性包括但不限于“具有实心翅膀图案”、“具有斑点状翅膀图案”、“具有浅黄色”。该数据集通过亚马逊的Mechanical Turk进行注释。

[0086] 在训练中,使用了150个鸟类。使用经预训练的卷积神经网络(CNN)(参见论述CNN的参考文献No.3)从图像中提取特征。所述系统的参数选定如下:

[0087] $Q=312$ (属性维)

[0088] $P=4096$ (图像特征维)

[0089] $L=7000$ (词典基元的数量)

[0090] $K=200$ (类的数量)

[0091] 词典 $D_x \in R^{P \times L}$ 和 $D_z \in R^{Q \times L}$ 最初设定成随机矩阵,并且根据训练数据迭代地学习所述词典。在测试阶段,将属于50个提出的类别(即,类)的测试图像馈送至系统,并且该系统能够以约35%的准确度(该准确度远高于偶然(即,2%))来对这50个类进行分类。

[0092] 应注意,报告结果是初步测试并且可以借助于更细致地调整参数(例如,式(1)中的 λ 和 L)而得到显著改善。另外,使用来自专门训练成检测和识别鸟的CNN的图像特征也可以改善结果。

[0093] 本文所公开的发明直接解决了对自主平台(例如,无人机、无人驾驶飞行器(UAV))和自主运载工具上的鲁棒机器视觉系统的需求。所述系统提供有效适应新颖场景(例如,新颖对象、天气状况)的能力。当前最先进的机器视觉系统需要覆盖了所有可能场景的各种类型的大量训练数据。获得所有视觉类的注释数据实际上是不可能的,因此,迫切需要如下学习系统,即,该学习系统可以传递和调整它们的学习知识以推理新颖类。根据本公开的实施方式联合稀疏视觉特征和语义属性建模系统使得能够利用较少训练示例来执行零样本机器视觉。

[0094] 图6是例示使用处理器600利用未见过的实例的类标签来控制装置602的流程图。可以借助于处理器600和未见过的实例的标签控制的装置602的非限制性示例包括运载工具或运载工具组件(诸如,制动器、转向机构、悬架或安全装置(例如,安全气囊、安全带张紧器等))。此外,运载工具可以是无人驾驶飞行器(UAV)、自主地面车辆或者由驾驶员或远程操作员控制的人工操作车辆。如本领域技术人员可以理解的,其它装置类型的控制也是可以的。

[0095] 本发明提供了将自主运载工具操作扩展至新情况和新道路状况的关键功能。这将使自主运载工具能够在精心绘制和控制的地理区域和状况之外(例如,新交通标志、新区域、新道路类型或者在训练时间内未观察到的新对象类)有效地运行。例如,在生成未见过的实例的标签时,本文所述系统可以将未见过的实例与新交通标志相关联,并使自主运载工具根据新交通标志执行与驾驶参数一致的驾驶操作/操纵。例如,如果标志是停下标志,那么系统可以使自主运载工具应用功能响应(诸如,制动操作),以停下运载工具。其它适当响应可以包括转向操作、加速或减速的油门操作或者保持航向和速度不改变的决定中的一者或更多者。响应可以适合于规避碰撞、提高行进速度或提高效率。

[0096] 另外,所述系统可以嵌入自主机器人运载工具(诸如,UAV和UGV)和自动驾驶(self-driving)运载工具中。例如,在自动驾驶运载工具中,可以使用所述系统来规避碰撞。在该示例中,如果系统在其路径中检测到对象(例如,行人、另一运载工具),则将警报发送给运载工具操作系统以使运载工具执行制动操作。另选地,该警报可以发信号通知运载工具操作系统应当在对象周围执行迂回运动,这涉及到转向操作和加速操作或者提供碰撞规避所需的任何其它操作。此外,检测到的对象可以是道路标志(诸如,停止标志)。在对停止标志进行分类后,可以将警报发送至运载工具操作系统,从而使得运载工具制动或以其它方式遵守由该道路标志传达的消息。

[0097] 最后,虽然已经根据多个实施方式对本发明进行了说明,但本领域普通技术人员应当容易地认识到,本发明可以在其它环境中具有其它应用。应注意,可以有许多实施方式和实现。此外,所附权利要求绝不旨在将本发明的范围限于上述具体实施方式。另外,“用于…的装置”的任何用语旨在引发要素和权利要求的装置加功能的解读,而未特别使用“用于…的装置”用语的任何要素不应被解读为装置加功能要素,即使权利要求以其它方式包括了“装置”一词。此外,虽然已经按特定顺序陈述了特定方法步骤,但这些方法步骤可以按任何期望的顺序进行,并且落入本发明的范围内。

100

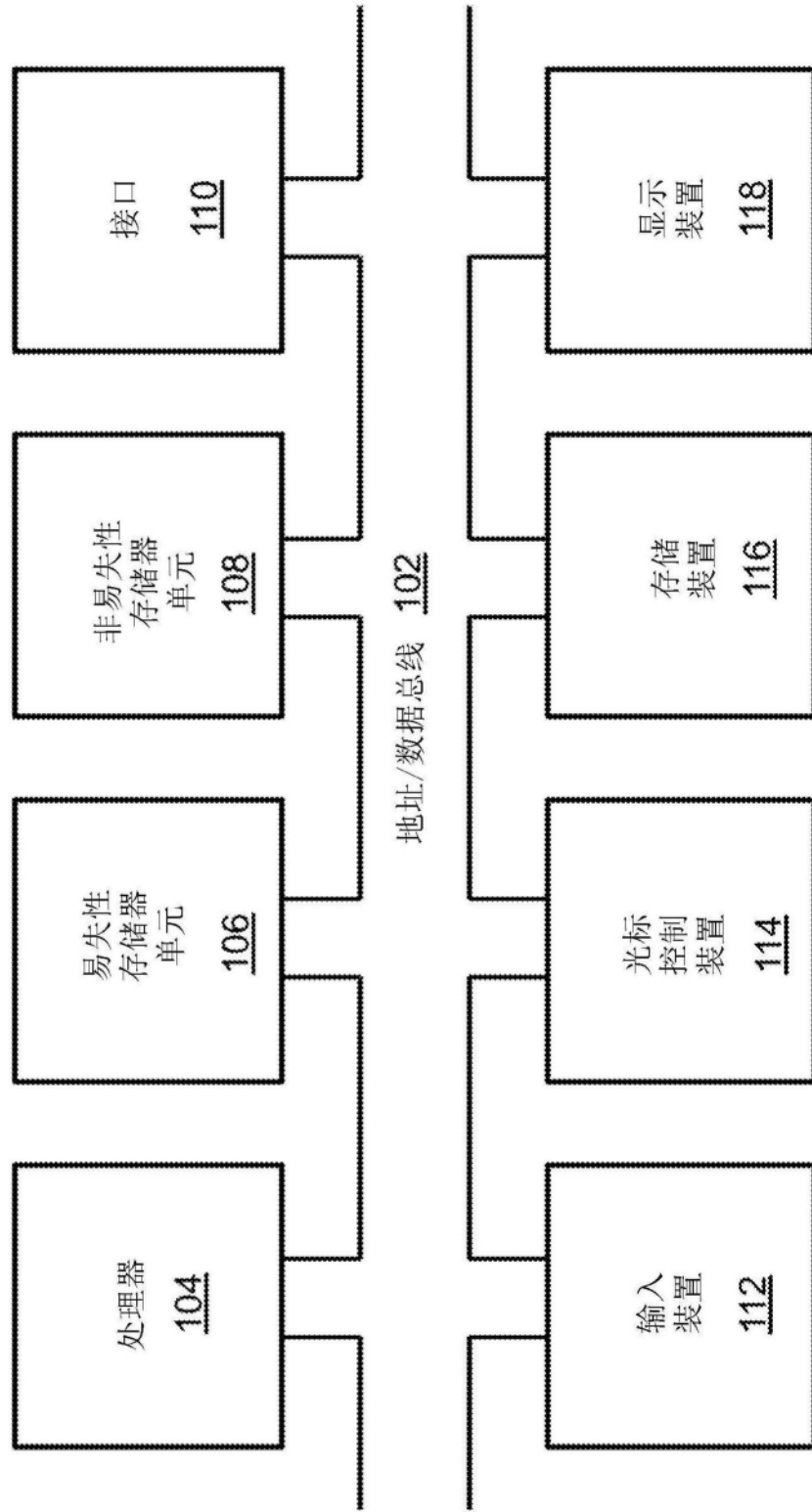


图1

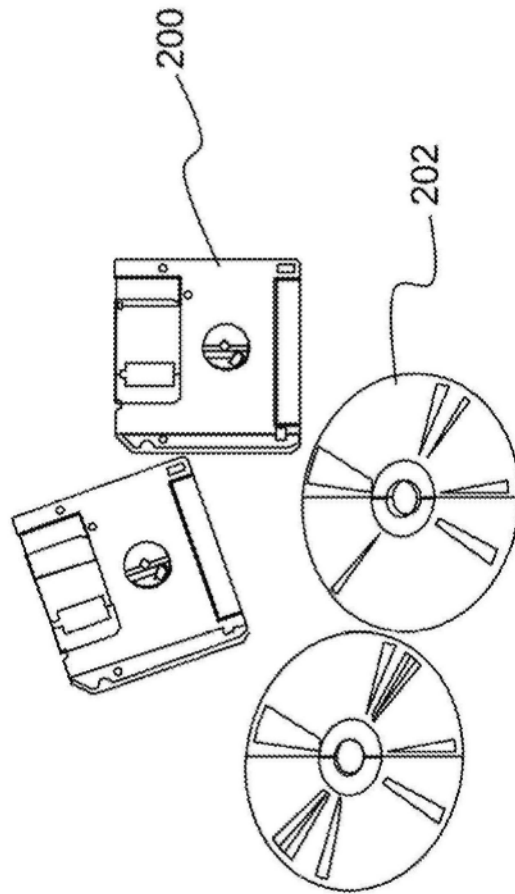


图2

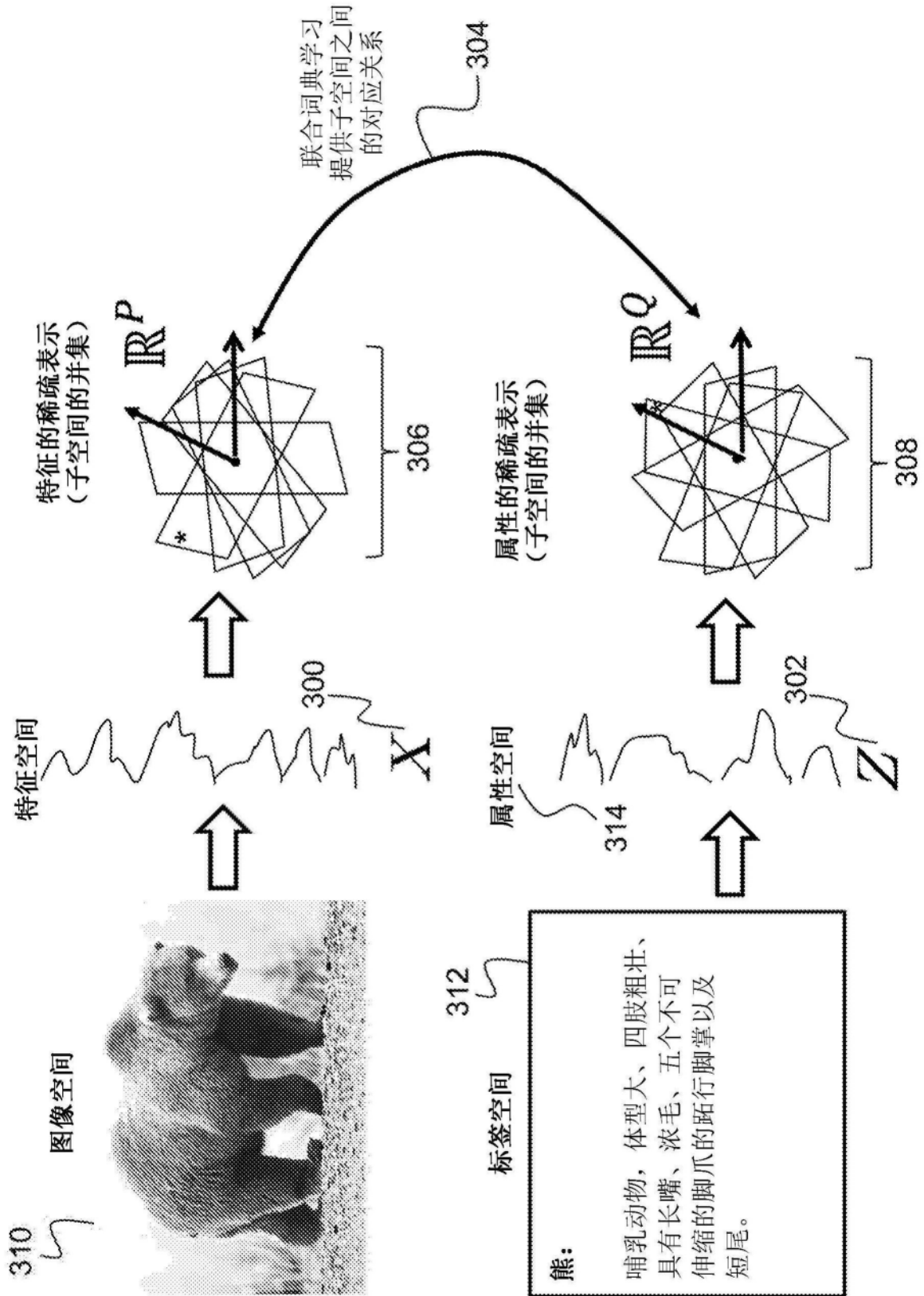


图3

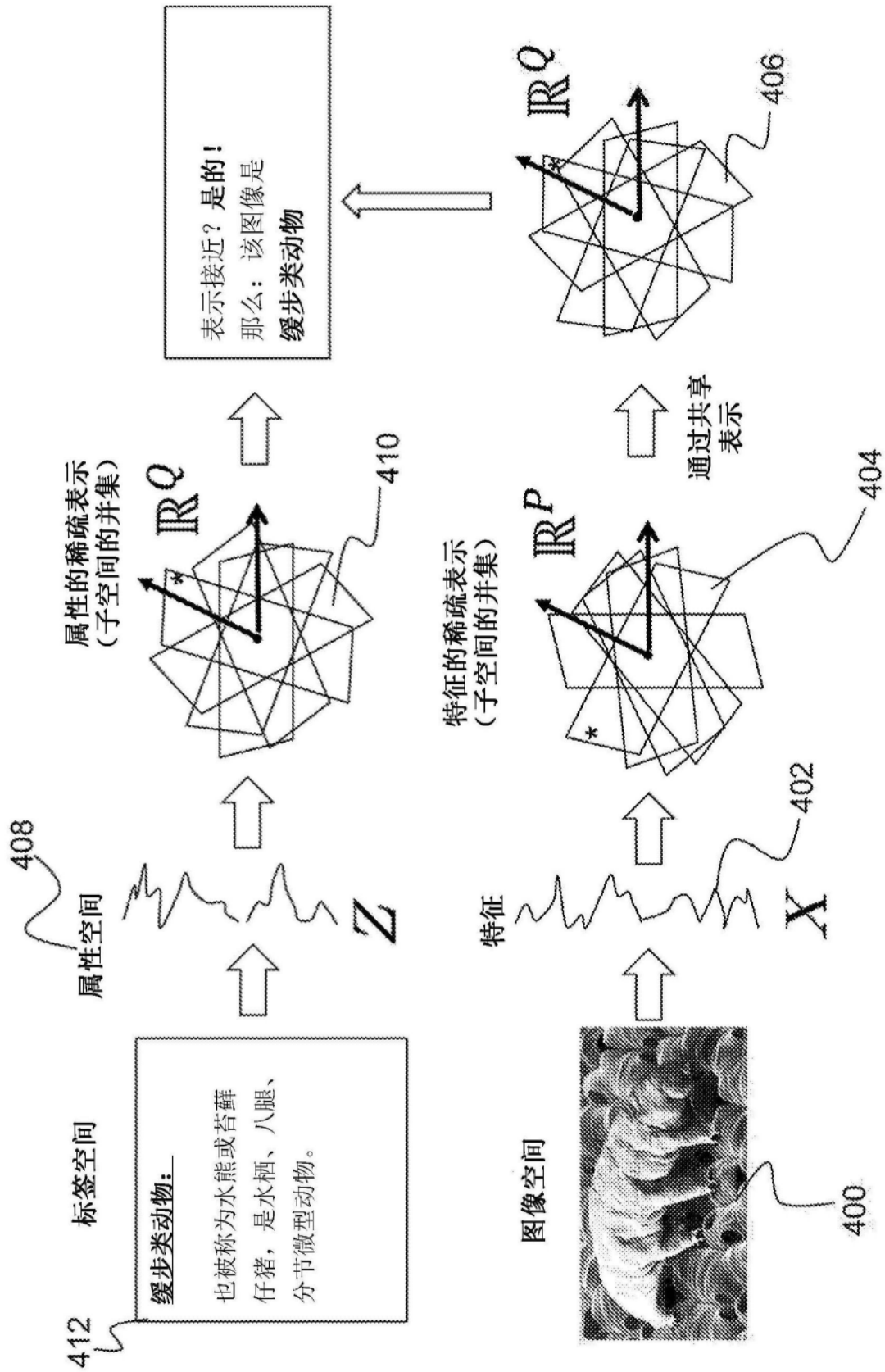


图4

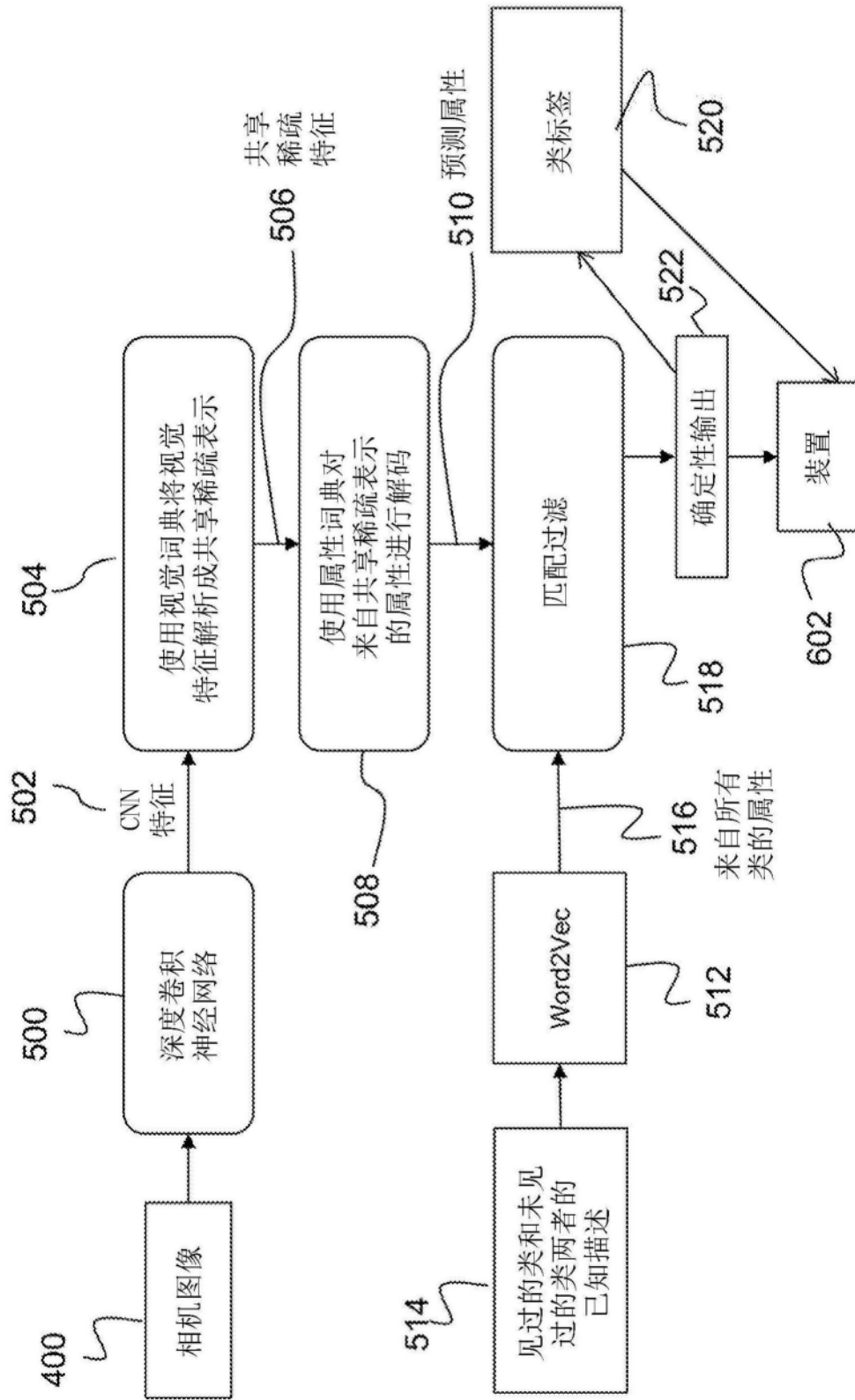


图5

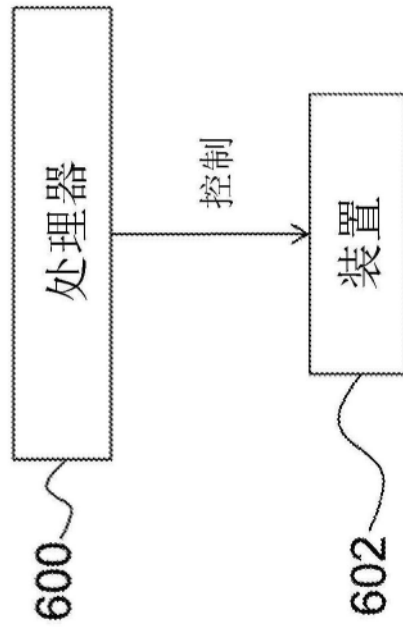


图6