



(12) 发明专利

(10) 授权公告号 CN 111225558 B

(45) 授权公告日 2023.08.11

(21) 申请号 201880064637.3  
 (22) 申请日 2018.08.07  
 (65) 同一申请的已公布的文献号  
 申请公布号 CN 111225558 A  
 (43) 申请公布日 2020.06.02  
 (30) 优先权数据  
 62/542,180 2017.08.07 US  
 62/661,610 2018.04.23 US  
 (85) PCT国际申请进入国家阶段日  
 2020.04.02  
 (86) PCT国际申请的申请数据  
 PCT/US2018/045676 2018.08.07  
 (87) PCT国际申请的公布数据  
 W02019/032622 EN 2019.02.14  
 (73) 专利权人 杰克逊实验室  
 地址 美国缅因州  
 (72) 发明人 V·库玛 B·Q·格瑟 J·皮德森  
 G·丘吉尔  
 (74) 专利代理机构 北京市金杜律师事务所  
 11256  
 专利代理师 陈文平 吕小羽

(51) Int.Cl.  
 A01K 1/03 (2006.01)  
 A01K 11/00 (2006.01)  
 A01K 29/00 (2006.01)  
 G06V 40/20 (2022.01)  
 G06V 20/40 (2022.01)  
 G06V 20/52 (2022.01)  
 G06V 10/82 (2022.01)  
 G06V 10/26 (2022.01)  
 G06V 10/764 (2022.01)  
 G06N 3/0455 (2023.01)  
 G06N 3/09 (2023.01)

(56) 对比文件  
 FR 2948738 A1, 2011.02.04  
 CA 2593741 A1, 2008.01.13  
 US 2004105264 A1, 2004.06.03  
 CN 201097536 Y, 2008.08.13  
 US 2003105412 A1, 2003.06.05  
 US 2016150758 A1, 2016.06.02  
 US 2016232440 A1, 2016.08.11  
 US 2010324861 A1, 2010.12.23 (续)

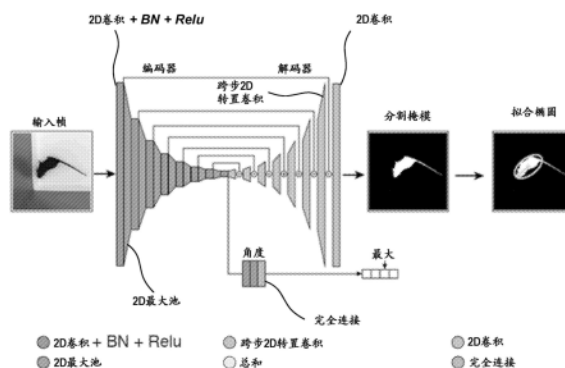
审查员 程婷婷

权利要求书3页 说明书31页 附图27页

(54) 发明名称  
 长期且连续的动物行为监测

(57) 摘要  
 本发明提供了用于连续监测动物如小型啮齿动物的行为的系统和方法。监测可包括视频、音频和其他传感器模式。在一个实施方案中，该系统可包括相机、活动场所设计、环境传感器和超声传感器。该系统独特地提供了适于小鼠行为研究的连续长期监测系统。还提供了基于神经网络的跟踪器，该跟踪器配置为与由监测系统获取的视频数据一起使用。测试了3种不同的神经网络架构以确定它们在不同环境条件下对遗传多样性小鼠的影响。已观察到，编码器-解码器分割神经网络以最少的训练数据实现了高的精度和

速度。通过迁移学习，这种通用的神经网络跟踪器可容易地扩展到其他实验范例，并且甚至可扩展到其他动物，从而为生物行为研究形成一个稳健的、可泛化的解决方案。



CN 111225558 B

[接上页]

**(56) 对比文件**

CN 101311947 A, 2008.11.26

US 2012150651 A1, 2012.06.14

US 2009122058 A1, 2009.05.14

CN 102077268 A, 2011.05.25

吉益民.《网络变异语言现象的认知研究》.  
南京师范大学出版社, 2012, (第1版), 119.

1. 一种用于在复杂环境中进行长期动物跟踪的方法,所述方法包括:
  - 由处理器接收表示对至少一只动物的观察的视频数据;
  - 由所述处理器确定来自于所述视频数据的输入视频帧;
  - 由所述处理器采用神经网络架构处理所述输入视频帧,其中所述神经网络架构包括卷积神经网络;
  - 基于采用所述神经网络架构对所述输入视频帧的处理,由所述处理器确定所述至少一只动物的椭圆描述,所述椭圆描述由预定的椭圆参数定义;和
  - 由所述处理器确定数据,所述数据包括对应于所述至少一只动物的所述椭圆描述的所述椭圆参数的一个或多个值。
2. 根据权利要求1所述的方法,其中所述椭圆参数为表示所述动物在平面内的位置的坐标、所述动物的长轴长度和短轴长度、以及所述动物的头朝向的角度,所述角度相对于所述长轴的方向定义。
3. 根据权利要求1所述的方法,其中所述神经网络架构为编码器-解码器分割网络,所述编码器-解码器分割网络配置为:
  - 从输入视频帧预测前景-背景分割图像;
  - 基于所述分割图像逐像素地预测在所述输入视频帧中是否存在动物;
  - 基于所述逐像素的预测输出分割掩模;和
  - 将其中预测存在所述动物的所述分割掩模的部分拟合成椭圆以确定表征所述预定的椭圆参数的值。
4. 根据权利要求3所述的方法,其中所述编码器-解码器分割网络包括:
  - 特征编码器,所述特征编码器配置为将所述输入视频帧抽象为小空间分辨率特征集合;
  - 特征解码器,所述特征解码器配置为将所述特征集合转换为与所述输入视频帧相同的形状并输出所述前景-背景分割图像;和
  - 角度预测器,所述角度预测器配置为预测所述动物的头朝向的角度。
5. 根据权利要求1所述的方法,其中所述神经网络架构包括分箱分类网络,所述分箱分类网络配置为预测所述椭圆描述的每个椭圆参数的最概然值的热图。
6. 根据权利要求5所述的方法,其中所述分箱分类网络包括特征编码器,所述特征编码器配置为将所述输入视频帧抽象为小空间分辨率,其中采用抽象化来生成所述热图。
7. 根据权利要求1所述的方法,其中所述神经网络架构包括回归网络,所述回归网络配置为从输入视频帧提取特征并直接预测表征每个椭圆参数的值。
8. 根据权利要求1所述的方法,其中所述动物为啮齿动物。
9. 根据权利要求1所述的方法,其中所述动物为啮齿动物,并且其中所述神经网络架构配置为识别不同颜色的啮齿动物。
10. 根据权利要求1所述的方法,其中所述神经网络架构配置为识别不同体形的动物。
11. 一种用于在复杂环境中进行长期动物跟踪的系统,所述系统包括:
  - 数据存储设备,所述数据存储设备配置为存储表示对至少一只动物的观察的视频数据;
  - 处理器,所述处理器配置为从所述数据存储设备接收所述视频数据并配置为:

确定来自于所述视频数据的输入视频帧；  
采用神经网络架构处理所述输入视频帧，其中所述神经网络架构包括卷积神经网络；  
基于采用所述神经网络架构对所述输入视频帧的处理，确定所述至少一只动物的椭圆描述，所述椭圆描述由预定的椭圆参数定义；和  
确定数据，所述数据包括对应于所述至少一只动物的所述椭圆描述的所述椭圆参数的一个或多个值。

12. 根据权利要求11所述的系统，其中所述椭圆参数为表示所述动物在平面内的位置的坐标、所述动物的长轴长度和短轴长度、以及所述动物的头朝向的角度，所述角度相对于所述长轴的方向定义。

13. 根据权利要求11所述的系统，其中所述神经网络架构为编码器-解码器分割网络，所述编码器-解码器分割网络配置为：

从输入视频帧预测前景-背景分割图像；  
基于所述分割图像逐像素地预测在所述输入视频帧中是否存在动物；  
基于所述逐像素的预测输出分割掩模；和  
将其中预测存在所述动物的所述分割掩模的部分拟合成椭圆以确定表征所述预定的椭圆参数的值。

14. 根据权利要求13所述的系统，其中所述编码器-解码器分割网络包括：  
特征编码器，所述特征编码器配置为将所述输入视频帧抽象为小空间分辨率特征集合；

特征解码器，所述特征解码器配置为将所述特征集合转换为与所述输入视频帧相同的形状并输出所述前景-背景分割图像；和

角度预测器，所述角度预测器配置为预测所述动物的头朝向的角度。

15. 根据权利要求11所述的系统，其中所述神经网络架构包括分箱分类网络，所述分箱分类网络配置为预测所述椭圆描述的每个椭圆参数的最概然值的热图。

16. 根据权利要求15所述的系统，其中所述分箱分类网络包括特征编码器，所述特征编码器配置为将所述输入视频帧抽象为小空间分辨率，其中采用抽象化来生成所述热图。

17. 根据权利要求11所述的系统，其中所述神经网络架构包括回归网络，所述回归网络配置为从输入视频帧提取特征并直接预测每个椭圆参数的一个或多个值。

18. 根据权利要求11所述的系统，其中所述动物为啮齿动物。

19. 一种存储指令的计算机可读存储介质，所述指令在由至少一个计算系统的至少一个数据处理器执行时实施一种在复杂环境中进行长期动物跟踪的方法，所述方法包括：

接收表示对至少一只动物的观察的视频数据；  
确定来自于所述视频数据的输入视频帧；  
采用神经网络架构处理所述输入视频帧，其中所述神经网络架构包括卷积神经网络；  
基于采用所述神经网络架构对所述输入视频帧的处理，确定所述至少一只动物的椭圆描述，所述椭圆描述由预定的椭圆参数定义；和  
确定数据，所述数据包括对应于所述至少一只动物的所述椭圆描述的所述椭圆参数的一个或多个值。

20. 根据权利要求19所述的计算机可读存储介质，其中所述椭圆参数为表示所述动物

在平面内的位置的坐标、所述动物的长轴长度和短轴长度、以及所述动物的头朝向的角度，所述角度相对于所述长轴的方向定义。

21. 根据权利要求19所述的计算机可读存储介质，其中所述神经网络架构为编码器-解码器分割网络，所述编码器-解码器分割网络配置为：

从输入视频帧预测前景-背景分割图像；

基于所述分割图像逐像素地预测在所述输入视频帧中是否存在动物；

基于所述逐像素的预测输出分割掩模；和

将其中预测存在所述动物的所述分割掩模的部分拟合成椭圆以确定表征所述预定的椭圆参数的值。

22. 根据权利要求21所述的计算机可读存储介质，其中所述编码器-解码器分割网络包括：

特征编码器，所述特征编码器配置为将所述输入视频帧抽象为小空间分辨率特征集合；

特征解码器，所述特征解码器配置为将所述特征集合转换为与所述输入视频帧相同的形状并输出所述前景-背景分割图像；和

角度预测器，所述角度预测器配置为预测所述动物的头朝向的角度。

23. 根据权利要求19所述的计算机可读存储介质，其中所述神经网络架构包括分箱分类网络，所述分箱分类网络配置为预测所述椭圆描述的每个椭圆参数的最概然值的热图。

24. 根据权利要求23所述的计算机可读存储介质，其中所述分箱分类网络包括特征编码器，所述特征编码器配置为将所述输入视频帧抽象为小空间分辨率，其中采用抽象化来生成所述热图。

25. 根据权利要求19所述的计算机可读存储介质，其中所述神经网络架构包括回归网络，所述回归网络配置为从输入视频帧提取特征并直接预测表征每个椭圆参数的值。

26. 根据权利要求19所述的计算机可读存储介质，其中所述动物为啮齿动物。

## 长期且连续的动物行为监测

### [0001] 相关申请的交叉引用

[0002] 本申请要求2017年8月7日提交的标题为“Long-Term and Continuous Animal Behavioral Monitoring”的美国临时申请号62/542,180和2018年4月23日提交的标题为“Robust Mouse Tracking In Complex Environments Using Neural Networks”的美国临时申请号62/661,610的权益。这些申请中的每一者通过引用全文并入本文。

### 背景技术

[0003] 动物行为可理解为神经系统响应于内部或外部刺激的输出。准确跟踪动物的能力作为对动物行为进行分类的过程的一部分可能会很有价值。例如,行为的变化是衰老、精神障碍或甚至代谢性疾病的标志,并可揭示关于对动物生理、神经认知和情绪状态的影响的重要信息。

### 发明内容

[0004] 传统上,评估动物行为的实验是侵入性地进行的,其中研究人员直接与动物互动。例如,研究人员可从居住环境(例如,笼子)移出动物如小鼠并将该动物转移到不同的环境(例如,迷宫或其他装置)中。然后,在执行任务时,研究人员可能会让他们自己待在新环境附近并跟踪动物以观察它们。然而,已认识到动物在新环境中或在进行试验的实验人员面前的行为可能有所不同。这常会导致数据的混乱并造成无法再现和误导性的结果。

[0005] 已开发出较少侵入性的监测技术以将行为监测实验期间的人为干扰最小化。例如,已探索了视频监控用于监测动物行为。然而,视频监控存在挑战。一方面,在很长一段时间内并在广泛的环境条件下连续地捕获时空分辨率高的视频数据的能力仍然是一个关键的障碍。在长时间段如几天、几周和/或几月内进行的动物观察研究可能会产生大量的数据,这些数据的获取和存储成本很高。另一方面,即使假设具有获取和存储足够质量的视频数据的能力,但研究人员手工查阅在如此长的时间段内长时间观察和跟踪动物期间产生的大量视频镜头在经济上也不可行。当所观察的动物数量增加(如在筛选新药或进行基因组学实验时可能是必需的)时,这一挑战会被放大。

[0006] 为了解决这一问题,已开发了基于计算机的技术来分析捕获的动物行为视频。然而,现有的基于计算机的系统缺乏在复杂且动态的环境中准确跟踪不同动物的能力。在一个实例中,现有的用于动物跟踪的基于计算机的技术可能无法准确地将单个动物与背景(例如,笼壁和/或地板、笼子内的物体如水碗)或多个动物彼此区分开。在最好的情况下,如果在观察期内未准确跟踪给定的动物,则可能失去有价值的观察数据。在最坏的情况下,如果在观察期内未正确跟踪给定的动物或该动物的一部分并将其误认为是另一只动物,则会向根据获取的视频数据分类的行为中引入误差。虽然已采用了诸如改变动物的毛色等技术来方便跟踪,但改变动物的毛色可能改变其行为。结果,在复杂且动态的环境中或与遗传异质动物一起进行的现有视频跟踪方法需要用户的高度参与,从而抵消了视频观察的上述优势。因此,大规模和/或长期的动物监测实验仍然不可行。

[0007] 随着神经科学和行为进入到大行为数据和计算行为学的时代,动物跟踪需要更好的技术以方便对动物长时间段内在半自然且动态的环境中的行为分类。

[0008] 相应地,已开发了采用神经网络的系统和方法,其能够提供对旷场中动物(例如,小鼠)稳健且可扩容的跟踪。作为一个实例,提供了便于以高的时空分辨率采集动物运动的视频数据的系统和方法。可在很长一段时间内并在广泛的环境条件下以连续的方式捕获此视频数据。

[0009] 获取的视频数据可用作卷积(convolutional)神经网络架构的输入以进行跟踪。可对神经网络加以训练,使得在面临新的环境或动物时,当经过训练时,该神经网络高度稳健并且能够在多种实验条件下跟踪而无需用户参与的调整。此类实验条件的实例可包括不同品系的小鼠,而与各种毛色、体形和行为以及不同的笼养环境无关。因此,本公开的实施方案可便于极小侵入性的动物跟踪,以便于在完全不同的条件下长时间地对大量动物进行行为监测。

[0010] 在某些实施方案中,可组合使用所公开的视频观察和动物跟踪技术。然而,可以理解,这些技术中的每一者可单独使用或者以彼此或与其他技术的任何组合使用。

[0011] 在一个实施方案中,提供了一种用于动物跟踪的方法。所述方法可包括:由处理器接收表示对动物的观察的视频数据,和由处理器执行神经网络架构。神经网络架构可配置为接收从视频数据提取的输入视频帧,基于输入视频帧生成至少一只动物的椭圆描述(ellipse description)(该椭圆描述由预定的椭圆参数定义),和提供数据,包括表征所述至少一只动物的预定的椭圆参数的值。

[0012] 在所述方法的另一个实施方案中,椭圆参数可以是表示动物在平面内的位置的坐标、动物的长轴长度和短轴长度、以及动物的头朝向的角度,该角度相对于长轴的方向定义。

[0013] 在所述方法的另一个实施方案中,神经网络架构可以是编码器-解码器分割网络(encoder-decoder segmentation network)。编码器-解码器分割网络可配置为从输入视频帧预测前景-背景分割图像,基于分割图像逐像素地预测在输入视频帧中是否存在动物,基于逐像素的预测输出分割掩模,和将其中预测存在动物的分割掩模的部分拟合成椭圆以确定表征预定的椭圆参数的值。

[0014] 在所述方法的另一个实施方案中,编码器-解码器分割网络可包括特征编码器、特征解码器和角度预测器。特征编码器可配置为将输入视频帧抽象为小空间分辨率特征集合。特征解码器可配置为将特征集合转换为与输入视频帧相同的形状并输出前景-背景分割图像。角度预测器可配置为预测动物的头朝向的角度。

[0015] 在所述方法的另一个实施方案中,神经网络架构可包括分箱(binned)分类网络,该分箱分类网络配置为预测椭圆描述的每个椭圆参数的最概然值的热图。

[0016] 在所述方法的另一个实施方案中,分箱分类网络可包括特征编码器,该特征编码器配置为将输入视频帧抽象为小空间分辨率,其中采用抽象化来生成热图。

[0017] 在所述方法的另一个实施方案中,神经网络架构可包括回归网络,该回归网络配置为从输入视频帧提取特征并直接预测表征每个椭圆参数的值。

[0018] 在所述方法的另一个实施方案中,动物可以是啮齿动物。

[0019] 在一个实施方案中,提供了一种用于动物跟踪的系统。所述系统可包括数据存储

设备,该数据存储设备保持表示对动物的观察的视频数据。所述系统还可包括处理器,该处理器配置为从数据存储设备接收视频数据并实施神经网络架构。神经网络架构可配置为接收从视频数据提取的输入视频帧,基于视频帧生成至少一只动物的椭圆描述(该椭圆描述由预定的椭圆参数定义),和提供数据,包括表征所述至少一只动物的预定的椭圆参数的值。

[0020] 在所述系统的另一个实施方案中,椭圆参数可以是表示动物在平面内的位置的坐标,动物的长轴长度和短轴长度,以及动物的头朝向的角度,该角度相对于长轴的方向定义。

[0021] 在所述系统的另一个实施方案中,神经网络架构可以是编码器-解码器分割网络。编码器-解码器分割网络可配置为从输入视频帧预测前景-背景分割图像,基于分割图像逐像素地预测在输入视频帧中是否存在动物,基于逐像素的预测输出分割掩模,并将其中预测存在动物的分割掩模的部分拟合成椭圆以确定表征预定的椭圆参数的值。

[0022] 在所述系统的另一个实施方案中,编码器-解码器分割网络可包括特征编码器、特征解码器和角度预测器。特征编码器可配置为将输入视频帧抽象为小空间分辨率特征集合。特征解码器可配置为将特征集合转换为与输入视频帧相同的形状并输出前景-背景分割图像。角度预测器可配置为预测动物的头朝向的角度。

[0023] 在所述系统的另一个实施方案中,神经网络架构可包括分箱分类网络。该分箱分类网络可配置为预测椭圆描述的每个椭圆参数的最概然值的热图。

[0024] 在所述系统的另一个实施方案中,分箱分类网络可包括特征编码器,该特征编码器配置为将输入视频帧抽象为小空间分辨率,其中采用抽象化来生成热图。

[0025] 在所述系统的另一个实施方案中,神经网络架构可包括回归网络,该回归网络配置为从输入视频帧提取特征并直接预测表征每个椭圆参数的值。

[0026] 在所述系统的另一个实施方案中,动物可以是啮齿动物。

[0027] 在一个实施方案中,提供了一种存储指令的非暂态计算机程序产品。所述指令在由至少一个计算系统的至少一个数据处理器执行时可实施一种方法,该方法包括:接收表示对动物的观察的视频数据,和执行神经网络架构。神经网络架构可配置为接收从视频数据提取的输入视频帧,基于输入视频帧生成至少一只动物的椭圆描述(该椭圆描述由预定的椭圆参数定义),和提供数据,包括表征所述至少一只动物的预定的椭圆参数的值。

[0028] 在另一个实施方案中,椭圆参数可以是表示动物在平面内的位置的坐标,动物的长轴长度和短轴长度,以及动物的头朝向的角度,该角度相对于长轴的方向定义。

[0029] 在另一个实施方案中,神经网络架构可以是编码器-解码器分割网络。编码器-解码器分割网络可配置为从输入视频帧预测前景-背景分割图像,基于分割图像逐像素地预测在输入视频帧中是否存在动物,基于逐像素的预测输出分割掩模,并将其中预测存在动物的分割掩模的部分拟合成椭圆以确定表征预定的椭圆参数的值。

[0030] 在另一个实施方案中,编码器-解码器分割网络可包括特征编码器、特征解码器和角度预测器。特征编码器可配置为将输入视频帧抽象为小空间分辨率特征集合。特征解码器可配置为将特征集合转换为与输入视频帧相同的形状并输出前景-背景分割图像。角度预测器可配置为预测动物的头朝向的角度。

[0031] 在另一个实施方案中,神经网络架构可包括分箱分类网络,该分箱分类网络配置

为预测椭圆描述的每个椭圆参数的最概然值的热图。

[0032] 在另一个实施方案中,分箱分类网络可包括特征编码器,该特征编码器配置为将输入视频帧抽象为小空间分辨率,其中可采用抽象化来生成热图。

[0033] 在另一个实施方案中,神经网络架构可包括回归网络,该回归网络配置为从输入视频帧提取特征并直接预测表征每个椭圆参数的值。

[0034] 在另一个实施方案中,动物可以是啮齿动物。

[0035] 在一个实施方案中,提供了一种系统并且其可包括活动场所和采集系统。活动场所可包括框架和安装到框架的围栏。围栏尺寸可定为容纳动物并且其可包括配置为允许进入到围栏内部的门。采集系统可包括相机、至少两个光源组、控制器和数据存储设备。两个光源组中的每一个可配置为发出波长彼此不同的光并发出入射到围栏上的光。相机可配置为当由所述光源组中的至少一个照射时获取围栏的至少一部分的视频数据。控制器可与相机和光源组电通信。控制器可配置为生成控制信号,该控制信号可用来控制相机的视频数据采集和光源组的光发射并接收相机获取的视频数据。数据存储设备可与控制器电通信并且其可配置为存储从控制器接收的视频数据。

[0036] 在所述系统的另一个实施方案中,围栏的至少一部分对于可见光可以是大致不透明的。

[0037] 在所述系统的另一个实施方案中,围栏的至少一部分可由对可见光波长大致不透明的材料形成。

[0038] 在所述系统的另一个实施方案中,围栏的至少一部分可由对红外光波长大致不反射的材料形成。

[0039] 在所述系统的另一个实施方案中,围栏的至少一部分可由聚氯乙烯(PVC)或聚甲醛(POM)的板材形成。

[0040] 在所述系统的另一个实施方案中,第一光源组可包括一个或多个配置为以一个或多个可见光波长发射光的第一灯,第二光源组可包括一个或多个配置为以一个或多个红外(IR)光波长发射光的第二灯。

[0041] 在所述系统的另一个实施方案中,红外光的波长可为大约940nm。

[0042] 在所述系统的另一个实施方案中,相机可配置为以至少480×480像素的分辨率获取视频数据。

[0043] 在所述系统的另一个实施方案中,相机可配置为以大于小鼠移动频率的帧速率获取视频数据。

[0044] 在所述系统的另一个实施方案中,相机可配置为以至少29帧每秒(fps)的帧速率获取视频数据。

[0045] 在所述系统的另一个实施方案中,相机可配置为获取具有至少8位深度(bit depth)的视频数据。

[0046] 在所述系统的另一个实施方案中,相机可配置为在红外波长下获取视频数据。

[0047] 在所述系统的另一个实施方案中,控制器可配置为压缩从相机接收的视频数据。

[0048] 在所述系统的另一个实施方案中,控制器可配置为使用具有采用基于方差的背景减除的滤波器的MPEG4编解码器来压缩从相机接收的视频数据。

[0049] 在所述系统的另一个实施方案中,MPEG编解码器滤波器可为Q0 HQDN3D。

[0050] 在所述系统的另一个实施方案中,控制器可配置为根据模拟明/暗周期的时间表来命令第一光源照射围栏。

[0051] 在所述系统的另一个实施方案中,控制器可配置为命令第一光源在明/暗周期的明部分期间用具有约50勒克斯(lux)至约800勒克斯的强度的可见光照射围栏。

[0052] 在所述系统的另一个实施方案中,控制器可配置为命令第二光源用红外光照射围栏,使得围栏的温度在红外照射下升高不到5℃。

[0053] 在所述系统的另一个实施方案中,控制器可配置为命令第一光源根据对数标度的1024个亮度级来照射围栏。

[0054] 在一个实施方案中,提供了一种方法,该方法可包括由至少一个光源组照射配置为容纳动物的围栏。每个光源组可配置为发射波长彼此不同的光。该方法还可包括由相机获取被所述光源组中的至少一个照射的围栏的至少一部分的视频数据。该方法可另外包括由与相机和光源组电通信的控制器生成可用来控制相机对视频数据的采集以及光源组的光发射的控制信号。该方法可还包括由控制器接收由相机获取的视频数据。

[0055] 在所述方法的另一个实施方案中,围栏的至少一部分对于可见光可以是大致不透明的。

[0056] 在所述方法的另一个实施方案中,围栏的至少一部分可由对可见光波长大致不透明的材料形成。

[0057] 在所述方法的另一个实施方案中,围栏的至少一部分可由对红外光波长大致不反射的材料形成。

[0058] 在所述方法的另一个实施方案中,围栏的至少一部分可由聚氯乙烯(PVC)或聚甲醛(POM)的板材形成。

[0059] 在所述方法的另一个实施方案中,第一光源组可包括一个或多个配置为以一个或多个可见光波长发射光的第一灯,第二光源组可包括一个或多个配置为以一个或多个红外(IR)光波长发射光的第二灯。

[0060] 在所述方法的另一个实施方案中,红外光的波长可为大约940nm。

[0061] 在所述方法的另一个实施方案中,相机可配置为以至少480×480像素的分辨率获取视频数据。

[0062] 在所述方法的另一个实施方案中,相机可配置为以大于小鼠移动频率的帧速率获取视频数据。

[0063] 在所述方法的另一个实施方案中,相机可配置为以至少29帧每秒(fps)的帧速率获取视频数据。

[0064] 在所述方法的另一个实施方案中,相机可配置为获取具有至少8位深度的视频数据。

[0065] 在所述方法的另一个实施方案中,相机可配置为在红外波长下获取视频数据。

[0066] 在所述方法的另一个实施方案中,控制器可配置为压缩从相机接收的视频数据。

[0067] 在所述方法的另一个实施方案中,控制器可配置为使用具有采用基于方差的背景减除的滤波器的MPEG4编解码器来压缩从相机接收的视频数据。

[0068] 在所述方法的另一个实施方案中,MPEG编解码器滤波器可为Q0 HQDN3D。

[0069] 在所述方法的另一个实施方案中,控制器可配置为根据模拟明/暗周期的时间表

来命令第一光源照射围栏。

[0070] 在所述方法的另一个实施方案中,控制器可配置为命令第一光源在明/暗周期的明部分期间用具有约50勒克斯至约800勒克斯的强度的可见光照射围栏。

[0071] 在所述方法的另一个实施方案中,控制器可配置为命令第二光源用红外光照射围栏,使得围栏的温度在红外照射下升高不到5℃。

[0072] 在所述方法的另一个实施方案中,控制器可配置为命令第一光源根据对数标度的1024个照明等级来照射围栏。

## 附图说明

[0073] 通过以下结合附图的详细描述,将更容易理解这些及其他特征,在附图中:

[0074] 图1为流程图,示意了用于动物跟踪的操作环境的一个示例性实施方案。

[0075] 图2为用于动物行为监测的系统的实施方案的示意图。

[0076] 图3A-3F为示意由图2的系统获取的样本帧的图像;(A-C)可见光;(D-F)红外(IR)光。

[0077] 图4A-4B为两种相机模型的量子效率与波长的关系图;(A) Sentech STC-MC33USB的相对响应;(B) Basler acA1300-60gm-NIR的量子效率。

[0078] 图5为IR长通滤波器的透明度-波长分布图。

[0079] 图6A-6D为示意经受不同压缩技术的视频帧的示例性实施方案的图像;(A) 无压缩;(B) MPEG4 Q0, (C) MPEG4 Q5; (D) MPEG4 Q0 HQDN3D。

[0080] 图7为示意适合与图2的系统一起使用的采集系统的部件的实施方案的图。

[0081] 图8A为根据本公开分析的观察环境的示例性实施方案的示意图,包括黑小鼠、灰小鼠、白化小鼠和花斑小鼠。

[0082] 图8B为造成动物跟踪不力的情况的示意图。

[0083] 图8C为包括椭圆形式的目标跟踪的小鼠跟踪的一个示例性实施方案的示意图。

[0084] 图9为分割网络架构的一个示例性实施方案的示意图。

[0085] 图10为分箱分类网络架构的一个示例性实施方案的示意图。

[0086] 图11为回归分类网络架构的一个示例性实施方案的示意图。

[0087] 图12A为图形用户界面的一个示例性实施方案,示意了用于前景(F)和背景(B)的两个标记的放置。

[0088] 图12B为图形用户界面的一个示例性实施方案,示意了由图12A的标记产生的分割。

[0089] 图13A为图9-11的分割、回归和分箱分类网络的实施方案的训练曲线图。

[0090] 图13B为图9-11的分割、回归和分箱分类网络的实施方案的验证(validation)曲线图。

[0091] 图13C的图示意了图9的分割网络架构的训练和验证性能。

[0092] 图13D的图示意了图11的回归网络架构的训练和验证性能。

[0093] 图13E的图示意了图10的分箱分类网络架构的训练和验证性能。

[0094] 图14A的图示意了根据本公开的实施方案,不同大小的训练集合(training set)的训练误差与步长(step)的关系。

[0095] 图14B的图示意了根据本公开的实施方案,不同大小的训练集合的验证误差与步长的关系。

[0096] 图14C的图示意了对于训练样本的完整训练集合,训练误差和验证误差与步长的关系。

[0097] 图14D的图示意了对于包含10,000(10k)个训练样本的训练集合,训练误差和验证误差与步长的关系。

[0098] 图14E的图示意了对于包含5,000(5k)个训练样本的训练集合,训练误差和验证误差与步长的关系。

[0099] 图14F的图示意了对于包含2,500(2.5k)个训练样本的训练集合,训练误差和验证误差与步长的关系。

[0100] 图14G的图示意了对于包含1,000(1k)个训练样本的训练集合,训练误差和验证误差与步长的关系。

[0101] 图14H的图示意了对于包含500个训练样本的训练集合,训练误差和验证误差与步长的关系。

[0102] 图15A-15D为捕获的视频数据的帧,其具有将各小鼠彼此区分开的重叠颜色指示;(A-B)可见光照射,(C-D)红外光照射。

[0103] 图16为比较图9的分割网络架构与光束中断系统的性能的图。

[0104] 图17A的图示意了来自根据本公开的实施方案和Ctrax的预测。

[0105] 图17B为由图9的分割网络架构确定的短轴预测的相对标准偏差的图。

[0106] 图18A为由图9的分割网络架构确定的基因多样性动物大品系调查所跟踪的总距离的图。

[0107] 图18B为由图9的分割网络架构确定的在动态环境中连续跟踪4天的六只动物中观察到的昼夜活动模式的图。

[0108] 应指出,附图不一定按比例绘制。附图仅旨在描绘本文公开的主题的典型方面,因此不应视为限制本公开的范围。

## 具体实施方式

[0109] 为清楚起见,本文在小型啮齿动物如小鼠的背景下讨论了用于视频捕获一只或多只动物并跟踪一只或多只动物以便于行为监测的系统和相应方法的示例性实施方案。然而,可采用和/或改变所公开的实施方案来监测其他动物而无限制。

[0110] 图1为示意图,示意了操作环境100的一个示例性实施方案,操作环境100包括活动场所200、采集系统700和配置为实施神经网络跟踪器的跟踪系统。如下文更详细地讨论的,可在活动场所200中容纳一只或多只小鼠。获取至少一只动物如小鼠的视频数据。视频数据可单独获取,也可与其他与动物监测有关的数据组合获取,如音频和环境参数(例如,温度、湿度、光强度等)。可通过采集系统700来执行获取该数据的过程,如对相机、麦克风、照明、其他环境传感器、数据存储和数据压缩的控制。可将获取的视频数据输入到跟踪系统,该跟踪系统可执行卷积神经网络(CNN)以基于视频数据跟踪一只或多只动物。

[0111] I. 视频数据采集

[0112] 在一个实施方案中,提供了用于捕获包括动物的运动在内的视频数据的系统和方

法。如下文所讨论,可在预定的时间段(例如,一分钟或多分钟、一小时或多小时、一天或多天、一周或多周、一月或多月、一年或多年等)内连续获取视频数据。视频数据的特性——包括但不限于分辨率、帧速率和位深度中的一项或多项——可能足以便于后续分析以提取行为模式。已提供了一种有效的解决方案并证实比现有的视频捕获系统更稳健且质量更高。本公开的实施方案在可视地标记小鼠的若干方法中测试。还呈现了视频和超声发声数据的同步采集的工作实施例。

[0113] 在一个实施方案中,可将视频监测系统部署用于动物监测达约4-6周的时间段。部署可包括以下中的一项或多项:图像捕获和活动场所设计、腔室设计优化、视频采集软件的开发、音频数据的采集、相机的压力测试、腔室和软件、以及部署阶段的腔室生产决策。下文将详细描述这些内容中的每一个。应理解,上面提及的4-6周的观察期是出于示例的目的提供的,并可根据需要采用本公开的实施方案达更长或更短的时间段。

[0114] a. 活动场所设计

[0115] 适当的活动场所设计对于高质量行为数据的采集可能很重要。该活动场所是动物的“家”并且其可配置为提供以下中的一种或多种:隔离环境干扰、适当的昼夜照明、食物、水、垫料,并且通常是无压力的环境。

[0116] 从行为角度来看,可能需要使该区域的压力和环境干扰最小化并允许表达自然的行为。

[0117] 从养殖角度来看,可能需要活动场所便于清洁、添加或移除以及移除小鼠、添加和移除食物和水。

[0118] 从兽医角度来看,可能需要活动场所便于健康检查和提供医疗处理,以及监测环境条件(例如,温度、湿度、光线等)而基本上不干扰所关注的行为。

[0119] 从计算机视觉角度来看,可能需要活动场所便于高质量视频和音频的采集而无实质性的遮挡、失真、反射和/或噪声污染并不实质性地干扰所关注的行为的表达。

[0120] 从设施角度来看,可能需要活动场所实质性地最小化占地面积并提供相对容易的存放而无需拆卸或重新组装。

[0121] 因此,活动场所可配置为提供行为、养殖、计算和设施之间的平衡。图2中示意了活动场所200的一个示例性实施方案。活动场所200可包括框架202,其上安装围栏204。围栏可包括配置为允许进入到围栏204内部的门206。一个或多个相机210和/或灯212可邻近于框架202安装或直接安装到框架202(例如,围栏204上方)。

[0122] 如下文详细讨论的,在某些实施方案中,灯212可包括至少两组光源。每组光源可包括一个或多个配置为以与其他光源组不同的波长发射光并入射到围栏204上的灯。作为实例,可将第一组光源配置为以一个或多个可见波长(例如,约390nm至约700nm)发射光,并且可将第二组光源配置为以一个或多个红外(IR)波长(例如,约大于700nm至约1mm)发射光。

[0123] 相机210和/或灯212可与用户界面214电通信。用户界面214可以是显示器,配置为查看由相机210获取的视频数据。在某些实施方案中,用户界面214可以是触摸屏显示器,配置为显示一个或多个用户界面以控制相机210和/或灯212。

[0124] 或者或另外,相机110、灯212和用户界面214可与控制器216电通信。控制器216可配置为生成控制信号,该控制信号可用来控制相机210的视频数据采集、灯212的光发射和/

或用户界面214对所获取视频数据的显示。在某些实施方案中,可任选地省略用户界面。

[0125] 控制器216还可与数据存储设备220通信。控制器216可配置为接收由相机210获取的视频数据并将获取的视频数据发送到数据存储设备220以便存储。可使用有线通信链路、无线通信链路及其组合来执行相机210、灯212、用户界面214、控制器216和数据存储设备220中一个或多个之间的通信。

[0126] 如下文所述,活动场所200可具有旷场设计,该旷场设计配置为实现行为、养殖、计算和设施的所需平衡,同时还允许在预定的时间段(例如,约5个月)内完成。

[0127] 材料

[0128] 在某些实施方案中,形成围栏204(例如,围栏204的下部)的材料的至少一部分对于可见光波长可以是基本上不透明的。这样,可减少和/或近似消除围栏204内动物可观察到的除灯212外的光源发出的可见光以及视觉线索(例如,物体和/或用户的运动)。在另外的实施方案中,形成围栏204的材料可对红外波长大致不反射,从而便于视频数据的采集。可在适合于提供机械支承的范围内选择围栏204的壁的厚度(例如,约1/8英寸至约1/4英寸)。

[0129] 在一个实施方案中,围栏204可使用由聚氯乙烯(PVC)或聚甲醛(POM)形成的泡沫板材构建。POM的一个实例为**Delrin®**(DuPont(美国特拉华州威明顿市))。有利地,这种泡沫板材可赋予活动场所200足够的通用性和耐久性以用于长期动物监测。

[0130] 在一个实施方案中,框架202可包括腿202a和在其间延伸(例如,水平地)的一个或多个架子202b。作为一个实例,框架202可以是具有预定尺寸的商业货架系统,其具有锁定轮以移动到存放区域。在一个实施方案中,预定尺寸可以是约2英尺乘2英尺乘6英尺(例如,Super Erecta Metroseal 3™,InterMetro Industries Corporation(美国宾夕法尼亚州威尔克斯巴里))。然而,在其他实施方案中,可采用不同尺寸的活动场所而无限制。

[0131] b. 数据采集

[0132] 视频采集系统可包括相机210、灯212、用户界面214、控制器216和数据存储设备220。可采用具有性能特征的预定平衡的视频采集系统,性能特征包括但不限于视频采集的帧速率、位深度、每一帧的分辨率和红外范围内的光谱灵敏度、以及视频压缩和存储中的一者或多者。如下文所讨论,可优化这些参数以使质量最大化并且使数据量最小化。

[0133] 在一个实施方案中,相机210可获取具有以下中至少之一的视频数据:约640×480像素分辨率、约29fps和约8位深度。使用这些视频采集参数,可生成约33GB/hr的未压缩视频数据。作为一个实例,相机210可以是Sentech USB2相机(Sensor Technologies America, Inc.(美国德克萨斯州卡罗尔顿市))。图3A-3F示意了从使用可见光(图3A-3C)和红外(IR)光(图3D-3F)的视频采集系统实施方案获取的样本帧。

[0134] 收集到的视频数据可由相机210和/或控制器216压缩,如下文所讨论。

[0135] 在另一个实施方案中,视频采集系统可配置为使获取的视频数据的分辨率大致翻倍(例如,至约960×960像素)。研究了比Sentech USB具有更高分辨率的四(4)种另外的相机,如下所示。

[0136] 表1-示例性相机和所选性质

相机	接口	分辨率	最大帧速率	最大位深度	快门	940nm下的量子效率	未压缩数据
Sentech STC-MB33USB	USB3	640×480	60	12	全局	6%	9.2MB/s @30fps, 8位
Basler acA1300-60gm-NIR	GigE	1280×1024	60	10	全局	21%	39.3MB/s @30fps, 8位
Raspberry Pi NoIR V2	串行	1280×720 1640×1232	90 30	8	全局	NA	H264 有损压缩, Max 3.1MB/s
PtGrey BFLY-U3-23S6M-C	USB3	1920×1200	41	12	全局	5%	69.1MB/s @30fps, 8位
Basler daA1600-60um	USB3	1600×1200	60	12	全局	7%	57.6MB/s @30fps, 8位

[0138] 这些相机的成本、分辨率、最大帧速率、位深度和量子效率可能有所不同。

[0139] 视频采集系统的实施方案可配置为以约30fps和约8位深度收集单色的视频数据。根据Shannon-Nyquist定理,帧速率应至少为所关注事件的频率的两倍(例如,参见Shannon,1949)。小鼠行为可从理毛的几赫兹到挥动的20赫兹不等(参见例如Deschenes et al.,2012;Kalueff et al.,2010;Wiltschko et al.,2015)。已观察到理毛以高至约7Hz的频率发生,因此,以大于小鼠运动频率的帧速率(例如,约29fps)记录视频被认为足以观察大多数小鼠行为。然而,相机可能在IR范围内迅速失去灵敏度。虽然可通过提高IR光的水平来克服此对比度损失,但提高IR光的强度会导致环境温度的升高。

#### [0140] 照明

[0141] 如上所述,灯212可配置为发射一种或多种类型的光,如可见白光和红外光。可见光可用于照明和被编程(例如,通过控制器216)以提供明/暗周期和可调节的强度。调节明周期的能力将允许模拟动物在野外遇到的太阳光。可调节明暗时段的长度以模拟季节并可进行光移以模拟时差(昼夜期提前和延迟)实验。另外,在某些动物中可采用强光诱导焦虑并可采用弱光引起不同的探索行为。因此,暂态地控制明/暗时长以及光强度的能力对于恰当的行为实验至关重要。

[0142] 在某些实施方案中,控制器216可配置为命令可见光源在明/暗周期的明部分期间用具有约50勒克斯至约800勒克斯的强度的可见光照射围栏204。所选光强度可随要观察的活动类型而异。一方面,可采用相对低的强度(例如,约200勒克斯至约300勒克斯)来促进小鼠的探索活动以便观察。

[0143] 在某些实施方案中,相机210可使用IR长通滤波器在IR范围内获取大致所有的视频数据。IR长通滤波器可去除输入到相机210的大致所有可见光。有利地,无论白天还是晚上,IR灯都可为围栏104提供均匀的照明。

[0144] 已评估了850nm和940nm LED这两种IR光波长。850nm的光可呈现出裸眼可见的明显红色调并可导致对动物的低光暴露。然而,这种昏暗的光线可能导致小鼠的情绪变化。因

此,选择940nm的光来进行记录。

[0145] 在940nm波长下进行记录可能在相机中具有非常低量子产率;由于增益高,这可能表现为看起来颗粒状的图像。因此,评估了使用不同相机的各种红外照明水平,以确定在没有因红外照射而实质性地提高围栏204的温度的情况下可获得的最大亮度级。在某些实施方案中,围栏204的温度可升高小于或等于约5°C(例如,小于或等于约3°C)。

[0146] 另外,还对Basler acA1300-60gm-NIR相机进行了评估。该相机在940nm下具有与表1中列出的其他相机相比大约3-4倍的光谱灵敏度,如图4A-4B所示。图4A以相对响应示意了作为代表性实例的Sentech相机的光谱灵敏度,而图4B以量子效率示意了Basler相机的光谱灵敏度。量子效率是与撞击传感器的光子相比释放的电子的量度。相对响应是以0至1的尺度表示的量子效率。图4A-4B中还以垂直线示意了940nm波长以供参考。

[0147] 灯212提供的可见光周期可由控制器216或与灯212通信的另一设备控制。在某些实施方案中,控制器216可包括光控制板(Phenome Technologies(伊利诺伊州斯科基))。该板具有对数标度的1024个亮度级,可经由RS485接口进行控制,并且能够实现黎明/黄昏事件。如下文更详细地讨论的,可见光的控制可集成到由控制器216执行的控制软件中。

#### [0148] 滤波器

[0149] 如上所述,任选地,为了在视频数据采集期间阻挡大致所有可见光到达相机210,可采用IR长通滤波器。作为一个实例,相机110可采用物理IR长通滤波器。无论活动场所200中是明阶段还是暗阶段,此配置都可提供基本均匀的照明。

[0150] 图5示出了潜在适用于所公开的系统和方法的实施方案中的滤波器特性曲线(profile)(例如,IR通滤波器092和093)。为了比较,示出了阻挡IR光的IR截止滤波器486。RG-850(玻璃,Edmunds Optics)和43-949(可激光切割塑料,Edmunds Optics)的其他特性曲线也可能适用。

#### [0151] 镜头

[0152] 在一个实施方案中,相机镜头可以是1/3" 3.5-8mm f1.4(CS接口)。该镜头可产生图3A-3B中看到的图像。也可采用C接口镜头中的类似镜头。

#### [0153] 视频压缩

[0154] 忽略压缩,相机210可以以下述速率生成原始视频数据:约1MB/帧、约30MB/秒、约108GB/小时、约2.6TB/天。在选择存储方法时,可考虑各种目标。根据视频的境况,在长期存储之前消除视频的某些元素可能是一个有价值的选择。另外,在考虑长期存储时,可能希望应用滤波器或其他形式的处理(例如,通过控制器216)。然而,对于以后更改处理方法的情况,保存原始或初始视频数据可能是一种有价值的方案。下面将描述视频压缩测试的一个实例。

[0155] 已对以约480×480的像素分辨率、约29fps和每个像素约8位收集约100分钟的视频数据评估了若干种压缩标准。从原始视频测试的两种无损格式为Dirac和H264。H264的文件大小略小,但转码所需的时间略长。Dirac可得到广泛支持,并随后将其转码成另一种格式。

[0156] MPEG4有损格式也已经过评估,因为它与H264密切相关,并且众所周知,它对位速率具有良好的控制。存在两种设置比特率的方法。第一种是设置在整个编码视频中恒定的静态比特率,第二种是基于与原始视频的偏差来设置可变比特率。在使用MPEG4编码器的

ffmpeg中,可通过选择质量值(从0到31,0为几乎无损)来轻松实现可变比特率的设置。

[0157] 图6A-6D中将三种不同的图像压缩方法与原始(初始)捕获的视频帧进行了比较。图4A中示出了原始图像。图4B-4D中由与原始图像的像素差异示出了其他三种方法并仅示出了压缩的效果。即,压缩图像与原始图像有多大不同。相应地,差异越小越好,压缩率越高越好。如图4B所示,使用Q0滤波器根据MPEG4编解码器执行的压缩呈现出1/17的压缩比。如图4C所示,使用Q5滤波器根据MPEG4编解码器执行的压缩呈现出1/237的压缩比。如图4D所示,使用HQDN3D滤波器根据MPEG4编解码器执行的压缩呈现出1/97的压缩比。

[0158] 当使用质量0参数(Q0滤波器,图4B;Q0 HQDN3D滤波器,图4D)时,根据所公开实施方案收集的的视频数据有大约0.01%的像素相对于原始视频数据发生了改变(最多增加或减少了其强度的4%)。每帧中这占约25个像素。这些像素中的大多数位于阴影的边界中。可以理解,如此小的图像变化是沿着干扰相机210自身的噪声标尺发生的。采用较大的质量值(例如,图4C的Q5),可引入伪像以更好地压缩视频数据。这些常以压缩过程中不小心时出现的块状像素化伪像为人们所熟悉。

[0159] 除了这些格式外,还可创建其他合适的无损格式来容纳单个用户的数据集。其中两个包括FMF编解码器(飞行电影格式)和UFMF编解码器(微型飞行电影格式)。这些格式的目的是最大限度地减少无关的信息并优化可读性以便跟踪。由于这些格式是无损的并可在静态背景模型上运行,故未经滤波的传感器噪声不允许进行任何实质性的数据压缩。压缩评估的结果如表2中所示。

[0160] 表2-压缩测试

	编解码器	滤波器	文件大小(字节)	注释
	原始	无	41,911,246,848	
	Dirac	无	15,568,246,208	无损
	H264	无	14,840,999,602	无损
	MPEG4	HQDN3D	2,471,731,352	有损
[0161]	Dirac	HQDN3D	10,113,632,202	滤波
	H264	HQDN3D	8,008,906,680	滤波
	MPEG4	HQDN3D Q0	429,522,590	有损 滤波
	FMF	无	41,720,668,160	无损
	UFMF	无	41,722,662,912	无损

[0162] 除了用于数据压缩的编解码器的选择外,还希望减少图像的背景噪声。背景噪声是所有相机所固有的,常以暗噪声表示,其指的是图像中的基线噪声。

[0163] 有许多消除此噪声的方法,包括更长的曝光时间、更大的光圈和降低增益。然而,如果这些方法直接影响实验,则不是可行的选择。因此,可采用ffmpeg的HQDN3D滤波器,该滤波器可获取时间和空间信息以消除小的波动。

[0164] 如图6B-6D所示,观察到HQDN3D滤波器提供所获取视频数据的文件大小的显著减小(例如,与原始视频数据的文件大小相比大约小100倍)。在采用带有HQDN3D滤波器的MPEG4编解码器进行压缩后,所得平均比特率可为约0.34GB/小时的压缩视频。此外,通过实验验证,基本上所有信息丢失都比传感器噪声(在没有小鼠的情况下获取的视频)产生的信

息丢失少几个数量级。这种类型的噪声消除大大增强了可压缩性。

[0165] 出乎意料的是,已发现HQDN3D滤波器显著提高由卷积神经网络(CNN)进行的跟踪的性能,这将在下文详细讨论。不受理论的束缚,据信由于HQDN3D滤波器是基于方差的背景减除方法,故可实现这种性能改进。方差较小时,前景更容易识别并产生更高质量的跟踪。

[0166] 超声波音频采集

[0167] 小鼠可在超声波范围内发声以进行社交、交配、进攻和幼崽饲养(参见例如Grimsley et al.,2011)。除嗅觉和触觉提示外,这种发声可能是小鼠交流中最主要的形式之一。虽然未在小鼠中进行测试,但在人类中,声音和发声的变化(老年嗓音)可定义如青春期和变衰老之类的转变(参见例如Decoster and Debruyne,1997;Martins et al.,2014;Mueller,1997)。

[0168] 因此,如下文详细讨论的,活动场所200的实施方案可还包括一个或多个麦克风222。麦克风222可安装到框架202并配置为从位于围栏204中的动物获取音频数据。可通过使用呈麦克风阵列形式的麦克风222来引导同步数据收集。麦克风222的这种配置允许精准定位哪只小鼠正在发声。最近已证实了使用麦克风阵列进一步确定一组小鼠中是哪只小鼠正在发声的能力(参见例如Heckman et al.,2017;Neunuebel et al.,2015)。

[0169] 可与Neunuebel et al.类似地提供数据收集装置。可将四个麦克风放在能够捕获声音的活动场所侧面上。当与视频数据集成时,可使用最大似然法来识别正在发声的小鼠(参见例如Zhang et al.,2008)。

[0170] 环境传感器

[0171] 在一个实施方案中,活动场所200可还包括一个或多个环境传感器224,其配置为测量一个或多个环境参数如温度、湿度和/或光强度(例如,可见光和/或IR)。在某些实施方案中,环境传感器224可被集成并配置为测量两个或更多个环境参数(例如,Phenome Technologies(伊利诺伊州斯科基))。环境传感器224可与控制器216电通信以收集每日温度和湿度数据以及亮度级。可输出收集的环境数据以显示在表明最低和最高温度以及灯光活动的用户界面中(参见下面的控制软件讨论)。

[0172] 软件控制系统

[0173] 可由控制器216执行软件控制系统以进行数据采集和灯控制。软件控制系统可配置为提供具有相应时间戳的视频、音频/超声波和环境数据的独立收集。可以以这种方式在任何预定的时间段(例如,一秒或多秒、一分钟或多分钟、一小时或多小时、一天或多天、一年或多年等)内不间断地收集数据。这可允许以后对所获取视频、音频/超声波和/或环境数据进行汇编或同步以便分析或呈现。

[0174] 操作系统

[0175] 操作系统的选择可由各种传感器的驱动器的可得性驱动。例如,Avisoft Ultrasonic麦克风驱动器仅与Windows操作系统兼容。然而,选择可能影响以下方面:

[0176] • 过程间通信-过程间通信的选项受底层OS的影响。同样,OS会影响线程(thread)之间的通信选择,但在像QT这样的跨平台框架上进行开发可弥补这一障碍。

[0177] • 访问系统时钟-访问高分辨率系统时钟的方法因OS不同而不同,这将在下文更详细地讨论。

[0178] 硬件选项

[0179] 在某些实施方案中,控制系统可由控制器216以单板计算机的形式实施。存在可高度稳健地连续运行的多种可用选项,包括MilSpec/工业计算机。

[0180] 外部时钟与系统时钟

[0181] 系统时钟可提供足够的实时时钟值而无需向系统中引入外部时钟。在Posix系统上,clock\_gettime(CLOCK\_MONOTONIC,…)函数可返回秒和纳秒。可使用clock\_getres()函数查询时钟的分辨率。可能希望控制系统的实施方案的时钟分辨率小于约33毫秒的帧周期。在一个实施方案中,系统时钟为Unix系统。

[0182] 已开发出GetTickCount64()系统函数用来获取自系统启动以来的毫秒数。该计时器的预期分辨率在约10毫秒至约16毫秒之间。这与clock\_gettime()系统调用的目的相同,但检查和考虑值换行可能是有利的。

[0183] 在Macintosh计算机上,对系统时钟的访问是类似的。评估以下代码段并观察到亚微秒的分辨率。

[0184] clock\_serv\_t cclock;

[0185] mach\_timespec\_t mts;

[0186] host\_get\_clock\_service(mach\_host\_self(),SYSTEM\_CLOCK,&cclock);

[0187] clock\_get\_time(cclock,&mts);

[0188] 在任何OS上,由于定期进行调整,故返回一天中时间的系统调用有时可能后移。在一个实施方案中,可采用单调增加的系统时钟。GetTickCount64()、clock\_gettime()和clock\_get\_time()都可满足此标准。

[0189] 视频文件大小

[0190] 相机供应商软件不大可能保存自动分割成合理大小并充分地带时间戳的输出文件。对于控制器116的实施方案,期望以不间断的方式收集视频数据,从相机110读取每个帧并以简单的形式提供所收集的视频数据。例如,控制器116可配置为以原始格式向数据存储设备120提供每个文件约10分钟的视频帧,具有时间戳报头或帧与帧之间的时间戳。这样每个文件将小于2GB。

[0191] 控制系统架构

[0192] 图7为框图,示意了采集系统700的部件。在某些实施方案中,采集系统700可由控制器216执行。每个框代表一个单独的过程或执行线程。

[0193] 控制过程

[0194] 控制过程可配置为启动和停止其他过程或线程。控制过程还可配置为向采集系统700提供用户界面。控制过程可配置为保存活动日志并可跟踪采集过程中遇到的误差(例如,在日志中)。控制过程也可配置为重新启动已变为非活动状态的过程或线程。

[0195] 部件之间的通信方法可在选择系统OS之后确定。控制过程的用户界面可以是命令行界面,或者其可以是图形界面。可在便携式框架如QT上构建图形界面,从而提供与OS的独立性。

[0196] 视频采集过程

[0197] 视频采集过程可配置为与相机210直接通信,从而将带时间戳的帧保存到数据存储设备220。视频采集过程可以高优先级运行,从而最大限度地降低丢帧的可能性。视频采集过程可保持相对简单,并且帧之间的处理极少。视频采集过程也可配置为控制灯212发出

的IR照射以确保以极小的有效快门速度进行适当的曝光。

#### [0198] 音频采集过程

[0199] 可将单独的音频采集过程配置为以适宜的时间戳获取超声音频数据。在一个实施方案中,音频系统可包括与围栏204音频通信的麦克风222阵列。在某些实施方案中,一个或多个麦克风222可位于围栏204内。麦克风阵列的麦克风可具有以下一项或多项功能:约500kHz的采样频率,约16位的ADC分辨率,约10kHz至约210kHz的频率范围,及8阶和210kHz的抗混叠滤波器。作为一个实例,麦克风阵列的麦克风可包括Pettersson M500麦克风(Pettersson Elektronik AB(瑞典乌普萨拉))或其功能等价物。如上文所讨论,可给麦克风222捕获的音频数据加时间戳并提供给控制器216以进行分析和/或提供给数据存储设备220以进行存储。

#### [0200] 环境数据采集过程

[0201] 可将单独的环境数据采集过程配置为收集环境数据,如温度、湿度和照明水平。可以以低频(例如,约0.01Hz-0.1Hz)收集环境数据。可由数据存储设备220将环境数据存储为带有每个记录的时间戳(例如,存储为一个或多个CSV文件)。

#### [0202] 照明控制过程

[0203] 可将照明控制过程配置为控制由灯212发出的可见光以便为小鼠提供日间/夜间周期。在一个实施方案中,如上文所讨论,相机210可配置为仅对IR响应而滤出基本上所有可见光,并且可对可见光滤波以不产生IR,从而该过程可避免影响视频捕获。

#### [0204] 视频汇编过程

[0205] 可将视频汇编过程配置为以预定的压缩将获取的视频数据重新打包成预定的格式。该过程可与视频采集保持分开以最大限度地减少丢帧的机会。此视频汇编过程可作为低优先级后台任务运行,或甚至可在数据采集完成后运行。

#### [0206] 看门狗(Watchdog)过程

[0207] 可将看门狗过程配置为监测数据采集过程的健康情况。例如,它可记录问题(例如,在日志中)并根据需要触发重启。看门狗过程还可侦听其正在监视的部件的“心跳”。通常,所述心跳可以是发送到控制器216的信号,其确认系统700的部件正在正常运行。作为一个实例,如果系统700的一个部件停止运转,则控制器216可检测到缺乏由该部件发送的心跳。在该检测之后,控制器216记录该事件并可使警报被发出。这样的警报可包括但不限于音频警报和视觉警报(例如,灯、字母数字显示等)。作为此类警报的替代或补充,控制器216可尝试重新启动部件的运行,如发送重新初始化信号或重启电源。系统700的部件与控制器216之间的通信方法可取决于OS的选择。

#### [0208] 小鼠标记

[0209] 在某些实施方案中,可标记小鼠以便于跟踪。然而,如下文更详细地讨论的,可省去标记而通过其他技术来便于跟踪。

[0210] 为视觉识别标记小鼠有着若干不易解决的参数。在一个实施方案中,可以对小鼠自身不可见的方式对小鼠进行长期(数周)标记,从而最大限度地减少对小鼠交流和行为的影响。作为一个实例,可采用在正常小鼠观察范围内不可见的长期IR敏感标记。

[0211] 在一个替代的实施方案中,可使用人发色和毛发漂白剂标记小鼠皮毛。这种手段可在数周内清楚地识别小鼠并可成功地用于行为实验中(参见例如Ohayon et al.,2013)。

然而,毛发标记过程需要麻醉小鼠,这一过程对于目前的小鼠监测系统是不可接受的。麻醉会改变生理,并且染发剂本身可能是刺激物,常常会改变小鼠行为。由于每只D0小鼠都是独特的,故这可能导致染料/麻醉x基因型效应并引入一个或多个未知变量。

[0212] 也可采用和优化使用基于IR染料的标记和纹身的其他替代方法。

[0213] 在又一个实施方案中,可采用剃毛来在小鼠的背部形成图案作为一种标记形式。

[0214] 数据存储

[0215] 在显影(development)阶段,可能需要少于2TB的总数据。这些数据可包括来自各种相机和压缩方法的原始样本和压缩视频。因此,可在压力测试期间实现集成式USV和视频数据以及较长期的长达7-10天视频数据的数据传输。视频的大小可根据选择的压缩标准来减小。下面提供了样本数据存储估算值。

[0216] 测试:

[0217] 1个活动场所

[0218] 至多5个相机

[0219] 视频持续时间:每个~1-2小时

[0220] 总共~10GB,高端。

[0221] 压力测试:

[0222] 1个活动场所

[0223] 1个相机

[0224] 视频持续时间:14天

[0225] 分辨率:2x电流(960x 960)

[0226] 总共~2TB

[0227] 生产:

[0228] 总共120个试验(12-16个活动场所,每组运行80只动物,交错实验)

[0229] 持续时间(每个):7天

[0230] 分辨率:2x电流(960x 960)

[0231] 32.25TB

[0232] II. 动物跟踪

[0233] 在没有高度用户参与的情况下,在现有动物监测系统中,不能在复杂且动态的环境中或与遗传异质动物一起进行动物如小鼠的视频跟踪,这使得大规模实验不可行。如下文所讨论,尝试使用现有系统和方法在多种环境中跟踪大量不同的小鼠品系,并证实这些系统和方法对于大规模实验的数据集是不够的。

[0234] 使用含不同皮毛颜色(包括黑色、野灰色、白化、灰色、棕色、裸色和花斑)的小鼠的示例性数据集进行分析。按JAX-IACUC批准的方案对所有动物进行测试,所述方案概述如下。对8至14周龄之间的小鼠进行测试。数据集包括59个品系的1857个视频,总共1702小时。

[0235] 所有动物均得自杰克逊实验室生产群体。按杰克逊实验室动物保护与使用委员会指南批准的方案对8至14周龄的成年小鼠进行行为测试。如Kumar, 2011所述进行旷场行为分析。简言之,在开始视频录制之前,对成组圈养小鼠称重并让在测试室中适应30-45分钟。这里给出了活动的前55分钟的数据。在可获得的情况下,从每个自交系和F1同基因型品系各测试8雄和8雌。

[0236] 在一个方面,可能希望在具有白色背景的另一旷场设备(例如,活动场所200)中跟踪多只动物。图8A的第1行(全帧)和第2行(裁剪)中示意了由视频采集系统采集的全帧和裁剪视频图像的实例。在每一环境中对各种遗传背景示出了理想和实际跟踪帧的实例(图8A第3行[理想跟踪]和第4行[实际跟踪])。

[0237] 在另一个方面,期望在具有挑战性的环境中进行行为的视频分析,如在包括食物和水杯的活动场所200的实施方案中和杰克逊实验室的基因敲除小鼠项目(KOMP2)(分别在图8A中第5列和第6列)。

[0238] 在24-小时装置中,将小鼠圈养在有白纸垫料和食物/水杯的活动场所200中。将小鼠保持在活动场所200中并使用灯212发出的红外光在明暗条件下进行连续记录。垫料和食物杯被小鼠移动,并在每一天的过程中改变灯212发出的可见光以模拟明/暗周期。

[0239] KOMP2项目已收集了超过五年的数据,并希望进行基于视频的记录作为一种附加的分析模式来检测光束中断系统所无法识别的步态影响。在步态分析中,分析了动物的运动。如果某动物步态异常,则可推导出骨骼、肌肉和/或神经等异常。KOMP2项目使用光束中断系统,其中将小鼠置于透明的聚碳酸酯盒子中,盒子四周都有红外光束。基材的地板也是聚碳酸酯的,下面的工作台表面为深灰色。可以观察到,一些放置在两张桌子交界处的盒子可能会留下接缝,顶上的灯(例如,LED灯)会在所有盒子上引起独特的高眩光。

[0240] 在一个方面,尝试使用Ctrax跟踪此视频数据集,Ctrax是一种现代的开源跟踪工具,其使用背景减除和斑点检测启发式。Ctrax在每帧的基础上将小鼠抽象为五个指标:长轴和短轴、小鼠中心的x和y位置、以及动物的方向(Branson, 2009)。它利用MOG2背景减除模型,其中软件估算视频背景的均值和变异以用于背景减除中。Ctrax使用预测的前景的形状来拟合椭圆。

[0241] 在另一个方面,尝试使用Lime Light跟踪此视频数据集,Lime Light是一种使用专有跟踪算法的市售跟踪软件。Lime Light使用单个关键帧背景模型进行分割和检测。一旦检测到小鼠,Lime Light就会使用专有算法将小鼠抽象到质心。

[0242] 数据集含有对这些现有分析系统的重大挑战。作为一个实例,使用Ctrax和Lime Light很难处理小鼠皮毛颜色和环境的组合。通常,在白色背景上表现出高对比度的环境——如较深色的小鼠(例如,黑色、野灰色)——会产生良好的跟踪结果。然而,在白色背景上对比度较低的环境——如较浅色的小鼠(例如,白化、灰色或花斑小鼠)——会产生不良结果。在白色旷场中的黑色小鼠可实现高的前景-背景对比度,因此实际跟踪非常接近理想情况。灰色小鼠在视觉上与活动场所的墙壁类似,因此在用后腿直立于墙上时其鼻子常常是消失的。白化小鼠与活动场所本身的背景类似,在跟踪过程中经常不可见。花斑小鼠由于其带图案的皮毛颜色而被分成两半。在尝试针对每个视频优化和微调Ctrax时,仍然观察到大量不良跟踪帧,如比较图8A第3行(理想跟踪)与第4行(实际跟踪)的实际跟踪结果所示。弃去不良跟踪帧会导致抽样偏差和生物学解释偏斜,这是不希望的。

[0243] 随着环境变得不太适合跟踪,如24-小时和KOMP2环境,观察到这些误差增加。此外,误差的分布不是随机的。例如,如图8第4行(实际跟踪)中所示,当小鼠在角落中、墙壁附近或在食物杯上时,发现跟踪是高度不准确的,但当动物在中心时,准确度会提高。在24-小时环境中向活动场所中放置食物杯将造成小鼠爬到上面时的跟踪问题。具有反射表面的活动场所,如KOMP2,也会因跟踪算法而产生误差。

[0244] 进一步探索不良跟踪的原因,发现在大多数情况下,不正确的跟踪是由于小鼠与背景的分割不良。这包括在将小鼠从前景移开或因不良对比度而把背景包括在前景中时的情况。传统上,通过改变用于优化视频数据收集的环境来解决其中的一些障碍。例如,要跟踪白化小鼠,可将旷场的背景颜色更改为黑色而增加对比度。然而,这样的环境改变在当前情况下是不合适的,因为环境的颜色会影响小鼠和人类的行为并且这样的操作可能潜在地混淆实验结果(Valdez, 1994; Kuleskaya, 2014)。另外,在24小时数据收集系统或KOMP2活动场所中,此类方案对花斑小鼠可能无法奏效。

[0245] 由于Ctrax使用单一背景模型算法,故进行测试以确定其他背景模型是否可改善跟踪结果。测试了26种不同的分割算法(Sobral, 2013)并发现这些传统算法中的每一种在某些情况下表现良好但在其他地方则失败,如图8B所示。用于动物跟踪的其他可用系统和方法包括依靠背景减除手段进行跟踪。由于所有26种背景减除方法均失败,故认为Ctrax和Limelight的结果代表了这些其他技术。据信这些分割算法由于不适当的分割而失败。

[0246] 因此,尽管存在许多跟踪方案以便分析视频数据,但克服适当的小鼠分割这一基本问题以用现有方案的代表性实例实现高保真小鼠跟踪的尝试并不成功。没有人适当地解决了小鼠分割这一基本问题并通常依靠环境优化来实现适当的分割,因此会造成潜在的混淆。

[0247] 此外,用于微调背景减除算法参数的时间成本可能令人望而却步。例如,在跟踪来自24小时装置的数据时,当小鼠以一种姿势长时间睡眠时,小鼠便成为背景模型的一部分而不能被跟踪。对于每一小时的视频,典型的监督工作将花费有经验的用户5分钟的互动时间以确保高质量的跟踪结果。虽然此用户互动水平对于较小且受限制的实验是易处理的,但大规模和长期的实验需要大量的时间投入来监督跟踪性能。

[0248] 本公开的实施方案克服了这些困难并构建了稳健的下一代跟踪器,该跟踪器适合于分析含动物如小鼠的视频数据。如下文将详细讨论的,采用了人工神经网络,该人工神经网络可在复杂且动态的环境条件下实现高的性能,不关心毛色的遗传,并且不需要用户进行持续的微调。

[0249] 卷积神经网络是包括多个处理层的计算模型,这些处理层学习具有多个抽象级别的数据表示。这些方法已在语音识别、视觉目标识别、目标检测及许多其他领域如药物发现和基因组学方面极大地改进了现有技术(LeCun, 2015)。一个优点是,一旦开发出具有合适的超参数的有效网络,则只需添加适当的训练数据就可将神经网络轻松扩展到其他任务。因此,所公开的实施方案提供了可高度泛化的小鼠跟踪解决方案。

[0250] 神经网络架构

[0251] 已开发出3种主要的网络架构以解决视觉跟踪问题。在一个实施方案中,目标跟踪可基于分割掩膜采用小鼠的椭圆描述形式,如图8C所示。参见Branson, 2005。在替代的实施方案中,可采用非椭圆的形状。

[0252] 椭圆表示可通过六个变量(本文中也称为参数)来描述动物的位置。在一个方面,变量之一可以是定义在预定坐标系中的位置的坐标(例如,笛卡尔坐标系的 $x$ 和 $y$ ),该坐标表示小鼠在所采集的视频帧中的像素位置(例如,平均中心位置)。即,平面内唯一的像素位置。任选地,如果需要,可在视频帧中检测界标(例如,围栏204的角)以帮助确定坐标。在另一个方面,变量可还包括小鼠的长轴长度和短轴长度,以及长轴矢量角的正弦和余弦。该角

度可相对于长轴的方向定义。长轴可在视频帧的坐标系中从大约动物头的尖端(例如,鼻子)延伸到动物身体的末端(例如,到大约动物的尾巴从其身体延伸至的点)。为了清楚起见,在本文中,将经裁剪的帧示意为神经网络的输入,而实际输入为未标记的完整帧。

[0253] 下面详细讨论用于利用神经网络架构确定椭圆参数的示范性系统和方法。应理解,可视情况通过所公开的实施方案采用和确定其他参数。

[0254] 在一个实施方案中,第一架构为编码器-解码器分割网络。如图9所示,该网络从给定的输入帧预测前景-背景分割图像并可逐像素地预测是有小鼠还是无小鼠,输出为分割掩模。

[0255] 该第一架构包括特征编码器,其配置为将输入抽象为小空间分辨率特征集合(例如,5×5,而不是480×480)。许多参数被分配给神经网络进行学习。可通过有监督的训练来进行学习,其中为神经网络提供示例并调整参数以产生正确的预测。下表3中示意了最终模型定义和训练超参数的完整说明。

[0256] 表3- 训练超参数

[0257]	模型 所有	参数 旋转	值 ±2.5 deg.
[0258]	编码器-解码器分割网络	平移 加性噪声 亮度 对比度	±5.0 px μ = 0.0, σ = 5.0 ±5% ±5%
	回归网络	学习率 批尺寸	10 <sup>-5</sup> 50
	分箱分类网络	学习率 批尺寸	10 <sup>-5</sup> 5 10 <sup>-3</sup> 25

[0259] 在特征编码器之后是特征解码器,其配置为将小空间分辨率特征集合转换回与原始输入图像相同的形状。即,神经网络中习得的参数反转特征编码操作。

[0260] 向编码特征添加3个全连接层以预测椭圆面对的基本方向。全连接层可指神经网络层,其中给定层中的每个数字都乘以不同的参数(例如,可学习的参数)并然后相加在一起以在新的层中产生单个值。可训练该特征解码器以产生前景-背景分割图像。

[0261] 网络的前半部分(编码器)利用2D卷积层,然后进行批量归一化、ReLU激活和2D最大池化层。更多详细信息可见于Goodfellow,2016中。

[0262] 采用初始为8的滤波器规格,其在每一个池化层之后翻倍。对于2D卷积层,使用的内核是5×5形状的,而对于最大池化层,使用的内核是2×2形状的。输入的视频是480×480

$\times 1$ 形状的(例如,单色),并在6个这些重复层之后,所得形状为 $15 \times 15 \times 128$ (例如,128色)。

[0263] 在替代的实施方案中,可采用其他形状的池化层,如 $3 \times 3$ 。重复层表示层的重复结构。神经网络针对每一层学习不同的参数并将各层堆叠在一起。虽然上面讨论了6个重复层,但可采用更多或更少数量的重复层。

[0264] 应用另一个2D卷积层(内核 $5 \times 5$ 、 $2 \times$ 滤波器),然后是2D最大池,其具有 $3 \times 3$ 的不同内核并且步长为3。 $15 \times 15$ 的空间形状可通过使用系数3来进一步减小。典型的最大池为内核 $2 \times 2$ 、步长2,其中每个 $2 \times 2$ 网格选择最大值并产生1个值。这些设置选择 $3 \times 3$ 网格中的最大值。

[0265] 应用最终的2D卷积层以产生形状为 $5 \times 5 \times 512$ 的特征瓶颈。特征瓶颈是指编码特征集合,即从所有这些矩阵运算输出的实际矩阵值。学习算法将编码特征集合优化至对于训练编码特征集合以执行良好的任务而言最有意义。然后将此特征瓶颈传递给分割解码器和角度预测器。

[0266] 分割解码器使用跨步转置2D卷积层反转编码器并通过求和结点进行预下采样的激活。应指出,该解码器不利用ReLU激活。预下采样的激活和求和结点也可称为跳跃连接。由于解码时一层处的特征与编码器层的形状匹配,故允许网络在更好地编码或在编码器状态下保持原有特征之间进行选择。

[0267] 在层恢复至 $480 \times 480 \times 8$ 形状后,应用内核大小为 $1 \times 1$ 的附加卷积将深度合并为2张单色图像:背景预测和前景预测。最终输出为 $480 \times 480 \times 2$ (2色)。指定第一颜色代表背景。指定第二颜色代表前景。取决于每个像素,两者中较大的一个是网络认为是输入像素的像素。如下文所讨论,softmax操作会重新缩放这些颜色至共计为1的累加概率。

[0268] 随后,在此深度上应用softmax.softmax是分类或分箱成组的一种形式。关于softmax的其他信息,可见于Goodfellow,2016中。

[0269] 从特征瓶颈也可创建角度预测。这可通过应用具有批量归一化和ReLU激活(内核尺寸 $5 \times 5$ ,特征深度128和64)的2个2D卷积层来实现。从这里开始,一个全连接层被展平并用来产生4个神经元的形状,这些神经元用来预测小鼠头所朝向的象限。有关批量归一化、ReLU激活和展平的更多详细信息,可参见Goodfellow,2016。

[0270] 由于角度是由分割掩模预测的,故仅需要选择正确的方向( $\pm 180$ 度)。即,因为被预测的是椭圆,所以只有一个长轴。长轴的一端在小鼠的头的方向上。假定小鼠沿头-尾轴较长。因此,一个方向为 $+180^\circ$ (头),而另一个方向为 $-180^\circ$ (尾)。编码器-解码器神经网络架构可选择的4个可能的方向是极坐标网格上的45-135、135-225、225-315和315-45度。

[0271] 选择这些边界以避免角度预测中的不连续性。值得指出的是,如上文所讨论,角度预测为长轴的矢量角的正弦和余弦的预测并采用atan2函数。atan2函数具有不连续性(在 $180^\circ$ 处),而选择的边界将避免这些不连续性。

[0272] 在网络产生分割掩模后,可如Branson,2009中所述将椭圆拟合算法应用于跟踪。虽然Branson使用加权样本均值和方差进行这些计算,但分割神经网络在描述改进之处的情况下仍保持不变。投射的阴影会给背景减除算法产生的分割掩模增加误差。神经网络学习而根本没有这些问题。另外,在使用加权和未加权样本均值和方差之间未观察到显著差异。使用由所公开的神经网络的实施方案预测的掩模从加权和未加权方法预测的椭圆拟合参数没有显著不同。

[0273] 给定分割掩模,计算像素位置的样本均值以表示中心位置。

$$[0274] \quad \mu_{x,y} = \frac{1}{N} \sum_i^N p_i \quad (1)$$

[0275] 类似地,计算像素位置的样本方差以表示长轴长度(a)、短轴长度(b)和角度( $\theta$ )。

$$[0276] \quad \sigma = \frac{1}{N} \sum_i^N (p - \mu_{x,y})(p - \mu_{x,y})^T \quad (2)$$

[0277] 为了获得轴长度和角度,必须求解特征值分解方程。

$$[0278] \quad \sigma = U^T D U, \quad U = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}, \quad D = \begin{pmatrix} \frac{a}{2} & 0 \\ 0 & \frac{b}{2} \end{pmatrix}^2 \quad (3)$$

$$[0279] \quad a = 2\sqrt{D_{11}}, \quad b = 2\sqrt{D_{22}}, \quad \theta = \text{atan}(U_{12}, U_{21}) \quad (4)$$

[0280] 第二网络架构为分箱分类网络。分箱分类网络架构的结构可预测每个椭圆拟合参数的最概然值的热图,如图10所示。

[0281] 该网络架构始于将输入图像抽象为小空间分辨率的特征编码器。虽然大多数回归预测器都是通过边界框(例如,正方形或矩形)实现解决方案的,但椭圆只是增加了一个附加参数:角度。由于角度是在 $360^\circ$ 和 $0^\circ$ 下等价的重复序列,故将角度参数转换为其正弦和余弦分量。这产生总共6个从网络回归的参数。该网络的前半部分编码一组与解决问题有关的特征。

[0282] 通过将表示特征的矩阵(阵列)转换为单个矢量,可将编码特征展平。展平的编码特征然后被连接到附加的全连接层,其输出形状由输出的期望分辨率决定(例如,通过将特征的矢量输入到全连接层中)。例如,对于小鼠的X-坐标位置,有480个箱,一个箱对应于 $480 \times 480$ 像素图像中的每一个x-列。

[0283] 当网络运行时,每个热图中的最大值被选择为最概然值。每个期望的输出参数可实现作为与编码特征相连的一组独立的可训练的全连接层。

[0284] 测试了多种预构建的特征检测器,包括Resnet V2 50、Resnet V2101、Resnet V2 200、Inception V3、Inception V4、VGG和Alexnet。特征检测器表示对输入图像进行运算的卷积。除了这些预构建的特征检测器外,还调查了各种各样的定制网络。观察到Resnet V2 200在此调查中表现最佳。

[0285] 最终的架构是回归网络,如图11所示。例如,它获取输入视频帧,通过Resnet200 CNN提取特征,并直接预测椭圆拟合的6个参数。每个值(对于椭圆拟合为6)是连续的,可具有无限范围。网络必须学习适宜的值的范围。这样,直接从输入图像预测椭圆数值以描述跟踪椭圆。即,回归网络不是直接预测参数,而是从分箱的可能值的选项中选择最概然值。

[0286] 其他神经网络架构以不同的方式运行。编码器-解码器神经网络架构输出每个像素是小鼠或者不是小鼠的概率。分箱分类神经网络架构输出描述小鼠位置的箱。每个参数的类别是预先确定的,并且网络(编码器-解码器或分箱)只须为每个类别输出概率。

[0287] 回归网络架构始于将输入抽象为小空间分辨率的特征编码器。与先前的架构相比,回归神经网络训练依赖于交叉熵损失函数,而不是均方误差损失函数。

[0288] 由于存储限制,仅测试了具有缩小的特征尺寸的定制VGG样网络。表现最佳的网络由2个2D卷积层、随后是2D最大池化层构造。对于2D卷积层,使用的内核是 $3 \times 3$ 形状的,而对于2D最大池化层,使用的内核是 $2 \times 2$ 形状的。最初使用为16的滤波器深度并在每一个2D最

大池化层之后翻倍。将此2卷积加最大池序列重复5次以产生 $15 \times 15 \times 256$ 的形状。

[0289] 该层被展平并连接到每个输出的全连接层。每个输出的形状取决于所需的分辨率和预测范围。例如，然后将这些编码特征展平并连接到全连接层以产生输出形状为6，即要求网络预测来拟合椭圆的值的数量。出于测试目的，仅观察中心位置并用整个图像范围(0-480)进行训练。另外的输出如角度预测可作为另外的输出矢量而轻松地添加。测试了多种现代的特征编码器，本文针对该网络讨论的数据来自具有200层的Resnet V2，其在该架构中实现了最佳性能结果(He, 2016)。

[0290] 训练数据集

[0291] 为了测试网络架构，如下文所讨论使用基于OpenCV的标注界面生成跨越多个品系和环境的16,234个训练图像和568个单独的验证图像的训练数据集。该标注界面允许快速标注前景和背景，以及椭圆拟合，并可用来快速生成训练数据以便通过迁移学习使任何网络适应新的实验条件。

[0292] 采用OpenCV库来创建基于分水岭的交互式分割和基于轮廓的椭圆拟合。使用该软件，用户左键单击将点标记为前景(例如，小鼠;F)，而右键单击将其他点标注为背景(B)，如图12A所示。击键时，执行分水岭算法来预测分割和椭圆，如图12B所示。如果用户需要对预测的分割和椭圆进行编辑，则他们可以只是标注更多的区域并再次运行分水岭算法。

[0293] 当在神经网络用户(例如，研究者)选择的预定容错范围内预测时，用户将选择椭圆的方向。他们通过选择四个基本方向——上、下、左、右中之一来做到这一点。由于确切的角度是由椭圆拟合算法选择的，故用户只需要确定 $\pm 90$ 度的方向即可。一旦选择了方向，所有相关数据将保存并为用户呈现一个新的帧以进行标注。

[0294] 带注释的数据集的目的旨在针对小鼠确定良好的椭圆拟合跟踪数据。在标注数据时，对椭圆拟合进行优化以使椭圆位于小鼠躯干的中心，长轴边缘与小鼠的鼻子大致相切。往往将尾巴从分割掩模去除以提供更好的椭圆拟合。

[0295] 为了训练网络以便推理，生成3个带注释的训练集。每个数据集都包含参考帧(输入)、分割掩模和椭圆拟合。生成每个训练集以跟踪不同环境中的小鼠。

[0296] 第一种环境是具有恒定白色背景的旷场，其含16,802个带注释的帧。前16,000帧是从65个独立视频标注的，这65个独立视频从24个相同装置之一获取。在第一次网络训练之后，观察到在未包括在带注释的数据中的特殊情况下网络的性能较差。在活动场所中跳到半空、奇怪的姿势和排尿的情况是通常观察到的故障。这些故障被识别、正确地标注、并且合并到带注释的训练集里以进一步泛化和改进性能。

[0297] 第二种环境是在两个不同的照明条件下(白天可见光照射和夜晚红外光照射)具有ALPHA-dri垫料(dri bedding)和食物杯的标准旷场。在此数据集里，4天内跨6个装置对总共2,192个帧进行了注释。带注释的帧中916个取自夜间照射，带注释的帧中1,276个取自日光照射。

[0298] 最终的带注释的数据集是使用KOMP数据集的Opto-M4旷场笼创建的。该数据集包括1083个标注帧。这些标签都是在不同的视频(每个视频标注1帧)和8种不同的装置中采样的。

[0299] 神经网络训练

[0300] a) 训练数据集扩充

[0301] 在训练过程中通过应用反射并在对比度、亮度、旋转方面应用小的随机变化来将此训练数据集扩充8倍,以使网络对于输入数据的微小波动具有稳健性。进行扩充是为了抑制神经网络记忆训练数据集,这可能导致数据集以外的实例不良的性能(验证)。进一步的讨论可见于Krizhevsky,2012中。

[0302] 自Alexnet (Krizhevsky,2012)以来,训练集扩充一直是训练神经网络的重要方面。利用一定量的训练集扩充来实现良好的正则化性能。由于数据来自鸟瞰图,故应用水平、垂直和对角反射来使等价训练集大小立即增加8倍将是直截了当的。另外,在运行时,对整个帧应用小的旋转和平移。从均匀分布对旋转扩充值采样。最后,也可对帧应用噪声、亮度和对比度扩充。从正态分布选择用于这些扩充的随机值。

[0303] b) 训练学习率和批尺寸

[0304] 对每次网络训练独立地选择培训学习率和批尺寸。虽然较大的网络如Resnet V2 200在 $480 \times 480$ 的输入大小下可能会遇到批尺寸的存储限制,但使用网格搜索手段而实验性地确定了良好的学习率和批尺寸。上表3中示意了选择用于训练这些网络的超参数。

[0305] 模型

[0306] 在Tensorflow v.0中构建、训练和测试模型。所提供的训练基准是在NVIDIA® Tesla® P100GPU架构上进行的。

[0307] 通过多次训练迭代来训练超参数。在第一次网络训练之后,观察到当在训练数据中欠表示的特殊情况下网络表现不佳。在活动场所中跳到半空、奇怪的姿势和排尿的情况是通常观察到的故障。识别出这些困难的帧并且合并到训练数据集里以进一步改善性能。最终模型定义和训练超参数的完整说明在上表3中示意。

[0308] 图13A-13E分别示意了由所有三个网络表明的训练和验证损失曲线的图。总体而言,训练和验证损失曲线表明所有三个网络均训练到性能介于1至2个像素的平均误差之间。出乎意料的是,分箱分类网络显示出不稳定的损失曲线,表明过度拟合和验证的泛化性差(图13B、13E)。回归架构收敛到1.2个像素的验证误差,显示出比验证更好的训练性能(图13A、13B、13D)。然而,Resnet V2200——给出最佳结果的特征提取器——是大而深的网络,具有200多个层和6270万个参数并导致每帧的处理时间大大延长(33.6ms)。其他预构建的通用网络(Zoph,2017)可在较快的计算时间方面作出折衷的情况下取得类似或更差的性能。因此,回归网络是准确但计算上昂贵的方案。

[0309] 如图13A、13B、13C中进一步所示,编码器-解码器分割架构收敛到0.9个像素的验证误差。分割架构不仅性能良好,而且GPU运算的运算效率很高,平均处理时间为每帧5-6ms。在Nvidia® Tesla® P100(一种服务器级GPU)上,视频数据可被处理至200fps(6.7X实时),而在Nvidia® Titan Xp(一种消费级GPU)上,视频数据可被处理至125fps(4.2X实时)。这样高的处理速度很可能是由于该结构仅有18层深和1,060万个参数。

[0310] 还对训练集大小进行了基准测试来确定必要的经标注训练数据的相对规模以实现编码器-解码器分割网络架构的良好网络性能。通过改组和随机采样训练集的子集(例如,10,000、5,000、2,500、1000和500)测试了此基准。每个子采样训练集都经过训练并与相同的验证集进行比较。此基准测试的结果在图14A-14H中示出。

[0311] 通常,训练曲线似乎难以区分(图14A)。即,训练尺寸设置在训练集误差率上没有

表现出性能变化(图14A)。令人惊奇的是,验证性能在2,500个训练样本以上收敛到相同的值,但在1,000个训练样本以下误差增加(图14B)。如进一步所示,在超过2,500个训练样本时,验证准确性优于训练准确性(图14C-14F),而仅通过在1,000处匹配训练准确性,则验证准确性开始显示弱泛化的迹象(图14G)。仅使用500个训练样本显然是过度训练,如由发散且增加的验证误差率所示意(图14H)。这表明训练集不再大到足以允许网络很好地泛化。因此,可从仅使用2,500个带注释的图像进行训练的网络获得良好的结果,这将花大约3小时来由标注界面生成。相应地,虽然训练样本的确切数量最终将取决于视觉问题的难度,但建议的起始点应为约2,500个训练样本。

[0312] 图15A-15B示意了表明根据所公开的实施方案跟踪的小鼠在可见光下的示例性视频帧,而图15C-15D示意了在红外光下的示例性视频帧。如图所示,各个小鼠的空间范围被逐像素地进行颜色编码。

[0313] 考虑到其运算效率、准确性、训练稳定性以及所需训练数据的数量少,故选择了编码-解码器分割架构来预测整个视频的小鼠位置以便与其他手段进行比较。

[0314] 通过从具有不同毛色和数据收集环境的小鼠推断整个视频(图8A)并从视觉上评估跟踪的质量来评估基于神经网络的跟踪的质量。还将基于神经网络的跟踪与独立的跟踪模式——KOMP2光束中断系统(图8A,第6列)进行了比较。

[0315] 实验活动场所

[0316] a) 旷场活动场所

[0317] 采用活动场所200的一个实施方案作为旷场活动场所。旷场活动场所尺寸为52cm乘52cm。地板为白色PVC塑料,墙壁为灰色PVC塑料。为了帮助清洁维护,向所有内缘均添加2.54cm的白色斜切面。由LED灯环(型号:F&V R300)提供照明。校准光环以在每个活动场所中产生600勒克斯的光。

[0318] b) 24-小时监测旷场活动场所

[0319] 扩充旷场活动场所以进行多日测试。灯212呈顶置式LED照明的形式,设置为标准的12:12LD周期。将ALPHA-dri放入活动场所中作为垫料。为了提供食物和水,在活动场所中放置单个Diet Ge176A食物杯。监测该营养源并在耗尽时更换。每个矩阵白天以250勒克斯照射,夜间则以小于约500勒克斯照射。为了在夜间记录视频,灯212包括IR LED(940nm)照明。

[0320] c) KOMP旷场活动场所

[0321] 除了定制的活动场所外,也在市售系统上所公开的系统和方法的实施方案进行了基准测试。Opto-M4旷场笼使用透明塑料墙制成。这样,由于相应发生的反射,视觉跟踪变得非常困难。笼子尺寸为42cm乘42cm。该活动场所的照明由100-200勒克斯的LED照射完成。

[0322] 视频采集

[0323] 通过结合图2和7讨论的视频采集系统的实施方案来采集所有视频数据。视频数据使用Sentech相机(型号:STC-MB33USB)形式的相机210和Computar镜头(型号:T3Z2910CS-IR)在640×480像素分辨率、8位单色深度和约29fps(例如,约29.9fps)下获取。使用190/255的目标亮度以数字方式控制曝光时间和增益。将光圈调整到其最宽,以便使用较低的模拟增益来达到目标亮度。这继而减小基线噪声的放大。使用“原始视频”编解码器和“pal8”像素格式将文件临时保存在本地硬盘驱动器上。分析运行大约2小时,产生大约50GB的原始

视频文件。晚上,使用ffmpeg软件来应用 $480 \times 480$ 像素裁剪去噪滤波器,并使用mpeg4编解码器(质量设置为最大)进行压缩,压缩后的视频大小为大约600MB。

[0324] 将相机210安装到框架202上、在架子202b上方大约100cm处以减轻透视变形。手动设置缩放和焦点以实现8像素/厘米的缩放。此分辨率既可最大限度地减少活动场所边界上未使用的像素,又可产生每只小鼠大约800的像素面积。尽管KOMP活动场所略小,但仍使用了8像素/厘米目标的相同缩放。

[0325] 使用编码器-解码器分割神经网络从KOMP2数据集跟踪了2002个视频(总计700小时),结果示意于图8中。这些数据包括在20分钟旷场试验中测试的C57BL/6NJ背景中的232只基因敲除系。由于基材透明,每个KOMP2活动场所的背景略有不同,故对8个测试腔室中的每一个(平均 $n=250$ ,图16)和所有组合盒比较了跟踪性能。在KOMP2使用的所有8个测试腔室中,两种方法之间在旷场中行进的距离之间观察到非常高的相关性( $R=96.9\%$ )。观察到与这种趋势高度不一致的两只动物(红色箭头)。观察视频显示,两只动物都表现出奇怪的姿势,一只动物步态蹒跚,另一只则驼着背。据信蹒跚和驼背步态会导致异常的光束中断,从而导致从光束中断系统测得行进过错误地高的总距离。此实例强调了神经网络的优点之一:不受动物姿势的影响。

[0326] 也将经过训练的分割神经网络的性能与Ctrax就来自上文结合图8A讨论的各种测试环境和毛色的各种视频进行了比较。进行Ctrax比较的动机是出于多种原因。一方面,Ctrax被认为是最好的常规跟踪器之一,它允许对许多跟踪设置进行微调。Ctrax也是开源的并提供用户支持。给定使用BGSLibrary的结果(图8B),预期将从其他跟踪器得到类似或最差的性能。使用经过训练的分割神经网络以及Ctrax跟踪了每组12只动物。如下所述,对于72个视频中的每一个,都对Ctrax的设置进行了微调。

[0327] Ctrax包含各种设置以优化跟踪能力(Branson,2009)。该软件的作者强烈建议确保按特定的标准设置活动场所以确保良好的跟踪。在本文讨论的大多数测试中(例如,白色背景上的白化小鼠),都采用了Ctrax不能在其上良好运行的环境。但是,通过精心调整参数,仍可实现良好的性能。随着大量的设置来操作,Ctrax可以很容易地成为大的时间成本以实现良好的跟踪性能。设置Ctrax以便在所公开的环境中跟踪小鼠的方案如下。

[0328] 第一项操作是创建背景模型。Ctrax的核心基于背景减除,故拥有稳健的背景模型对于功能是必不可少的。当小鼠移动时,模型可发挥最佳功能。为了创建背景模型,需要找一段其中小鼠在清晰地移动的视频并从该段中抽取帧的样本。这将确保小鼠不被包含在背景模型中。这种手段将极大地改善Ctrax对24小时数据的跟踪性能,因为小鼠不经常移动而通常会被结合到背景模型中。

[0329] 第二项操作是设置背景减除的设置值。这里,背景亮度归一化方法使用的Std范围为254.9至255.0。用于分割出小鼠的阈值在每个视频的基础上进行调整,因为曝光和毛色的轻微改变都会影响性能。为了调整这些阈值,应用一组良好的起始值并对视频进行检查以确保总体上良好的性能。在某些实施方案中,可检查每个视频是否有小鼠用后腿直立于墙上的情况,因为由于阴影,这些通常是最难跟踪的帧。另外,可应用形态过滤来去除环境中的稀疏变化,以及去除小鼠尾巴以便拟合椭圆。采用的打开半径为4、闭合半径为5。

[0330] 在进一步的操作中,手动调整Ctrax启用的各种跟踪参数以确保观察结果事实上是小鼠。在考虑时间的情况下,对这些参数进行一次充分的调整,并然后用于被跟踪的所有

其他小鼠。如果视频的效果明显不佳,则稍稍调整常规设置以改善性能。对于形状参数,从一个一个单独的黑色小鼠视频确定基于2个标准偏差的界限。由于预期某些小鼠在分割步骤中会表现不良,因此进一步降低最小值。尽管无法分割整个小鼠,但这允许Ctrax仍能找到小鼠的正确位置。这种方法效果很好,因为所有装置都具有相同的8倍缩放并且测试的小鼠通常具有相同的形状。运动设置非常宽泛,因为实验装置仅跟踪活动场所中的一只小鼠。在观察参数下,主要采用“最小面积忽略”,它会过滤掉较大的检测结果。这里,大于2500的检测结果将被过滤掉。在事后检查选项卡下,使用“修改伪检测”设置来去除长度小于500帧的检测。

[0331] 由于Ctrax无法产生有效的背景模型,故通过比较,从24小时装置手动略去其中动物整个小时都在持续睡觉的视频。计算了在Ctrax和神经网络之间行进的总距离的累积相对误差并示意于(图17A)中。对于视频中的每一分钟,都对来自神经网络和Ctrax二者的行进距离预测加以比较。此度量指标量度每只小鼠的质心跟踪的准确性。跟踪黑色、灰色和花斑小鼠显示出小于4%的误差。然而,在白化(14%)、24-小时活动场所(27%橙色)和KOMP2(10%,蓝色)中看到显著较高的误差水平(图17A)。因此,在没有神经网络跟踪器的情况下就无法充分跟踪白化、KOMP2或24-小时数据。

[0332] 还观察到,当前景分割预测不正确时,如当预测中包括阴影时,椭圆拟合将不能正确地表示小鼠的姿势。在这些情况下,即使质心跟踪是可接受的,椭圆拟合本身也是高度可变的。

[0333] 用于行为识别的现代机器学习软件,如JAABA (Kabra, 2013),利用这些特征进行行为分类。通过短轴的相对标准偏差来量化椭圆跟踪中的方差并示意于图17B中。该度量指标显示了所有实验室小鼠间的最小方差,因为当跟踪是准确的时,通过行为分析中表达的各种姿势,单独的小鼠的宽度仍保持相似。即使在总行进距离的误差累积相对误差较低的情况下,对灰色和花斑小鼠(图17A)也观察到高的跟踪变化(图17B)。正如预期的那样,对白化和KOMP2跟踪,观察到短轴高的相对标准偏差。因此,对于质心跟踪和椭圆拟合方差二者,均发现神经网络跟踪器性能优于传统跟踪器。

[0334] 在建立了编码器-解码器分割神经网络作为高精度跟踪器后,用两个大的行为数据集进一步测试了其性能。用横跨58个小鼠品系的1845只小鼠生成了旷场视频数据(1691小时),其中包括所有各种颜色、花斑鼠、裸鼠和肥胖鼠。该数据集包括47个自交品系、11个同基因型F1小鼠品系,并且是根据Mouse Phenome Database of Bogue, 2018生成的最大旷场数据集。

[0335] 图18A中示意了对行进的总距离的跟踪结果。点表示品系中的个体,方框表示平均值 $\pm$ 标准偏差。使用单个经过训练的网络,无需任何用户调整,就可以高精度跟踪所有小鼠。目视检查来自大多数品系的小鼠的跟踪保真度并观察到优异的性能。观察到的活动表型与先前公开的小鼠旷场行为数据集一致。

[0336] 采用相同的神经网络来跟踪对四只C57BL/6J和两只BTBR  $T^+1tpr3^{tf}$ /J小鼠收集的24-小时视频数据(图8A,第5列)。这些小鼠在具有垫料、食物和水杯的场所中于12:12明-暗条件下圈养了多天,在此期间食物改变了位置。使用可见和红外光源记录视频数据。使用同一网络在这些条件下跟踪所有动物的活动,在明暗条件下都观察到非常好的性能。

[0337] 结果示意于图18B中,其分别用8个亮点表示亮状态,用暗点表示暗状态。正如预期

的那样,在暗阶段观察到运动活动水平高的活动节律(曲线)。

[0338] 总之,在复杂环境中基于视频跟踪动物一直是动物行为领域的长期挑战(Egnor, 2016)。现有技术系统不能解决动物分割的根本问题并严重依赖于前景和背景之间的视觉对比度来进行精确跟踪。结果,用户必须限制环境以获得最佳结果。

[0339] 本文描述了能够在复杂且动态的环境中起作用的基于现代神经网络的跟踪器和相应的使用方法。通过使用可训练的神经网络,解决了跟踪、前景和背景分割中的根本问题。对三种不同架构的测试发现,编码器-解码器分割网络可实现高度的准确性并在高速度(超过6X实时)下发挥功能。

[0340] 还提供了标注界面,该界面允许用户通过标注少至2,500张图像(这大约花3小时)来针对其特定的环境训练新的网络。

[0341] 已将所公开的经过训练的神经网络与两个现有方案进行了比较并发现它在复杂环境中的性能大大优于后者。任何利用背景减除手段的现成系统都有望获得类似的结果。实际上,在测试26种不同的背景减除方法时,发现每种方法都会在某些情况下失败。然而,单个神经网络架构能够在多种环境下对所有毛色的小鼠起作用,而无需进行微调或用户输入。这种机器学习方法可在动态环境条件下以最少的用户输入实现长期跟踪,从而构成用于行为研究的下一代跟踪架构的基础。

[0342] 本文描述的控制系统的的一个或多个方面或特征可在数字电子电路、集成电路、专门设计的应用专用集成电路(ASIC)、现场可编程门阵列(FPGA)计算机硬件、固件、软件和/或其组合中实现。这些各个方面或特征可包括在一个或多个计算机程序中的实施,所述程序可在包括至少一个可编程处理器的可编程系统上执行和/或解读,所述可编程处理器可以是专用的或通用的,其耦合以从存储系统、至少一个输入设备和至少一个输出设备接收数据和指令及向存储系统、至少一个输入设备和至少一个输出设备发送数据和指令。可编程系统或计算机系统可包括客户端和服务端。客户端和服务端通常彼此远离并通常通过通信网络进行交互。客户端和服务端的关系是通过在相应计算机上运行并彼此具有客户端-服务端关系的计算机程序产生的。

[0343] 计算机程序,其也可称为程序、软件、软件应用程序、应用程序、组件或代码,包括用于可编程处理器的机器指令,并可以以高级程序语言、面向对象的编程语言、功能性编程语言、逻辑编程语言和/或汇编/机器语言来实施。如本文所用,术语“机器可读介质”是指用来向可编程处理器提供机器指令和/或数据的任何计算机程序产品、装置和/或设备,如磁盘、光盘、存储器和可编程逻辑设备(PLD),包括以机器可读信号接收机器指令的机器可读介质。术语“机器可读信号”是指用来向可编程处理器提供机器指令和/或数据的任何信号。机器可读介质可非暂态地存储此类机器指令,例如像非暂态固态存储器或磁性硬盘驱动器或任何等效存储介质将做的那样。机器可读介质可替代地或另外地以瞬时方式存储此类机器指令,例如像处理器缓存或与一个或多个物理处理器内核相关联的其他随机存取存储器将做的那样。

[0344] 为了提供与用户的交互,本文所述主题的一个或多个方面或特征可在具有显示设备的计算机上实施,显示设备例如为阴极射线管(CRT)或液晶显示器(LCD)或用于向用户显示信息的发光二极管(LED)监视器以及用户可通过它们向计算机提供输入的键盘和点击设备(例如,鼠标、轨迹球等)。也可使用其他种类的设备来提供与用户的交互。例如,提供给用

户的反馈可以是任何形式的感觉反馈,例如视觉反馈、听觉反馈或触觉反馈;并且来自用户的输入可以以任何形式接收,包括但不限于声音、语音或触觉输入。其他可能的输入设备包括但不限于触摸屏或其他触敏设备,如单点或多点电阻式或电容式触控板、语音识别硬件和软件、光学扫描仪、光学指示器、数字图像捕获设备及相关解释软件等。

[0345] 整个本申请引用的所有参考文献,例如专利文件,包括已发布或授权的专利或等同物、专利申请出版物及非专利文献文件或其他原始资料,通过引用全文并入本文,就好像一个一个单独地通过引用并入一样,至每个参考文献至少部分地不与本申请中的公开内容不一致的程度。例如,除了参考文献的部分不一致的部分外,通过引用并入了部分不一致的参考文献。

[0346] 当在本文中使用时,马库什组或其他分组,该组的所有单独的成员以及该组的所有可能的组合和子组合意在单独地包括在本公开中。

[0347] 如本文所用,单数形式“一个”、“一种”和“该”包括复数指代,上下文另有明确指除外。因此,例如,“一个细胞”的提及包括本领域技术人员已知的多个这样的细胞及其等同物,等等。另外,术语“一个”(或“一种”)、“一个或多个/一种或多种”和“至少一个/至少一种”在本文中可互换使用。

[0348] 如本文所用,术语“包含”与“包括”、“具有”、“含有”和“特征在于”同义并各自可互换使用。这些术语中的每一个还是包含性或开放性的而不排除另外的未引用的要素或方法步骤。

[0349] 如本文所用,术语“由……组成”不包括未在权利要求要素中指定的任何要素、步骤或成分。

[0350] 如本文所用,术语“基本上由……组成”不排除不会实质性影响权利要求的基本和新型特征的材料或步骤。在本文的每个实例中,术语“包括”、“基本上由……组成”和“由……组成”中的任何一个都可用另外两个术语中的任一个替换。

[0351] 本文说明性地描述的实施方案可在不存在本文未具体公开的任何一个或多个要素、一个或多个限制的情况下适当地实践。

[0352] 表述“权利要求XX-YY中的任何一项”(其中XX和YY表示权利要求编号)旨在提供替代形式的多个从属权利要求并且在一些实施方案中可与表述“如权利要求XX-YY中的任何一项”互换。

[0353] 除非另有定义,否则本文使用的所有技术和科学术语具有与所公开的实施方案所属领域普通技术人员通常所理解的相同的含义。

[0354] 任何时候说明书中给出范围例如温度范围、时间范围、组成范围或浓度范围时,所有中间范围和子范围以及包括在给定范围内的所有单个值均意在包括在本公开中。如本文所用,范围具体包括作为范围的端点值提供的值。例如,1至100的范围具体包括端点值1和100。应理解,本文的描述中包括的范围或子范围中的任何子范围或单个值可从本文的权利要求排除。

[0355] 在以上描述和权利要求中,可能出现表述如“至少一个”或“一个或多个”,其后是要素或特征的联合列表。也可能在两个或更多个要素或特征的列表中出现术语“和/或”。除非与使用其的上下文隐性或显性地冲突,否则这样的表述旨在指任何单独列出的要素或特征或者任何所述要素或特征与任何其他所述要素或特征的组合。例如,表述“A和B中的至

少一个”、“A和B中的一个或多个”和“A和/或B”分别意在指“单独的A、单独的B或者A和B一起”。类似的解释也适用于包含三个或更多个项目的列表。例如,表述“A、B和C中的至少一个”、“A、B和C中的一个或多个”和“A、B和/或C”分别意在指“单独的A、单独的B、单独的C、A和B一起、A和C一起、B和C一起、或者A和B和C一起”。另外,在上文和权利要求书中术语“基于”的使用意在指“至少部分地基于”,从而也允许未述及的特征或要素。

[0356] 本文已采用的术语和表达用作描述性术语而不是限制性的,并且无意于使用此类术语和表达排除所示和所述特征或其部分的任何等同物,但应认识到,可在要求保护的实施方案的范围内作各种修改。因此,应理解,尽管本申请可包括对优选实施方案、示例性实施方案和任选特征的讨论,但本领域技术人员可对本文公开的构思作修改和改变。这样的修改和改变被认为在所公开的实施方案的范围内,如由附随的权利要求所限定。本文提供的特定实施方案是本公开的可用实施方案的示例,并且对于本领域技术人员将显而易见的是,它们可使用本说明书中阐述的设备、设备部件和方法步骤的大量变型来执行。对于本领域技术人员将显而易见的是,方法和可用于本方法的设备可包括大量任选的组成及处理元件和步骤。

[0357] 本公开的实施方案可在不脱离其精神或实质性特征的情况下以其他特定的形式实施。因此,前述实施方案在所有方面都应认为是示意性的而不是要限制本文所述的主体。

[0358] 参考文献

[0359] 下面列出的每个参考文献均通过引用全文并入本文。

[0360] • Bargmann,C.,Newsome,W.,Anderson,A.,Brown,E.,Deisseroth,K.,Donoghue,J.,MacLeish,P.,Marder,E.,Normann,R.,and Sanes,J.,BRAIN 2025:a scientific vision.Brain Research Through Advancing Innovative Neurotechnologies (BRAIN) Working Group Report to the Advisory Committee to the Director,NIH(2014).

[0361] • Bogue,M.A.et al.,“Mouse Phenome Database:an integrative database and analysis suite for curated empirical phenotype data from laboratory mice,”Nucleic Acids Res,46,D843-D850(2018).

[0362] • Branson,K.,Robie,A.A.,Bender,J.,Perona,P.&Dickinson,M.H.,“High-throughput ethomics in large groups of Drosophila,”Nature methods,6(6),451-457(2009).

[0363] • Branson,K.and Belongie,S.,“Tracking Multiple Mouse Contours (Without Too Many Samples),”2005IEEE Computer Society Conference on Computer Vision and Pattern(CVPR’05),June 20-25,San Diego,CA,pp.1039-1046(2005).

[0364] • Decoster,W.,and Debruyne,F.,“The ageing voice:changes in fundamental frequency,waveform stability and spectrum,”Acta Otorhinolaryngol Belg 51,pp.105-112(1997).

[0365] • Deschenes,M.,Moore,J.,and Kleinfeld,D.,“Sniffing and whisking in rodents,”Curr Opin Neurobiol,22,pp.243-250(2012).

[0366] • Egnor,S.E.and Branson,K.,“Computational Analysis of Behavior,”Annu Rev Neurosci,39,pp.217-236(2016).

[0367] • Gomez-Marin,A.,Paton,J.J.,Kampff,A.R.,Costa,R.M.,and Mainen,Z.F.,

“Big behavioral data: psychology, ethology and the foundations of neuroscience,” *Nat Neurosci*, 17, pp.1455-1462 (2014).

[0368] • Goodfellow, I., Bengio, Y., and Courville, A. Deep Learning, MIT Press (2016)

[0369] • Grimsley, J.M., Monaghan, J.J., and Wenstrup, J.J., “Development of social vocalizations in mice,” *PLoS One*, 6, e17460 (2011).

[0370] • He, K., Zhang, X., Ren, S. & Sun, J. in European Conference on Computer Vision Springer, pp.630-645 (2016).

[0371] • Heckman, J.J., Proville, R., Heckman, G.J., Azarfar, A., Celikel, T., and Englitz, B., “High-precision spatial localization of mouse vocalizations during social interaction,” *Sci Rep*, 7(7), 3017 (2017).

[0372] • Kabra, M., Robie, A.A., Rivera-Alba, M., Branson, S. & Branson, K., “JAABA: interactive machine learning for automatic annotation of animal behavior,” *Nat Methods*, 10, pp.64-67 (2013).

[0373] • Kalueff, A.V., LaPorte, J.L., and Bergner, C.L., Neurobiology of grooming behavior Cambridge; New York: Cambridge University Press (2010).

[0374] • Krizhevsky, A., Sutskever, I., and Hinton, G.E., “ImageNet classification with deep convolutional neural networks,” Advances in neural information processing systems 25 (NIPS 2012).

[0375] • Kuleskaya, N. & Voikar, V., “Assessment of mouse anxiety-like behavior in the light-dark box and open-field arena: role of equipment and procedure,” *Physiol Behav* 133, 30-38 (2014).

[0376] • Kumar, V., Kim, K., Chryshanthi, J., Thomas, L.C., Hong, H., and Takahashi, J.S., “Second-generation high-throughput forward genetic screen in mice to isolate subtle behavioral mutants,” *Proceedings of the National Academy of Sciences*, 108 (Supplement 3), pp.15557-15564 (2011).

[0377] • LeCun, Y., Bengio, Y. & Hinton, G., “Deep learning,” *Nature* 521, 436-444 (2015).

[0378] • Martins, R.H., Goncalvez, T.M., Pessin, A.B., and Branco, A., “Aging voice: presbyphonia,” *Aging Clin Exp Res*, 26, pp.1-5 (2014).

[0379] • Mueller, P.B., “The aging voice,” *Semin Speech Lang*, 18(2), pp.159-169 (1997).

[0380] • Neunuebel, J.P., Taylor, A.L., Arthur, B.J., and Egnor, S.E., “Female mice ultrasonically interact with males during courtship displays,” *Elife*, 4: e06203, (2015).

[0381] • Ohayon, S., Avni, O., Taylor, A.L., Perona, P., and Roian Egnor, S.E., “Automated multi-day tracking of marked mice for the analysis of social behaviour,” *J. Neurosci. Methods*, 219(1), pp.10-19 (2013).

[0382] • Shannon, C.E., “Communication in the presence of noise,” *Proceedings*

of the IRE,37(1),pp.10-21(1949).

[0383] • Sobral,A.,“BGSLibrary:An OpenCV C++Background Subtraction Library,” in IX Workshop de Visão Computacional(WVC 2013),Vol.2 7(2013).

[0384] • Valdez,P.&Mehrabian,A.,“Effects of color on emotions,”Journal of experimental psychology:General,123(4),pp.394-409(1994).

[0385] • Wiltschko,A.B.,Johnson,M.J.,Iurilli,G.,Peterson,R.E.,Katon,J.M.,Pashkovski,S.L.,Abraira,V.E.,Adams,R.P.,and Datta,S.R.,“Mapping Sub-Second Structure in Mouse Behavior,”Neuron,88,pp.1121-1135(2015).

[0386] • Zhang,C.,Florêncio,D.,Ba,D.E.,and Zhang,Z.,“Maximum likelihood sound source localization and beamforming for directional microphone arrays in distributed meetings,”IEEE Transactions on Multimedia,10,pp.538-548(2008).

[0387] • Zoph,B.,Vasudevan,V.,Shlens,J.&Le,Q.V.,“Learning transferable architectures for scalable image recognition,”arXiv preprint arXiv:1707.07012(2017).

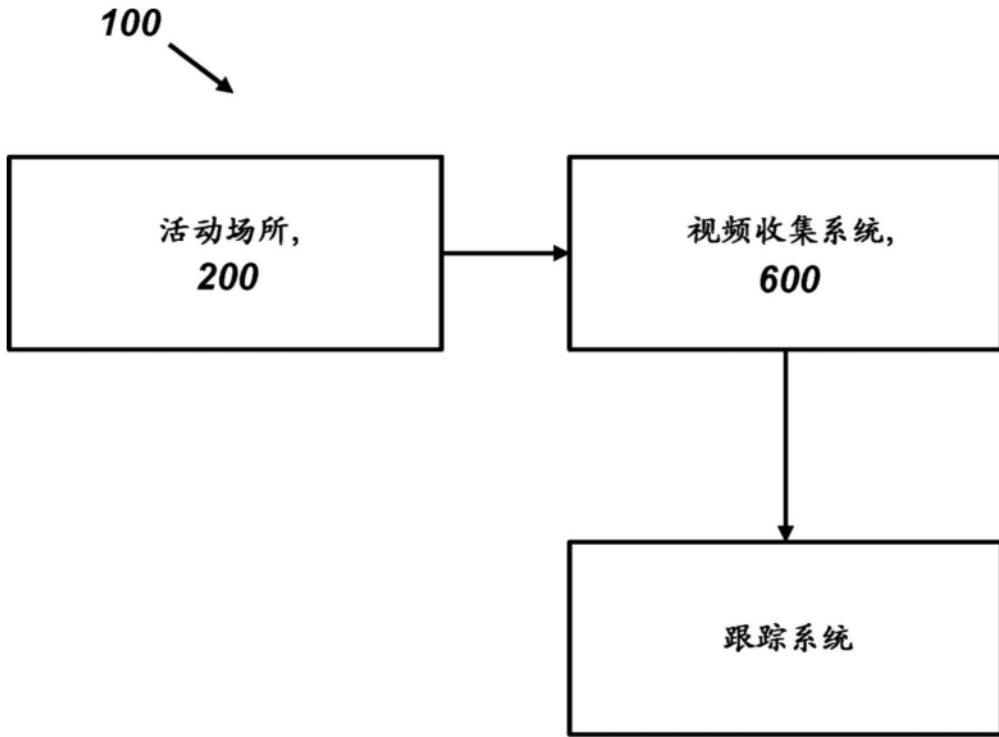


图1

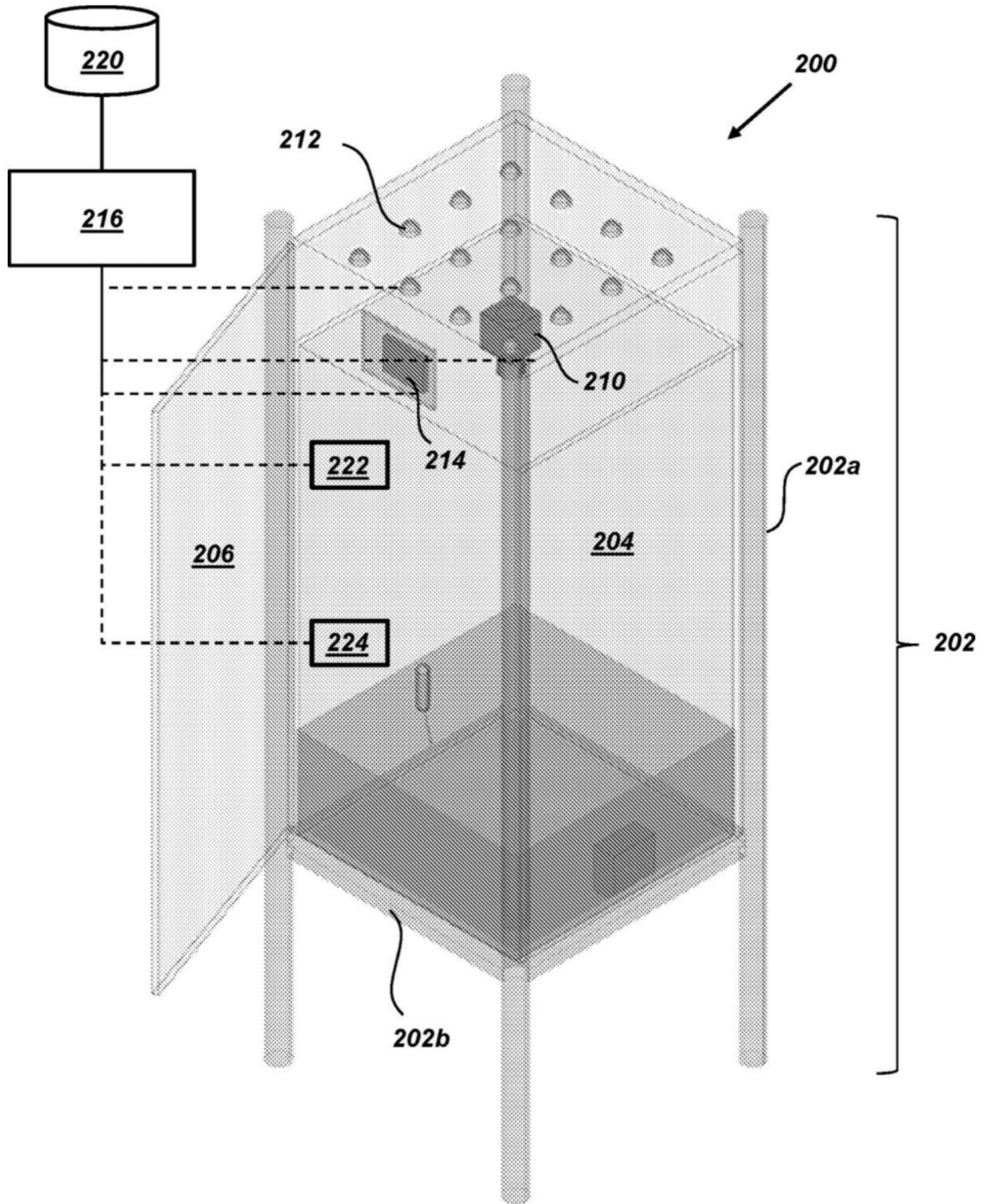


图2

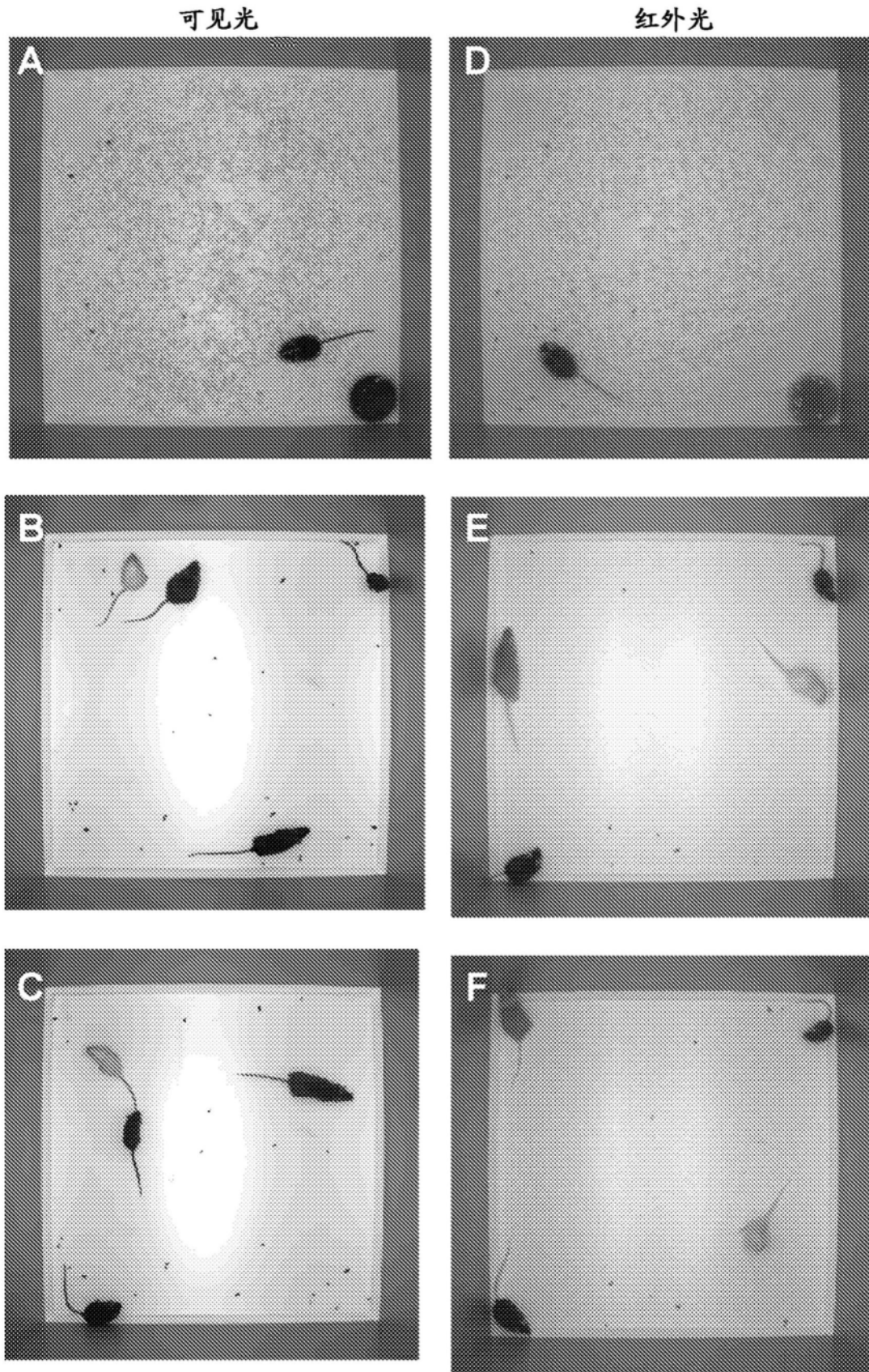


图3

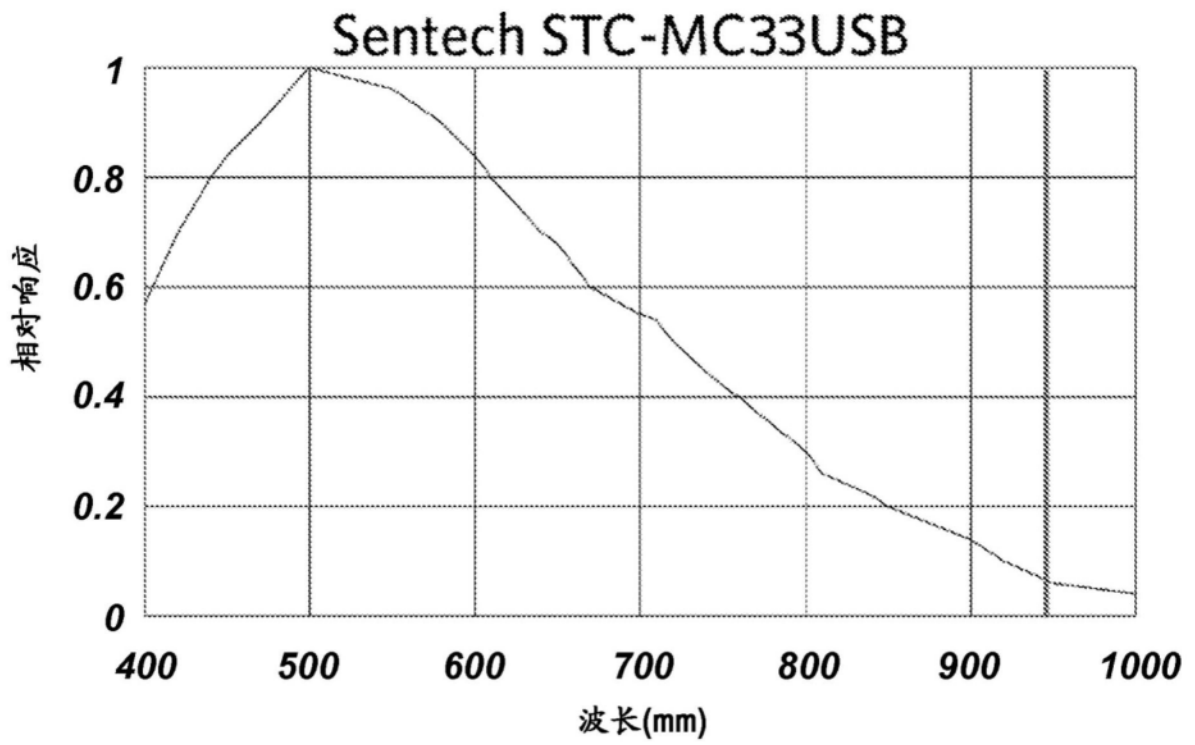


图4A

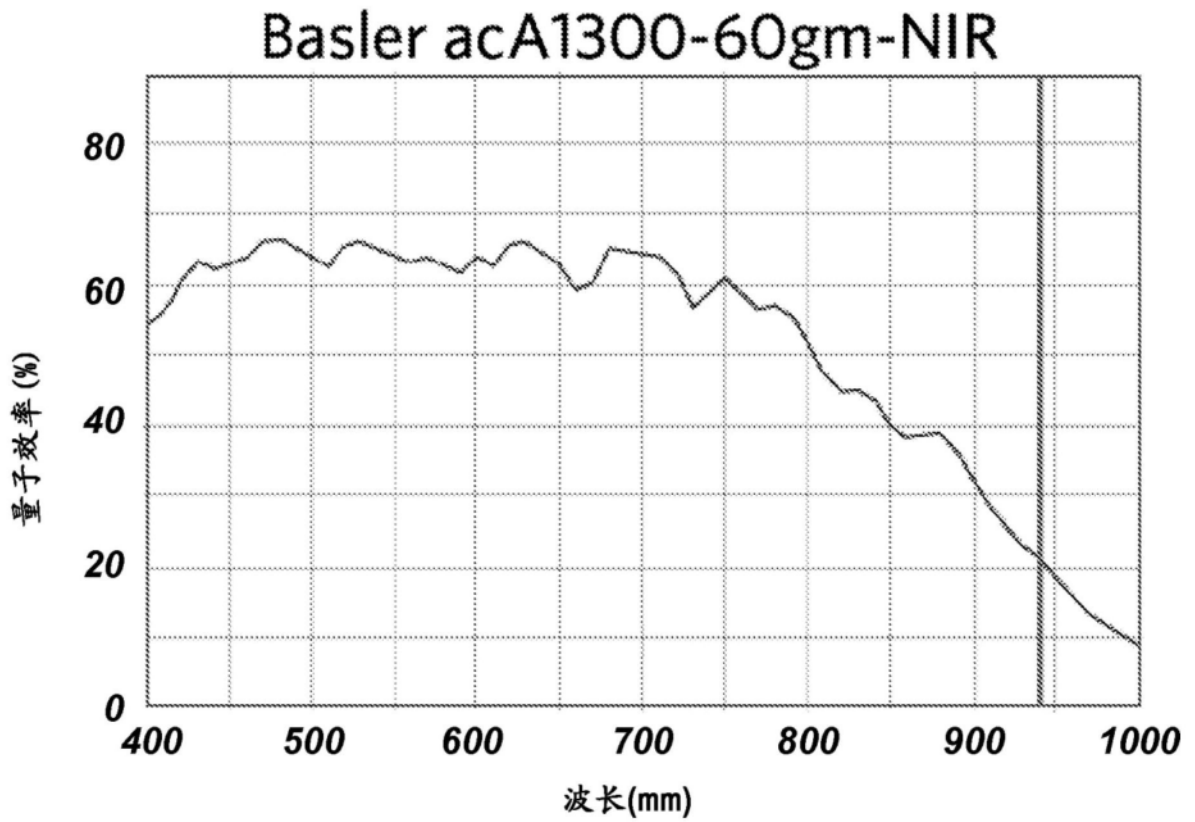


图4B

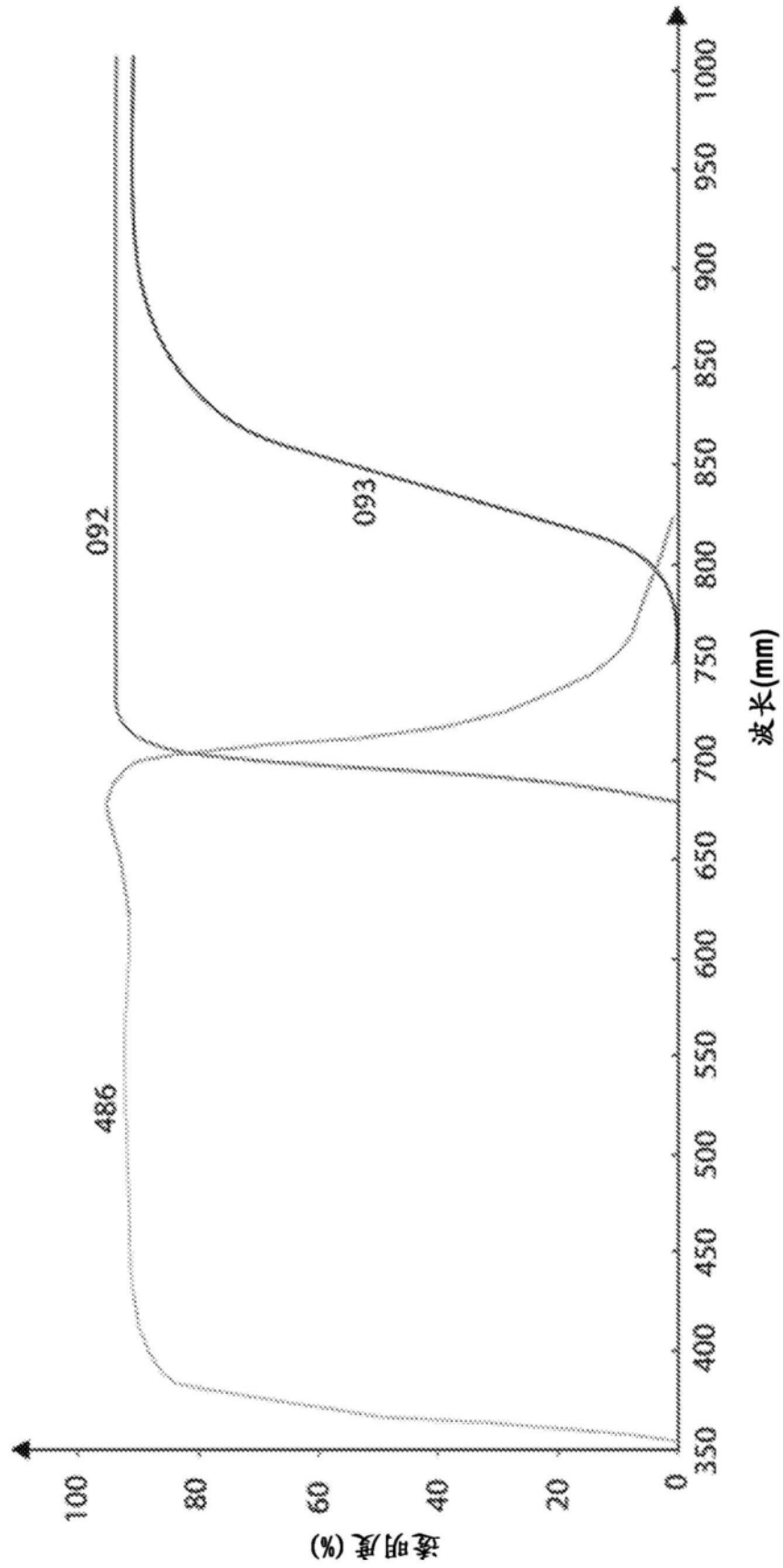


图5

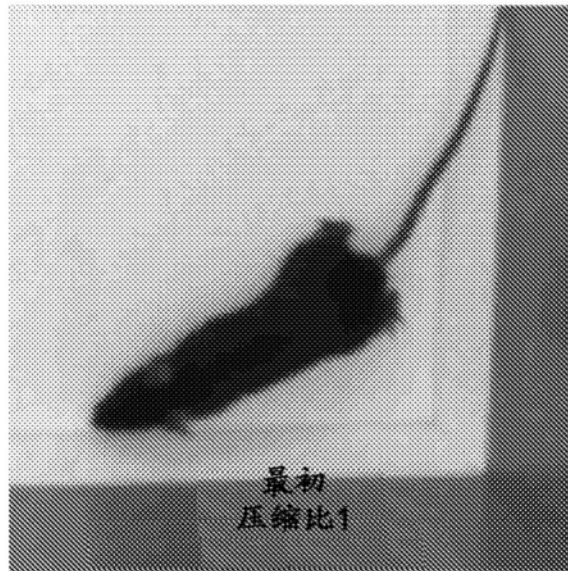


图6A

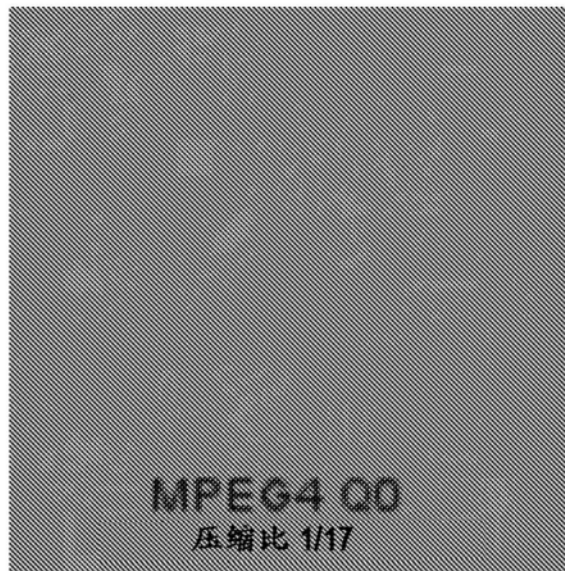


图6B

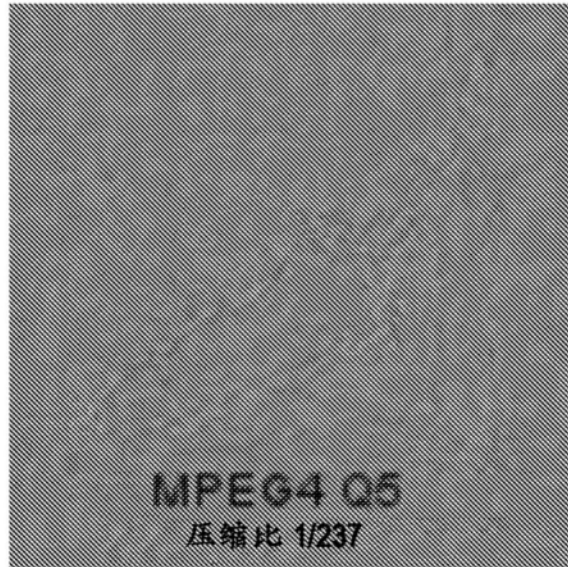


图6C

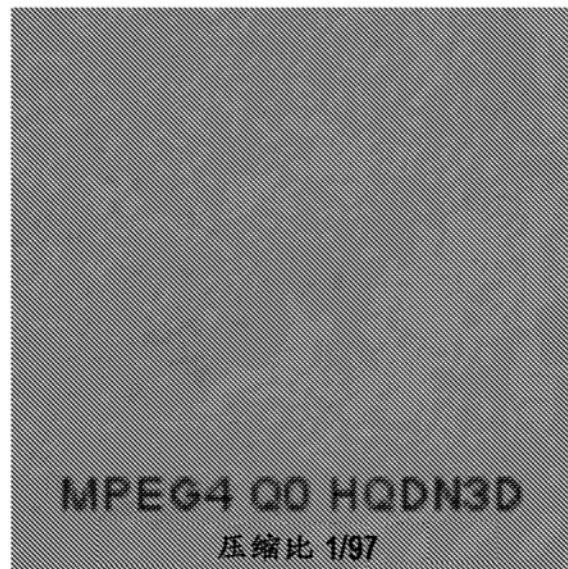


图6D

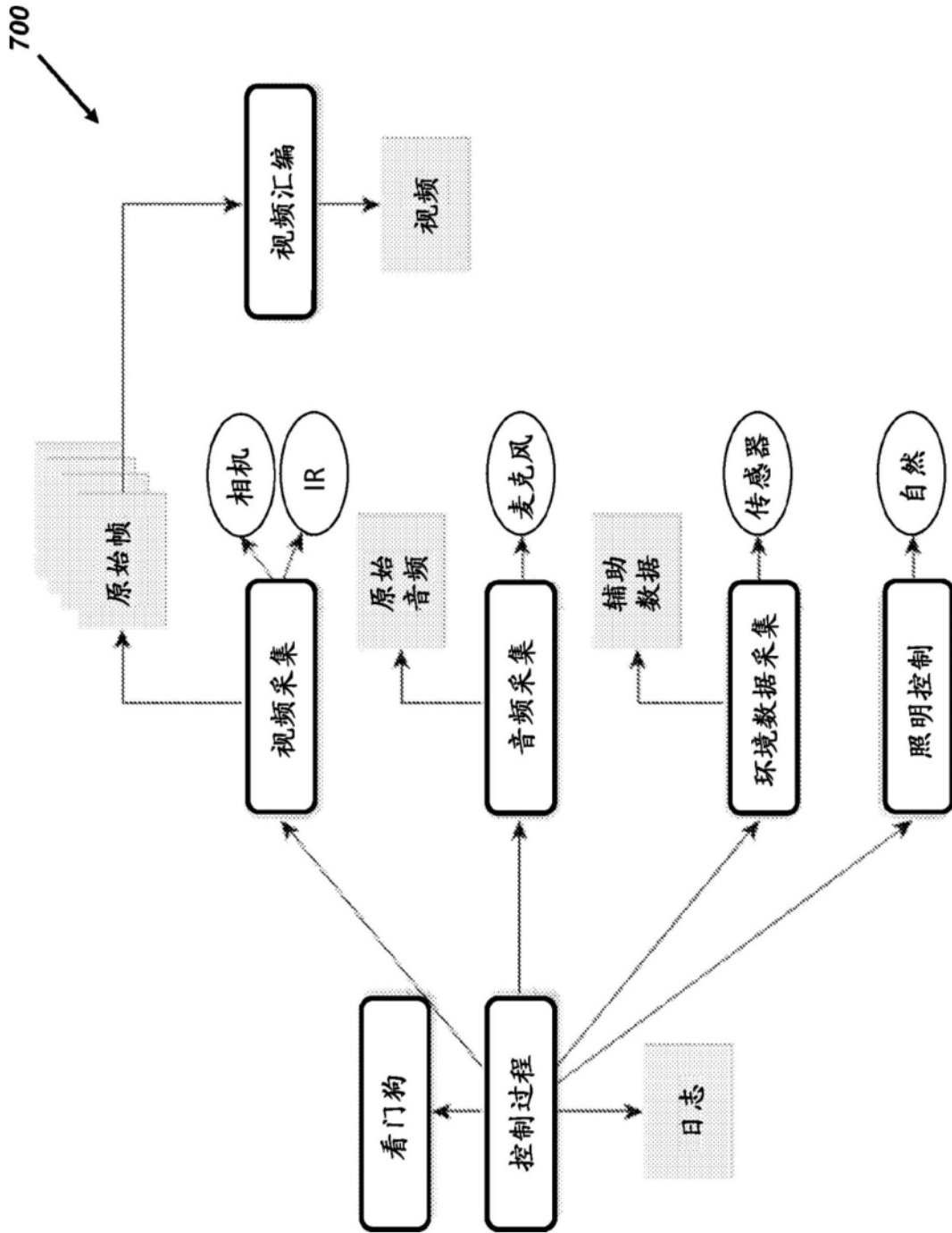


图7

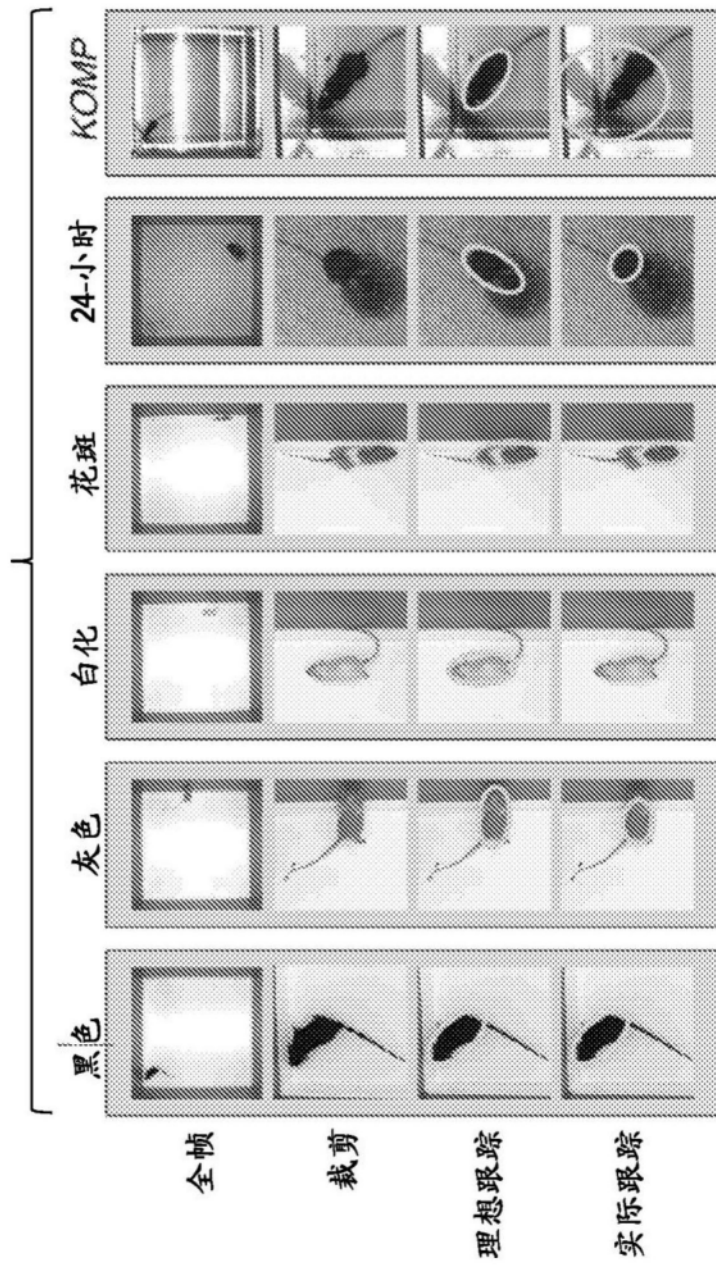


图8A

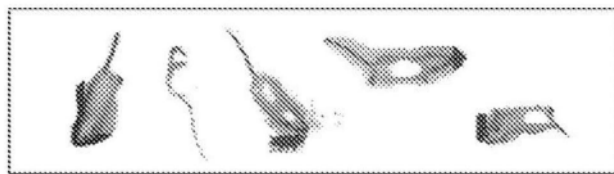
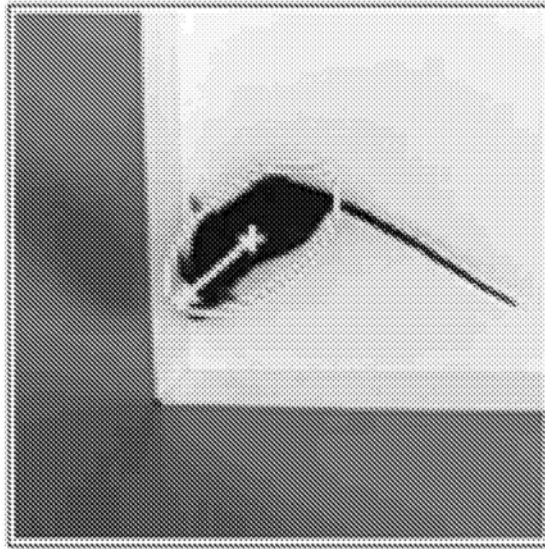


图8B



X-坐标, Y-坐标, 长轴, 短轴, 角度

图8C

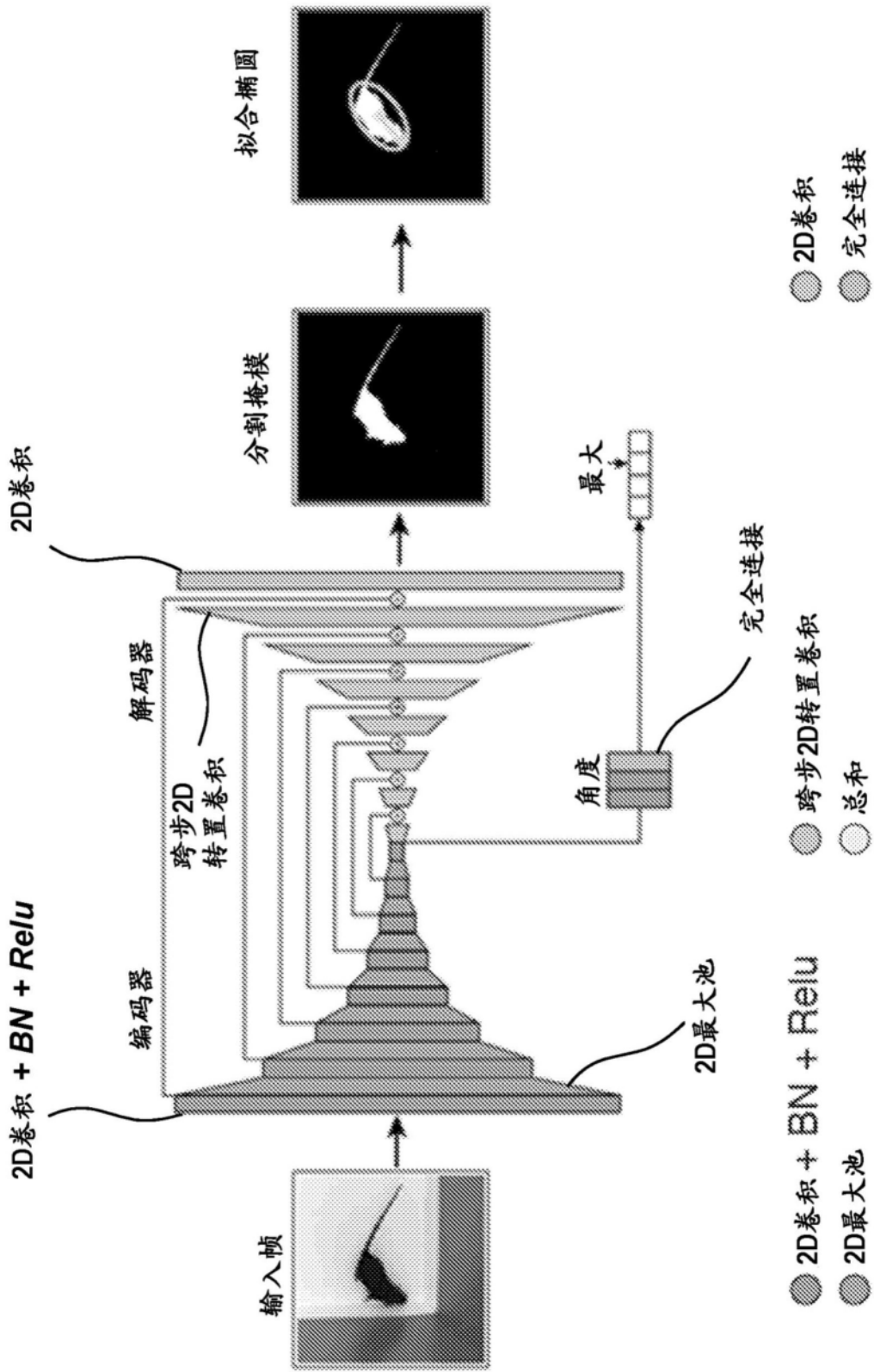


图9

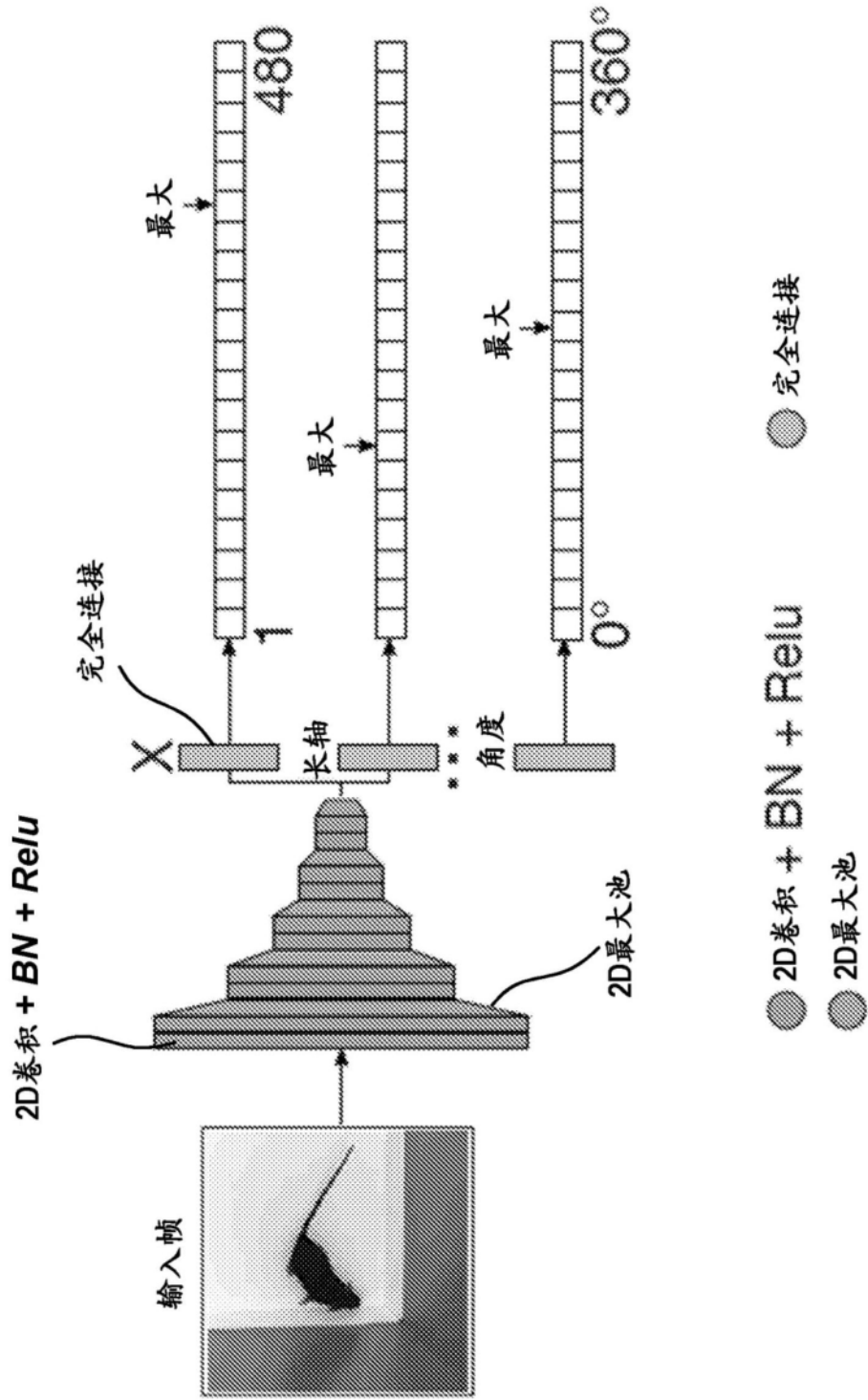


图10

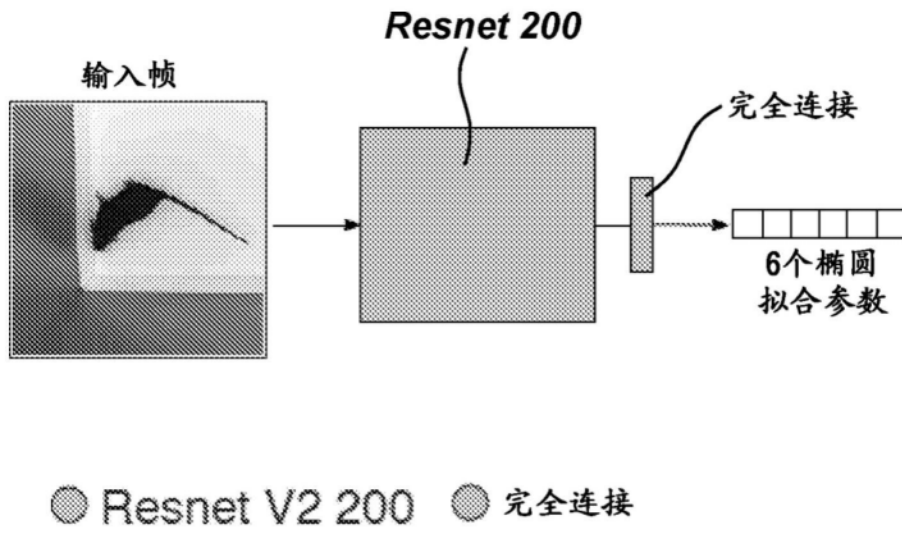


图11

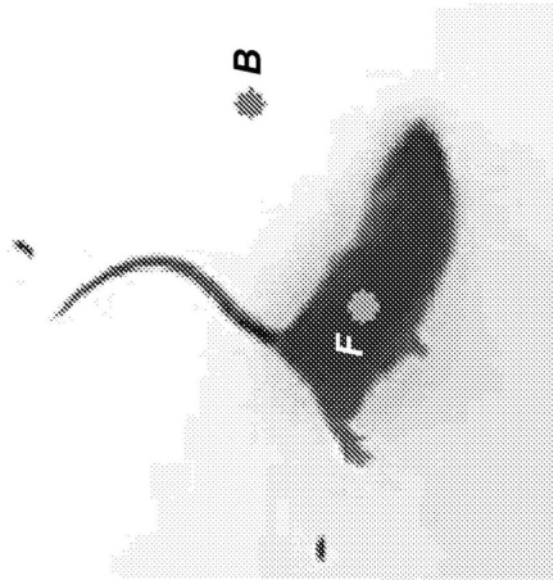


图12A

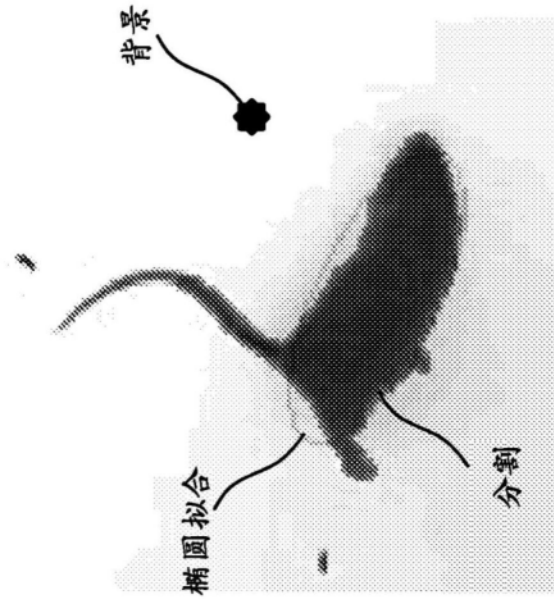


图12B

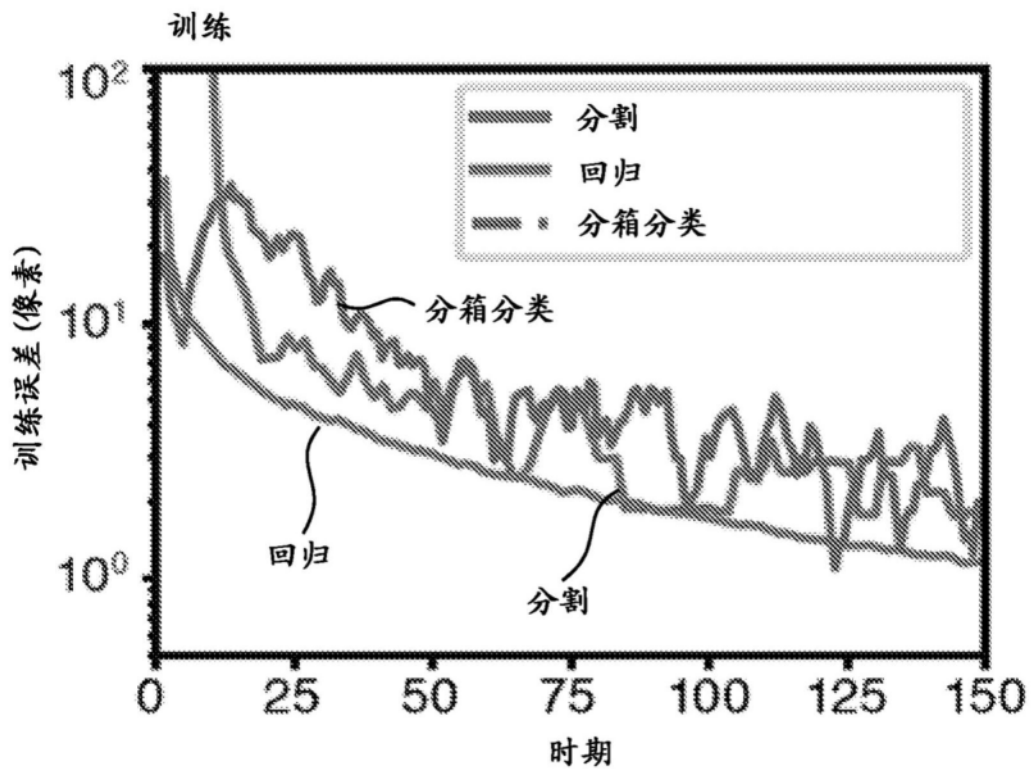


图13A

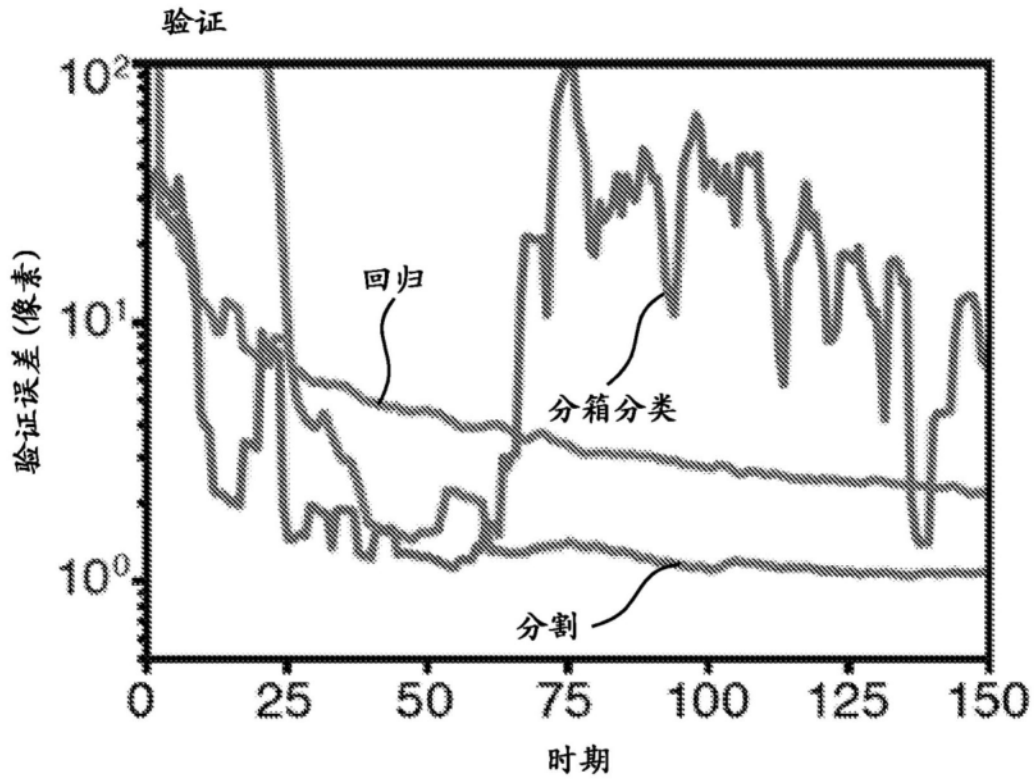


图13B

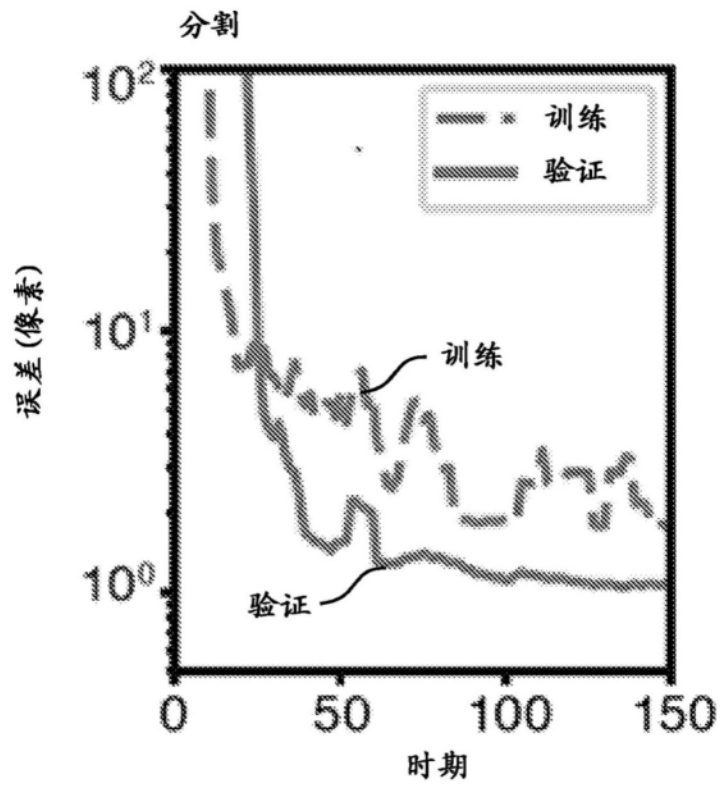


图13C

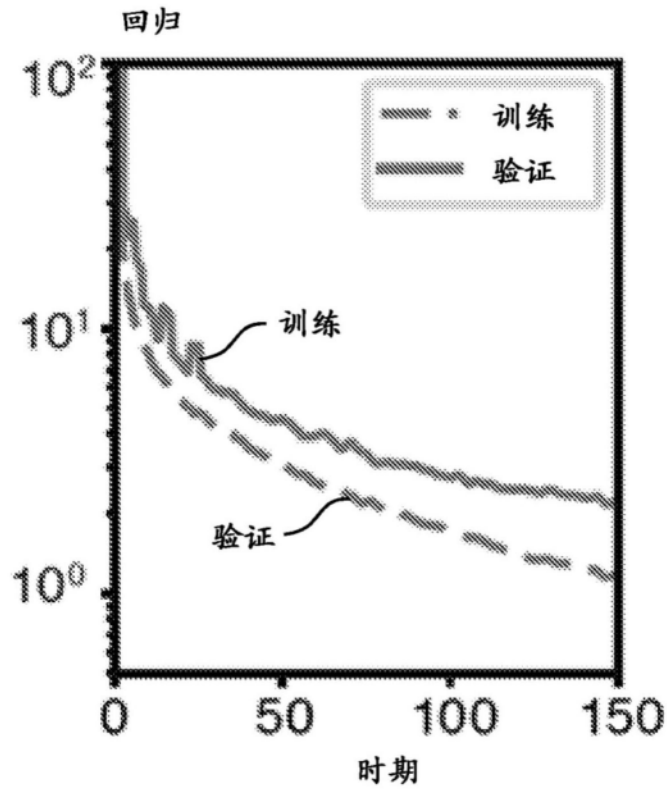


图13D

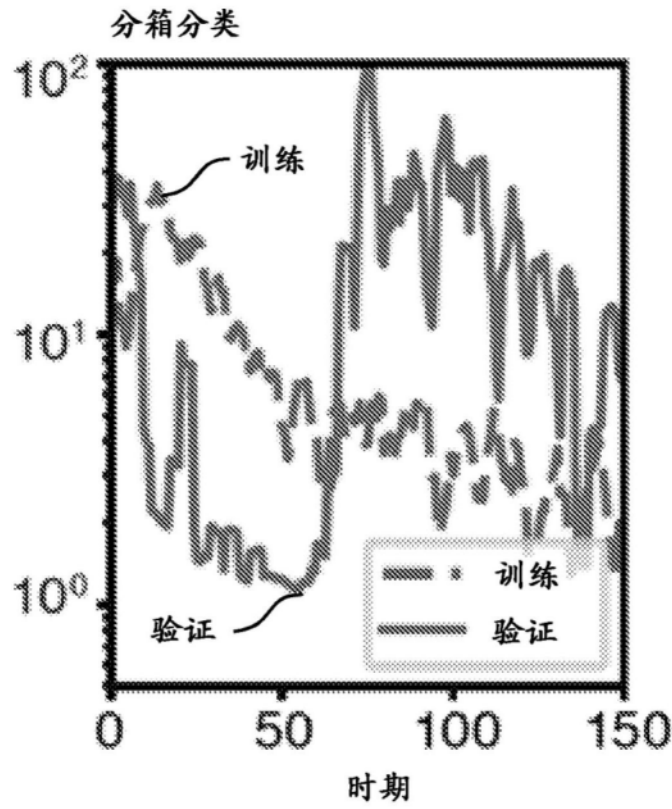


图13E

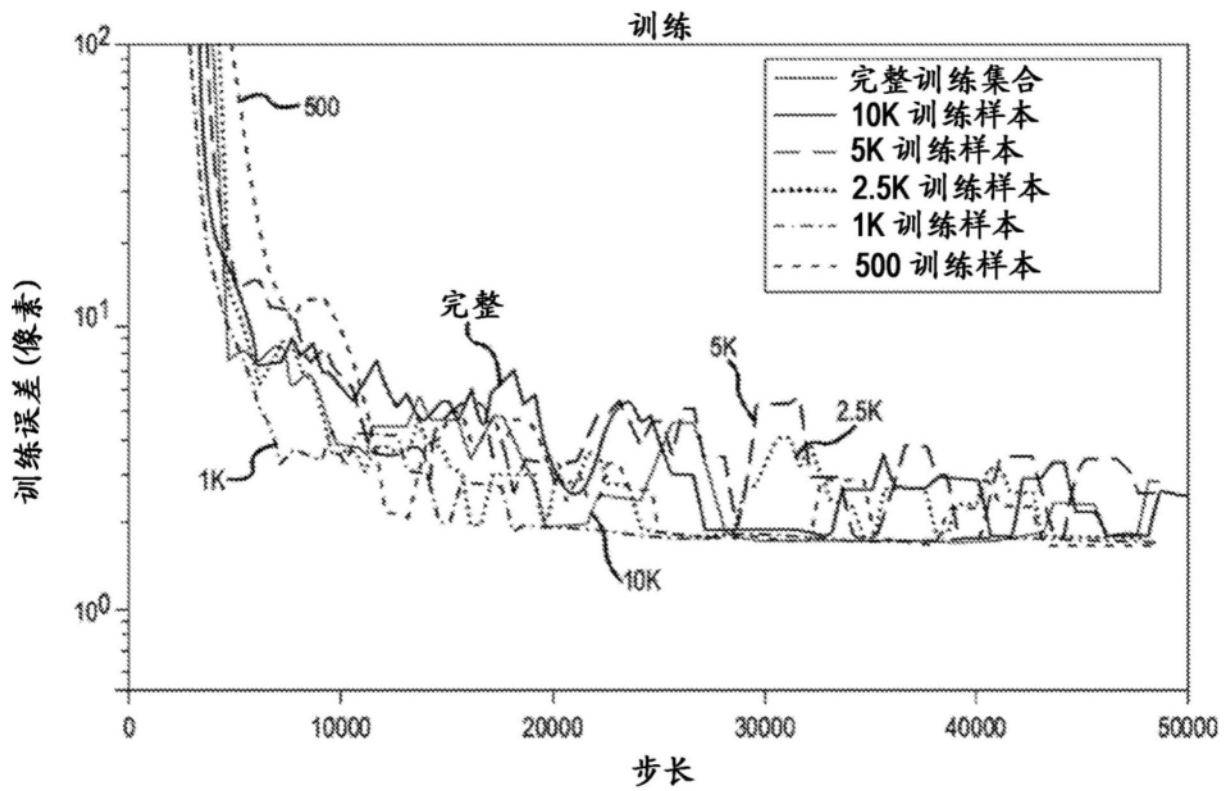


图14A

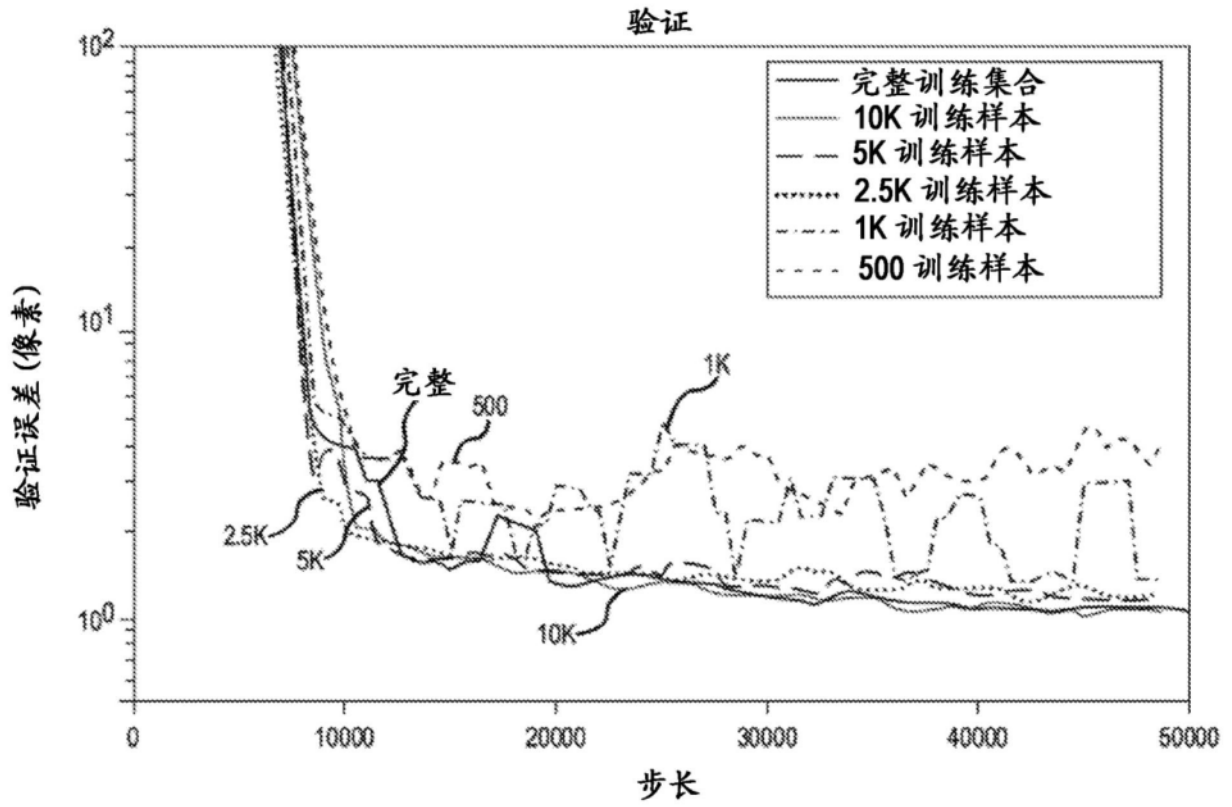


图14B

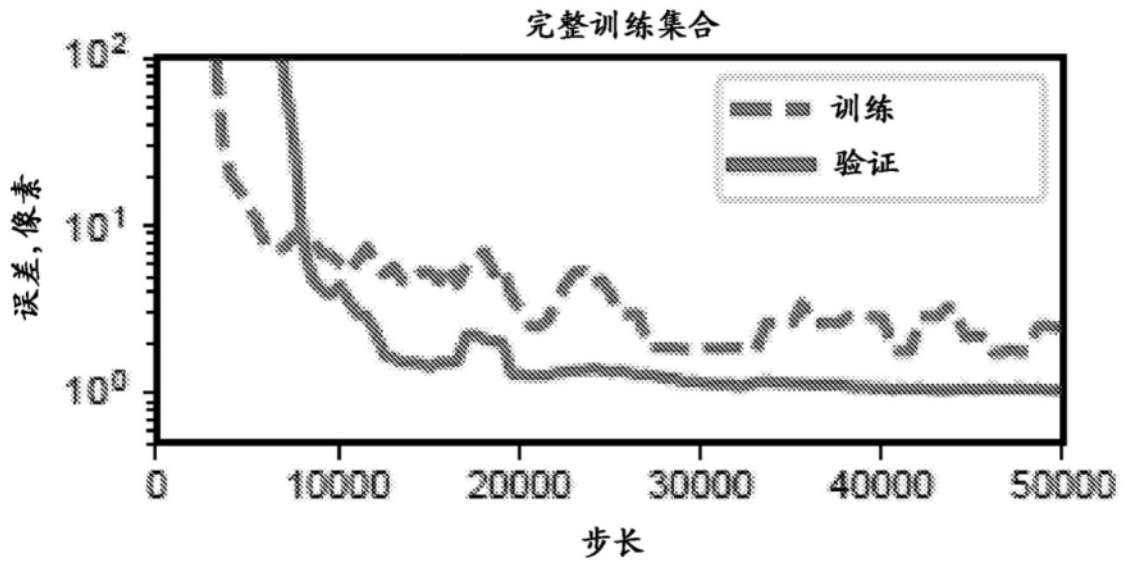


图14C

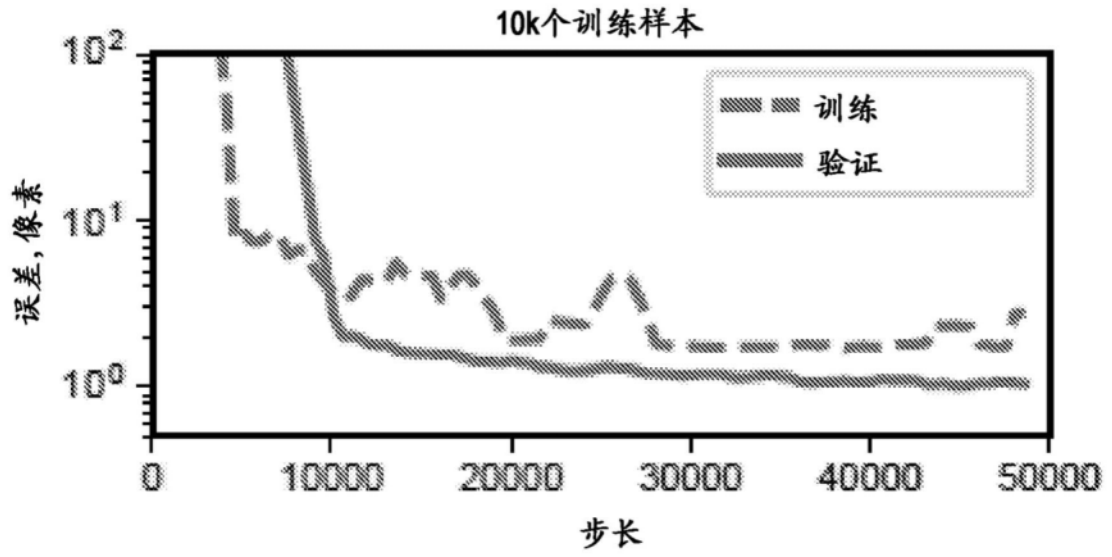


图14D

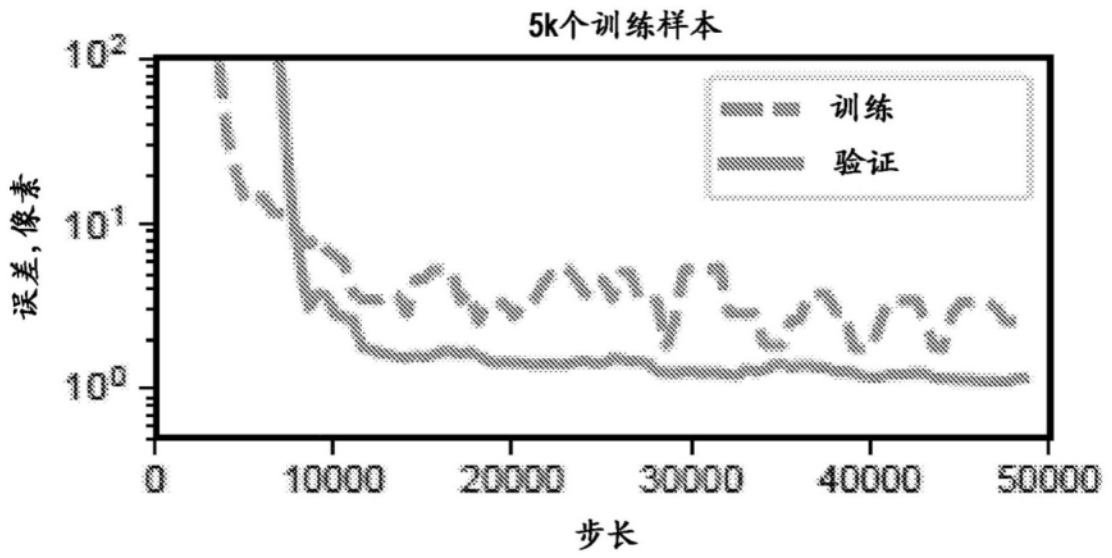


图14E

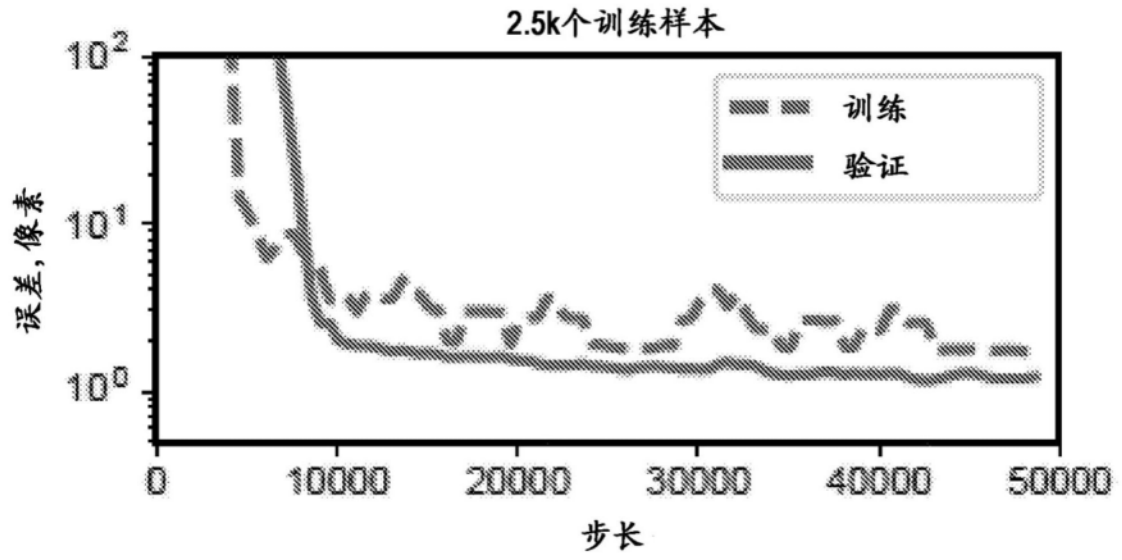


图14F

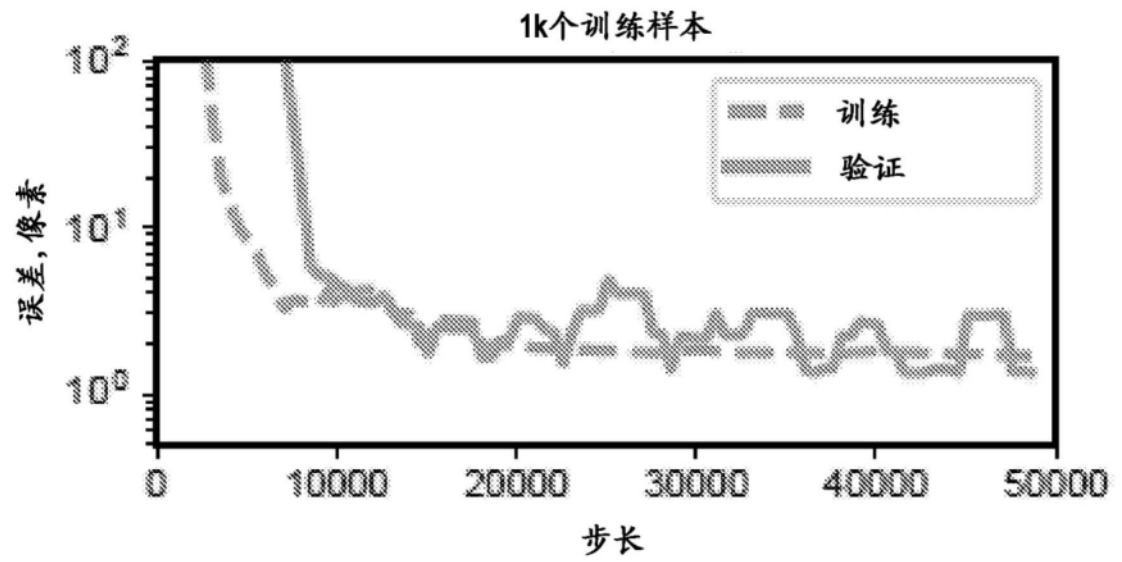


图14G

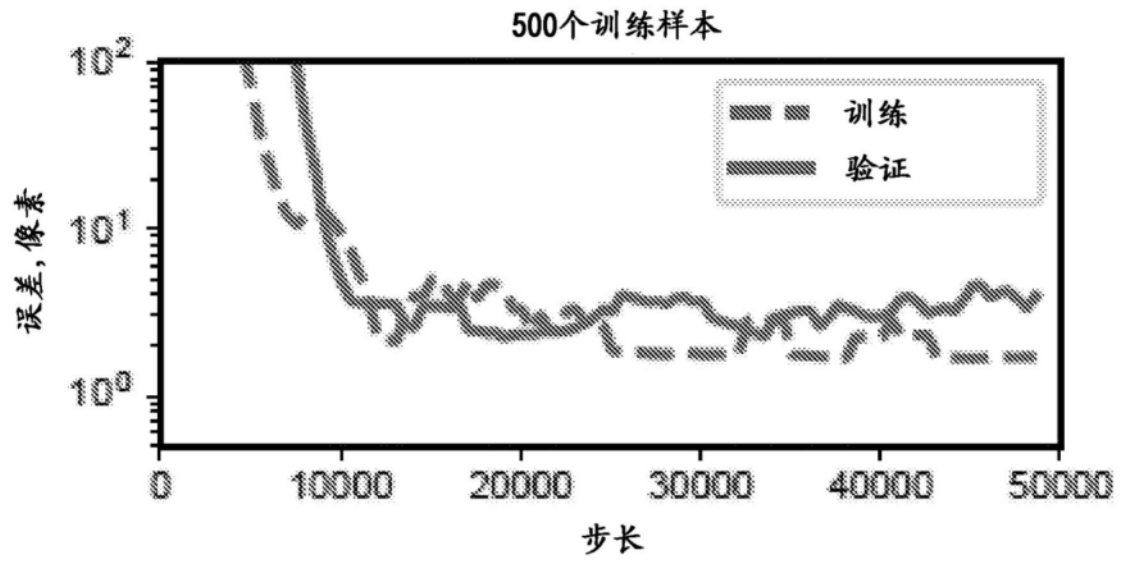


图14H

可见光

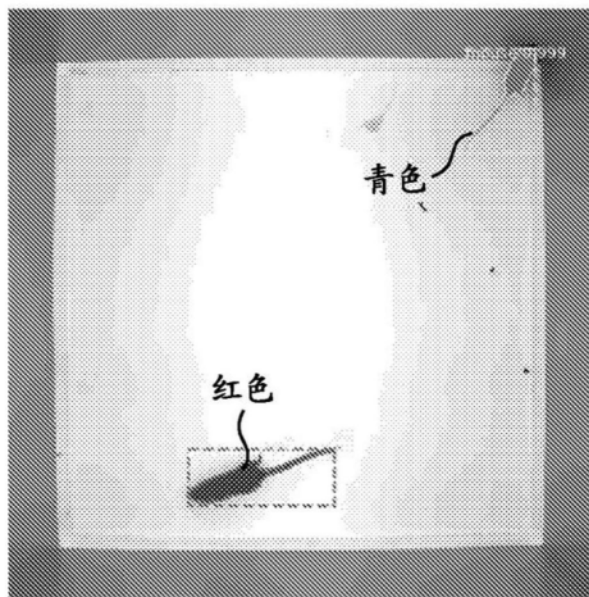


图15A

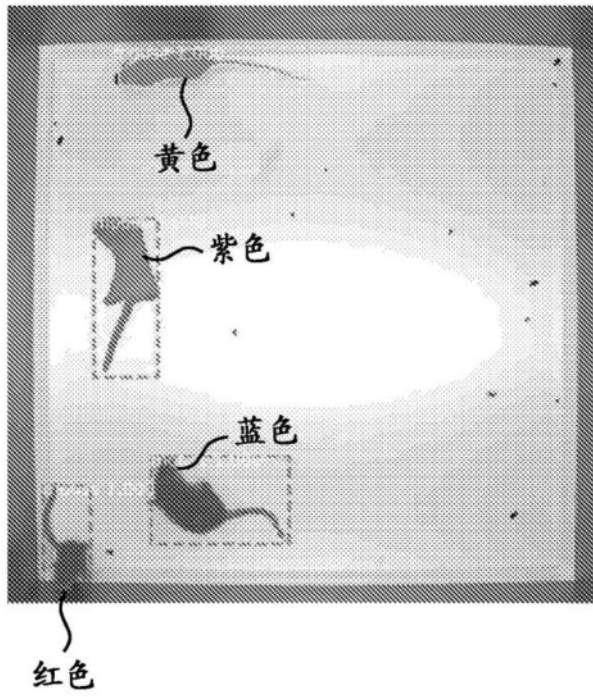


图15B

红外光

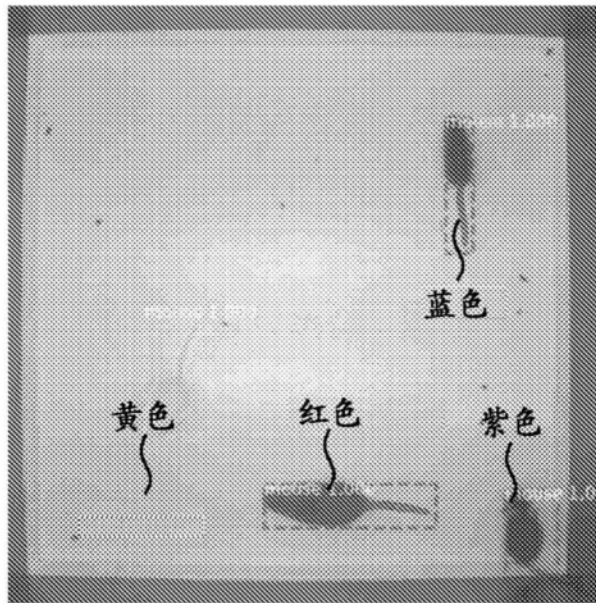


图15C

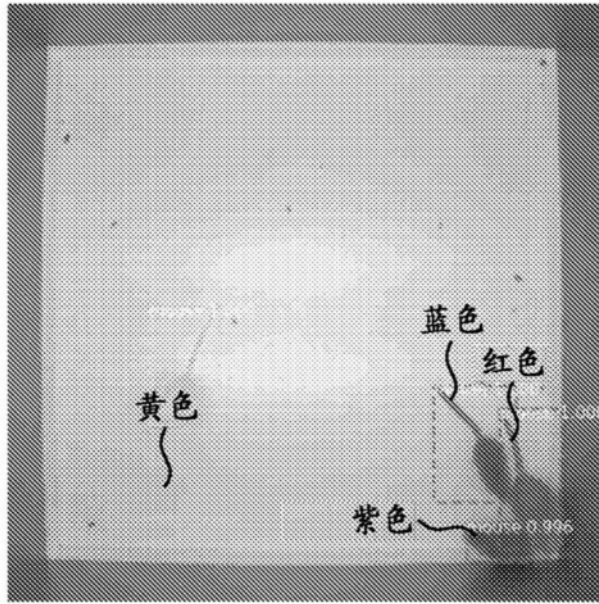


图15D

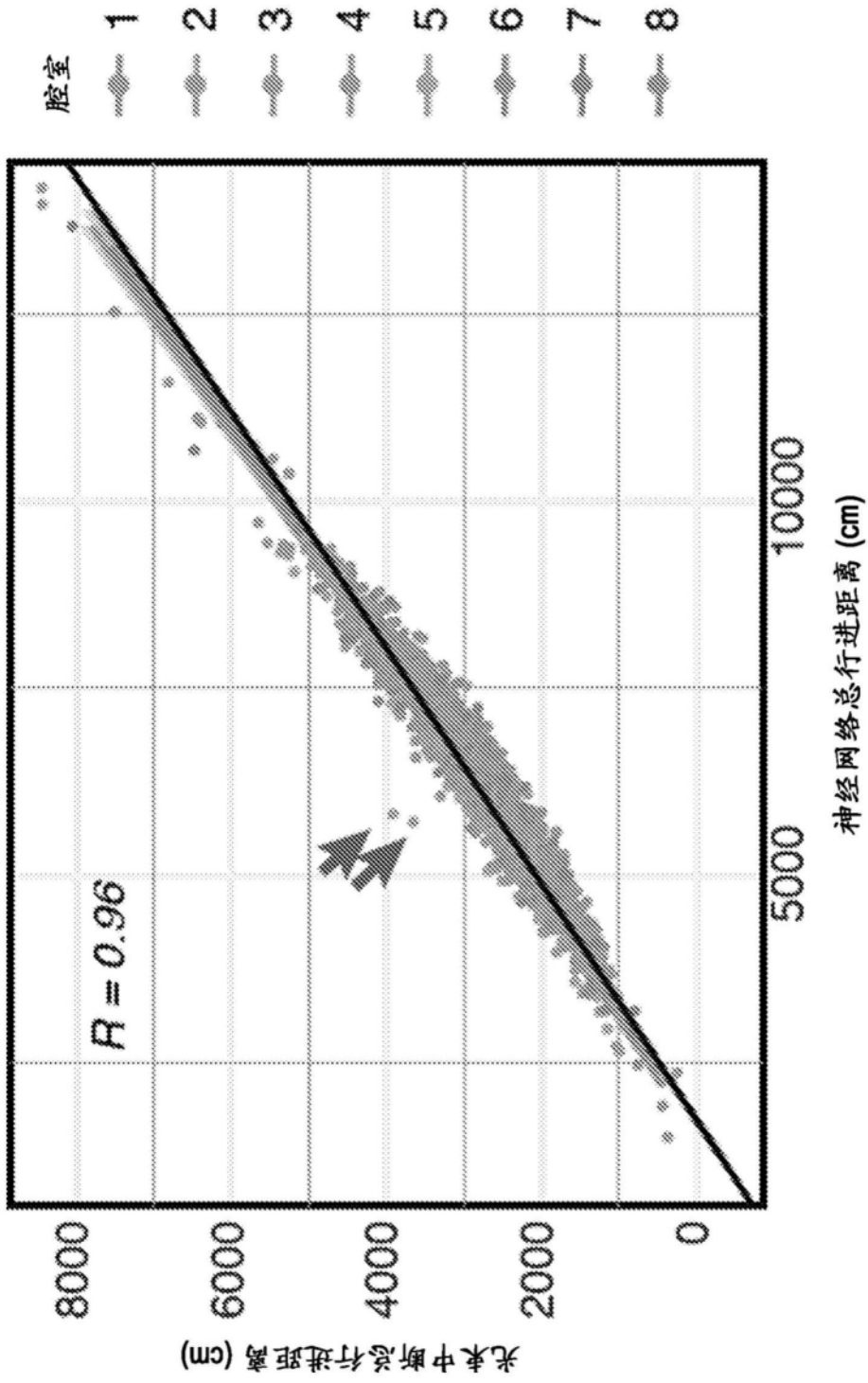


图16

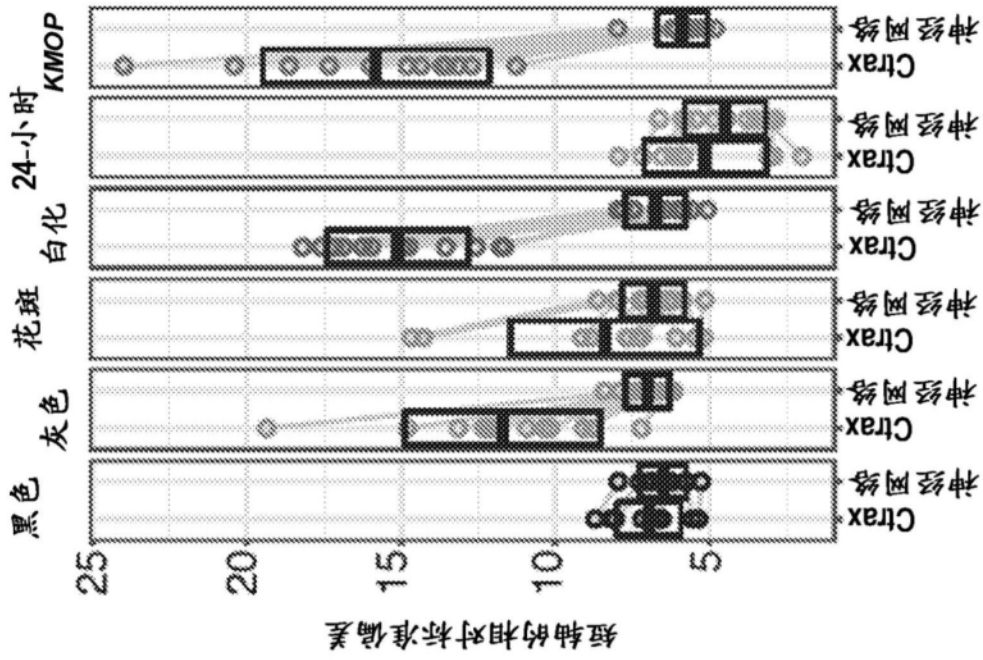


图17A

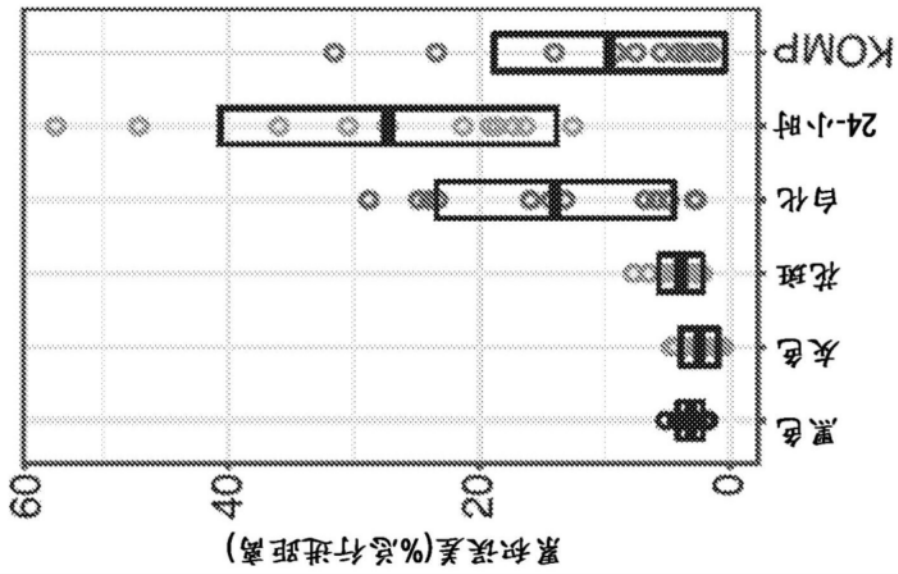


图17B

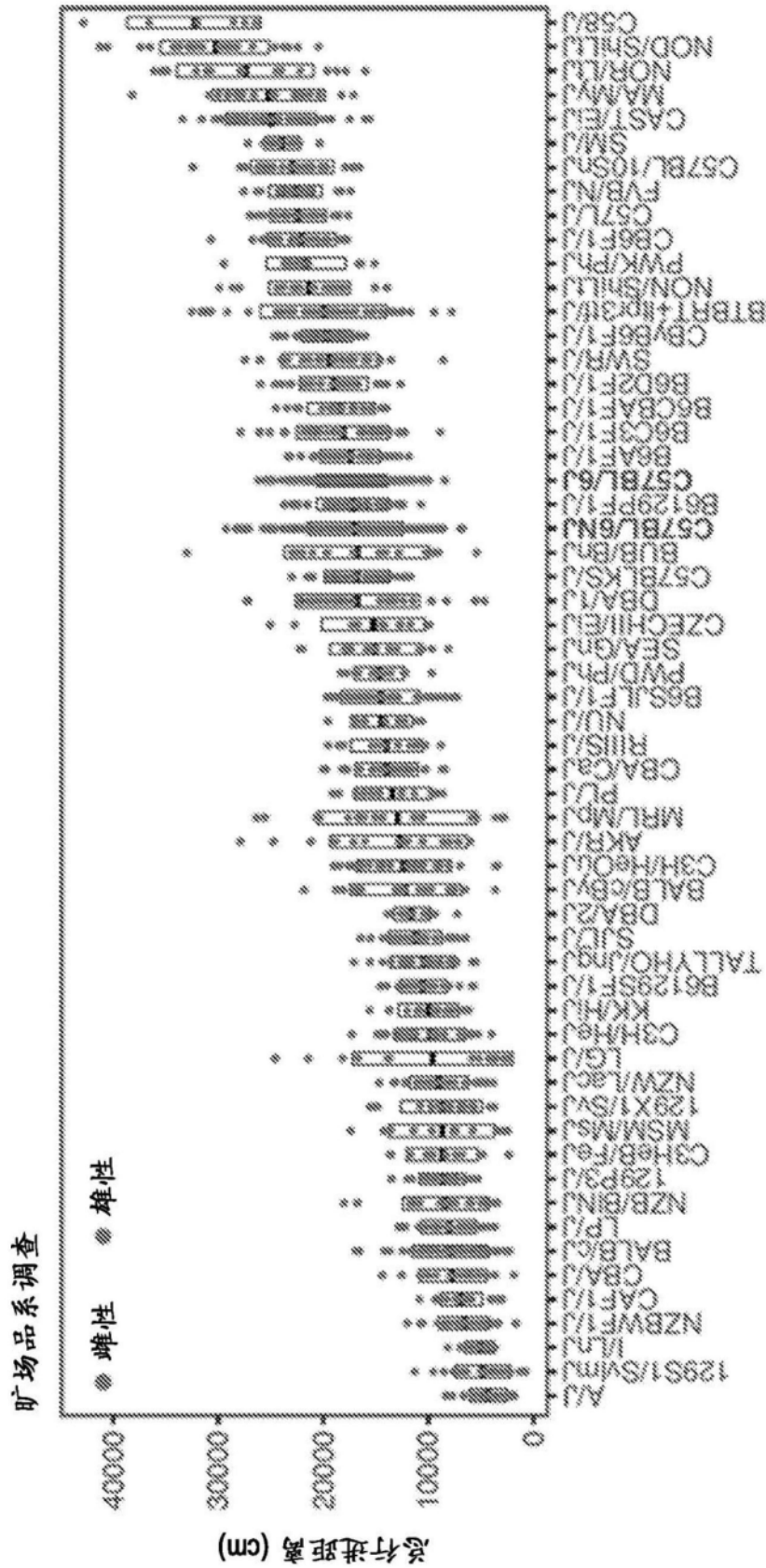


图18A

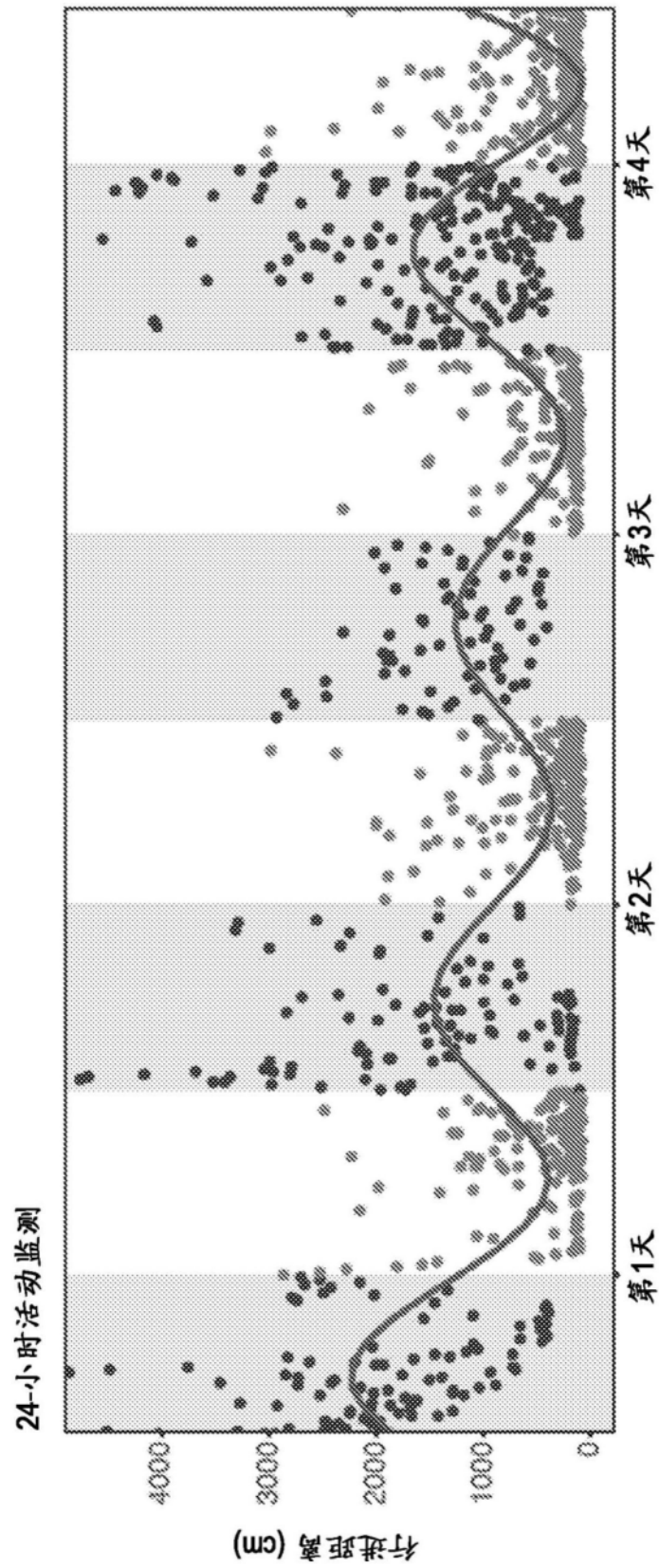


图18B