



(12)发明专利申请

(10)申请公布号 CN 108885682 A

(43)申请公布日 2018.11.23

(21)申请号 201680083751.1

(74)专利代理机构 北京市柳沈律师事务所

(22)申请日 2016.12.22

11105

(30)优先权数据

代理人 邵亚丽

15/055,446 2016.02.26 US

(51)Int.Cl.

(85)PCT国际申请进入国家阶段日

G06K 9/00(2006.01)

2018.09.18

G06N 3/02(2006.01)

(86)PCT国际申请的申请数据

PCT/US2016/068356 2016.12.22

(87)PCT国际申请的公布数据

W02017/146813 EN 2017.08.31

(71)申请人 谷歌有限责任公司

地址 美国加利福尼亚州

(72)发明人 P.C.尼尔森 E.M.克里斯琴森

M.伯恩德尔 M.弗伦金

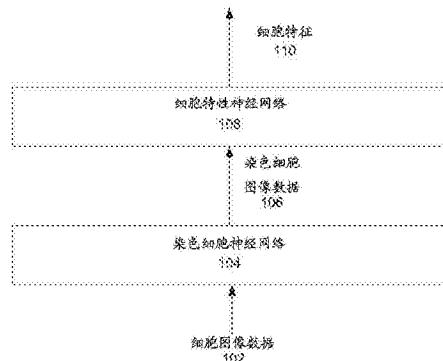
权利要求书1页 说明书6页 附图4页

(54)发明名称

使用神经网络处理细胞图像

(57)摘要

用于使用神经网络处理细胞图像的方法、系统和装置，包括编码在计算机存储介质上的计算机程序。一种方法包括获得包括用光学显微镜技术照射的一个或多个生物细胞的输入图像的数据；使用染色细胞神经网络处理数据；以及使用细胞特性神经网络处理一个或多个染色细胞图像，其中细胞特性神经网络通过训练已经被配置为接收一个或多个染色细胞图像并处理一个或多个染色细胞图像以生成表征在一个或多个染色细胞图像中染色的生物细胞的特征的细胞特性输出。



1. 一种方法,包括:

获得包括用光学显微镜技术照射的一个或多个生物细胞的输入图像的数据;

使用染色细胞神经网络处理数据,其中染色细胞神经网络通过训练已经被配置为接收包括用光学显微镜技术照射的所述一个或多个生物细胞的输入图像的数据,以及处理数据以生成一个或多个染色细胞图像,其中每个染色细胞图像描绘输入图像中的所述一个或多个生物细胞的相应染色版本;以及

使用细胞特性神经网络处理所述一个或多个染色细胞图像,其中细胞特性神经网络通过训练已经被配置为接收所述一个或多个染色细胞图像以及处理所述一个或多个染色细胞图像以生成表征在所述一个或多个染色细胞图像中染色的生物细胞的特征的细胞特性输出。

2. 根据权利要求1所述的方法,其中所述光学显微镜技术是以下中的一个:微分干涉对比显微镜、红光显微镜或紫外光显微镜。

3. 根据权利要求1或2所述的方法,其中所述染色细胞图像具有荧光通道。

4. 根据权利要求1至3中任一项所述的方法,其中,所述一个或多个染色细胞图像中的每一个中的每个像素对应于所述一个或多个生物细胞的输入图像中的像素。

5. 根据权利要求1至4中任一项所述的方法,其中细胞特性输出包括表征在所述一个或多个染色细胞图像中染色的所述一个或多个生物细胞的细胞类型、细胞健康、细胞群体、细胞表型或诱导多能干细胞协议中的细胞分化状态中的一个或多个的数据。

6. 根据前述权利要求中任一项所述的方法,其中染色细胞神经网络实施完全卷积的图像到图像的回归,并且包括所述一个或多个生物细胞的图像的数据是凹入的。

7. 根据前述权利要求中任一项所述的方法,其中细胞特性输出指定平均细胞健康或通过诱导多能干细胞协议正确分化的细胞百分比。

8. 根据前述权利要求中任一项所述的方法,其中细胞特性输出指定一个或多个生物细胞的每个生物细胞的细胞条件。

9. 根据前述权利要求中任一项所述的方法,其中包括所述一个或多个生物细胞的图像的数据还包括所述一个或多个生物细胞的附加图像,其中附加图像中的每一个具有用于所述一个或多个生物细胞的图像的不同的焦点。

10. 一种系统,包括一个或多个计算机和存储指令的一个或多个存储设备,所述指令在由所述一个或多个计算机运行时使所述一个或多个计算机执行根据前述权利要求中任一项所述的方法。

11. 存储指令的一个或多个非暂时性存储介质,当所述指令由一个或多个计算机运行时,使得所述一个或多个计算机执行根据权利要求1至9中任一项所述的方法。

## 使用神经网络处理细胞图像

### 背景技术

- [0001] 本说明书涉及用于处理生物细胞图像的神经网络。
- [0002] 神经网络是机器学习模型,其采用一个或多个模型层来为接收的输入生成输出,例如,一个或多个分类。除了输出层之外,一些神经网络还包括一个或多个隐藏层。每个隐藏层的输出用作网络中下一层的输入,即网络的下一个隐藏层或输出层。网络的每个层根据相应参数集的当前值从接收的输入生成输出。
- [0003] 处理细胞的现有方法包括用荧光标记物手动洗涤细胞以生产染色细胞。从染色细胞中,科学家们可以手动识别细胞的特性。

### 发明内容

- [0004] 一般地,本说明书描述了一种用于处理细胞图像的神经网络。
- [0005] 具体地,系统获得包括用光学显微镜技术(optical microscopy technique)照射的一个或多个生物细胞的输入图像的数据。
- [0006] 系统使用染色细胞神经网络处理数据。染色细胞神经网络是深度神经网络,其通过训练已经被配置为接收数据并处理数据以生成一个或多个染色细胞图像。每个染色细胞图像描绘输入图像中的一个或多个生物细胞的相应染色版本。
- [0007] 在本说明书中描述的主题的特定实施例可以被实施,以实现以下优点中的一个或多个。
- [0008] 染色细胞神经网络可以从细胞图像预测染色细胞图像,细胞特性神经网络可以从染色细胞图像预测细胞特性。因此,可以使用自动化过程从细胞图像预测细胞的特性,这可以降低成本和测试时间。也就是说,可以预测细胞的特性而不必对活细胞进行物理染色,这减少了需要修改以确定细胞特性的活细胞的数量。具体地,可以在没有物理染色并且可能杀死细胞的情况下生成活细胞的染色图像。然后可以使用这些染色图像来确定细胞的特性。另外,使用神经网络生成染色图像比物理染色细胞以生成染色图像更有效,并且需要更少的资源和时间。此外,神经网络可以被配置为同时对任意数量的通道进行染色,而物理协议只能同时染色少量通道。
- [0009] 在附图和以下描述中阐述了本说明书的主题的一个或多个实施例的细节。根据说明书、附图和权利要求,本主题的其他特征、方面和优点将变得显而易见。

### 附图说明

- [0010] 图1示出了用于处理细胞图像的示例系统。
- [0011] 图2是用于处理一个或多个细胞的图像以识别一个或多个细胞的特征的示例过程的流程图。
- [0012] 图3是用于训练染色细胞神经网络的示例性过程的流程图。
- [0013] 图4是用于训练细胞特性神经网络的示例过程的流程图。
- [0014] 在不同的附图中相同的附图标记和名称表示相同的元素。

## 具体实施方式

[0015] 图1示出了用于处理细胞图像的示例系统100。系统100是在一个或多个位置中的一个或多个计算机上实施为计算机程序的系统的示例，其中可以实施下面描述的系统、组件和技术。

[0016] 系统100包括染色细胞神经网络104。染色细胞神经网络104是被配置为处理细胞图像数据102以生成染色细胞图像数据106的神经网络。细胞图像数据102可以包括由显微相机拍摄的一个或多个生物细胞的一个或多个图像。染色细胞图像数据106包括预测细胞图像数据102中的细胞如果被染色会是什么样子的图像。

[0017] 细胞染色是一种用于，例如，使用显微镜，通过向细胞应用染色来使生物细胞可视化的技术。在应用染色之后，不同的染色剂可以使细胞的各种组分显示视觉标记，例如，将颜色从其原始颜色改变为不同的颜色。由于细胞被特定染色剂染色，因此可以显示视觉标记的细胞的示例组分包括细胞的细胞核或细胞膜。一些细胞染色技术在活细胞上进行，而其他染色技术可在非活细胞上进行。例如，一种细胞染色技术是将细胞样品浸入染料溶液中，然后在显微镜下漂洗(rinse)并捕获样本的图像。在对细胞进行染色后，科学家们可以直观地分析细胞或细胞的染色图像，以确定细胞的特性。

[0018] 在一些实施方式中，染色细胞神经网络104被训练以根据给定输入细胞图像生成多种类型的染色细胞图像。每个染色细胞图像对应于不同的类型的染色。例如，预测的染色图像可以是使用体内(*in vivo*)染色或体外(*in vitro*)染色生成的染色。作为说明，一个染色细胞图像可以是利用结晶紫染色剂染色的一个或多个细胞的预测图像。其他染色细胞图像可以是细胞图像的各种荧光通道。

[0019] 具体地，染色细胞神经网络104接收包括细胞图像的像素的数据作为输入。当生成染色图像时，染色细胞神经网络104针对细胞图像的每个输入像素生成相应的输出像素。每个细胞图像的输出像素形成相应的染色细胞图像。因此，染色细胞神经网络104可以实施完全卷积的图像到图像的回归。可以用于染色细胞神经网络104的示例神经网络体系结构在Hyeonwoo Noh、Seunghoon Hong和Bohyung Han发表的“Learning Deconvolution Network for Semantic Segmentation”中描述，可以访问<http://arxiv.org/abs/1505.04366>。

[0020] 系统100还包括细胞特性神经网络108。细胞特性神经网络108被配置为处理由染色细胞神经网络104生成的染色细胞图像数据106，以生成细胞图像数据102中的一个或多个细胞的细胞特征110。细胞特征110表征细胞图像数据102中的一个或多个细胞的特征。例如，细胞特征110可以指示以下中的一种或多种：细胞类型、健康、起源或状况、诱导多能干细胞方案中的分化状态、或细胞计数。例如，对于每个输入图像，细胞特性神经网络108可以生成输出的特征向量，其中每个特征对应于特征向量中输出的相应部分。

[0021] 为了配置神经网络104和108以生成染色细胞图像数据106和细胞特征110，系统100可以在训练数据上训练神经网络104和108，这将在下面参考图3和4进一步描述。

[0022] 图2是用于处理一个或多个细胞的图像以识别一个或多个细胞的特征的示例过程200的流程图。为了方便起见，将针对具有一个或多个运行软件以实施过程200的计算设备的系统(例如，图1的系统100)来描述过程200。

[0023] 系统获得包括利用光学显微镜技术照射的一个或多个细胞的一个或多个图像的

数据(步骤202)。

[0024] 例如,在一个或多个图像中,可以通过差分干涉对比显微镜、红光显微镜或紫外光显微镜照射细胞。在一些实施方式中,一个或多个细胞的每个图像都是凹入的(foveated)。在一些实施方式中,数据包括多个图像,每个图像由不同的显微镜技术照射。

[0025] 在一些实施方式中,数据包括一组焦点变动(varying focus)的图像。也就是说,每个图像聚焦在一个或多个细胞的不同的深度点处,并且该组图像中的不同的深度点跨越(span)一个或多个细胞的整个结构。图像可以由高分辨率相机拍摄。

[0026] 系统使用染色细胞神经网络(例如,图1的染色细胞神经网络104)处理数据,以生成一个或多个染色细胞图像(步骤204)。

[0027] 染色细胞神经网络是通过训练已经被配置为接收包括一个或多个细胞的一个或多个图像的数据并处理一个或多个图像以生成一个或多个染色细胞图像的神经网络。下面将参考图3进一步描述训练染色细胞神经网络。

[0028] 在数据包括一组焦点变动的图像的实施方式中,神经网络包括多个独立的神经网络层塔(tower)、组合每个塔的输出以生成组合输出的中间组合层、以及处理组合输出以生成染色细胞图像的一个或多个附加神经网络层。在这些实施方式中,系统将该组图像分组为多个图像组,其中每个组对应于相应的塔中的一个。然后,系统使用相应的塔处理每个组。在一些情况下,每组包括来自输入该组图像的单个图像。在一些其他情况下,一个或多个组可以包括多个图像,其中对应塔的输入是多维张量,例如,组中的图像的堆栈。

[0029] 系统使用细胞特性神经网络(例如,图1的细胞特性神经网络108)来处理一个或多个染色细胞图像,以产生表征一个或多个细胞的特征的输出(步骤206)。

[0030] 细胞特性神经网络是通过训练已经被配置为接收一个或多个染色细胞图像的神经网络,染色细胞图像描绘一个或多个细胞如果被染色会是什么样子,并且生成表征染色细胞图像中描述的一个或多个细胞的特征的输出。

[0031] 例如,细胞特性神经网络可以被配置为输出包括多个输出值的特征向量。神经网络被配置为表征的每个特征可以对应于特征向量中的输出值的相应部分。例如,特征可以指示以下中的一个或多个:细胞类型、健康、起源或状况、诱导多能干细胞方案中的分化状态,或细胞计数。作为一个示例,特征向量中的输出值中的一个预定输出值可以表示输入图像中的细胞数量的预测计数,而对应于细胞类型特征的部分可以包括多个值,每个值对应于相应类型的细胞。对于给定细胞类型的值可以表示图像中的细胞是该类型的可能性。

[0032] 下面将参照图4进一步描述细胞特性神经网络的训练。

[0033] 图3是用于训练染色细胞神经网络的示例过程300的流程图。为了方便起见,将针对具有一个或多个运行软件以实施过程300的计算设备的系统(例如,图1的系统100)来描述过程300。

[0034] 系统获得用于染色细胞神经网络的训练数据(步骤302)。训练数据包括多个训练示例,其中每个训练示例包括一个或多个细胞的一个或多个图像,并且对于每个训练示例,包括一个或多个相应的染色图像。在一些情况下,对于每个训练实例,相应的染色图像可以描绘利用各种染色剂染色的细胞。

[0035] 系统在训练数据上训练染色细胞神经网络(步骤304)。为了训练染色细胞神经网络,系统在每个训练示例中使用染色细胞神经网络处理图像以生成一个或多个染色细胞图

像。通常，每个染色细胞图像对应于不同的染色。然后，对于每个染色细胞图像的每个像素，系统确定由染色细胞神经网络生成的像素值和训练数据中相应的染色细胞图像中的像素值之间的误差。

[0036] 在一些实施方式中，染色图像可以包括利用靶向特定细胞隔室 (cellular compartment) 或结构的物质染色的细胞图像。例如，一些染色剂可以是特定地染色细胞核的染色剂，而其他染色剂可以结合细胞骨架的组分或一个或多个细胞膜，包括外细胞膜 (outer cell membrane) 和/或细胞内膜 (intracellular membrane) 的其他细胞膜。通过在这些图像上训练染色细胞神经网络，染色细胞神经网络可以来识别可能被这些不同的物质染色的特定细胞的部分。染色细胞神经网络可以反过来学习生成反映在将染色剂被应用于细胞的情况下这些细胞看起来是什么样子的图像。

[0037] 系统使用常规的神经网络的训练技术(例如，具有反向传播的随机梯度下降)，使用误差来更新染色细胞神经网络的每一层中使用的参数值。

[0038] 图4是用于训练细胞特性神经网络的示例过程400的流程图。为了方便起见，将针对具有一个或多个运行软件以实施过程400的计算设备的系统(例如，图1的系统100)来描述过程400。

[0039] 系统获得用于细胞特性神经网络的训练数据(步骤402)。训练数据包括多个训练图像，其中每个训练示例是一个或多个染色细胞的图像，并且对于每个训练图像是相应的特征向量。具体地，相应的特征向量包括与细胞特性神经网络被配置为预测的每个特征相对应的输出。

[0040] 在一些实施方式中，系统接收包括未染色细胞的图像的训练数据，然后使用训练的染色细胞神经网络处理图像，以生成将用作细胞特性神经网络的训练数据的训练图像。

[0041] 系统在训练数据上训练细胞特性神经网络(步骤404)。也就是说，系统使用细胞特性神经网络处理每个训练图像，以确定训练图像的预测特征向量。然后，系统确定训练数据中训练图像的预测特征向量和对应于训练图像的特征向量之间的误差，并使用常规神经网络训练技术使用该误差来更新细胞特性神经网络的每一层中使用的参数值。

[0042] 对于被配置为执行特定操作或动作的一个或多个计算机的系统，意味该系统已经在其上已经安装了在操作中使得系统执行操作或动作的软件、固件、硬件或它们的组合。对于要被配置为执行特定操作或动作的一个或多个计算机程序，意味着一个或多个程序包括当由数据处理装置运行时使装置执行操作或动作的指令。

[0043] 本说明书中描述的主题和功能操作的实施例可以在数字电子电路、有形实施的计算机软件或固件、计算机硬件中实施，包括本说明书中公开的结构及其结构等价物、或者它们中的一个或多个的组合。本说明书中描述的主题的实施例可以被实施为一个或多个计算机程序，即，在有形非暂时性程序载体上编码的一个或多个计算机程序指令模块，用于由数据处理装置运行或控制数据处理装置的操作。可替换地或附加地，程序指令可以被编码在人工生成的传播信号上，例如机器生成的电信号、光信号或电磁信号，该信号被生成以编码信息，用于传输到合适的接收器装置以由数据处理装置运行。计算机存储介质可以是机器可读存储设备、机器可读存储基板、随机或串行存取存储器设备、或者它们中的一个或多个的组合。然而，计算机存储介质不是传播信号。

[0044] 术语“数据处理装置”包括用于处理数据的所有类型的装置、设备和机器，包括例

如可编程处理器、计算机或多个处理器或计算机。该装置可以包括专用逻辑电路，例如FPGA (field programmable gate array, 现场可编程门阵列) 或ASIC(application-specific integrated circuit, 专用集成电路)。除了硬件之外，该装置还可以包括为所讨论的创建计算机程序运行环境的代码，例如，构成处理器固件、协议栈、数据库管理系统、操作系统、或者它们中的一个或多个的组合的代码。

[0045] 计算机程序(也可被称为或被描述为程序、软件、软件应用程序、模块、软件模块、脚本或代码)可以以任何形式的编程语言编写，包括编译或解释语言或者声明性或过程性语言，并且它可以以任何形式部署，包括作为独立程序或作为模块、组件、子程序或适合在计算环境中使用的其他单元。计算机程序可以但不必对应于文件系统中的文件。程序可以被存储在保存其他程序或数据的文件的一部分中，例如，存储在标记语言文档中的一个或多个脚本、专用于所讨论的程序的单个文件中或者存储在多个协调文件中，例如，存储一个或多个模块、子程序或代码的部分的文件中。计算机程序可以被部署成在一台计算机上或者在位于一个站点或者分布跨越多个站点并且通过通信网络互连的多台计算机上运行。

[0046] 如本说明书中所使用的，“引擎”或“软件引擎”是指提供不同于输入的输出的软件实施的输入/输出系统。引擎可以是编码的功能块，诸如库、平台、软件开发工具包 (software development kit, “SDK”) 或对象。每个引擎可以在任何适当类型的计算设备上实施，例如，服务器、移动电话、平板计算机、笔记本计算机、音乐播放器、电子书阅读器、膝上型或台式计算机、PDA、智能电话或包括一个或多个处理器和计算机可读介质的其他固定或便携式设备。附加地，两个或多个引擎可以在相同计算设备上或在不同的计算设备上实施。

[0047] 本说明书中描述的过程和逻辑流程可以由运行一个或多个计算机程序以通过对输入数据进行操作并生成输出来执行功能的一个或多个可编程计算机执行。过程和逻辑流程也可以由专用逻辑电路执行，并且装置也可以被实施为专用逻辑电路，例如FPGA (现场可编程门阵列) 或ASIC(专用集成电路)。

[0048] 适合于运行计算机程序的计算机包括，例如，可以基于通用或专用微处理器或者两者，或者任何其他种类的中央处理单元。通常，中央处理单元将从只读存储器或随机存取存储器或者两者接收指令和数据。计算机的基本元件是用于执行或运行指令的中央处理单元和用于存储指令和数据的一个或多个存储器设备。通常，计算机还将包括或可操作地耦合以从用于存储数据的一个或多个大容量存储设备(例如磁盘、磁光盘或光盘)接收数据或者将数据传输到一个或多个大容量存储设备，或者进行两者。然而，计算机不需要具有这样的设备。此外，计算机可以嵌入在另一个设备中，例如移动电话、个人数字助理 (personal digital assistant, PDA)、移动音频或视频播放器、游戏控制台、全球定位系统 (Global Positioning System, GPS) 接收器或便携式存储设备(例如，通用串行总线 (universal serial bus, USB) 闪存驱动器)，仅试举几例。

[0049] 适于存储计算机程序指令和数据的计算机可读介质包括所有形式的非易失性存储器、介质和存储设备，例如包括半导体存储设备，例如EPROM、EEPROM和闪存设备；磁盘，例如内部硬盘或可移动磁盘；磁光盘；以及光盘和DVD-ROM盘。处理器和存储器可以由专用逻辑电路补充或并入专用逻辑电路中。

[0050] 为了提供与用户的交互，本说明书中描述的主题的实施例可以在具有用于向用户

显示信息的显示设备(例如CRT(cathode ray tube,阴极射线管)监视器、LCD(liquid crystal display,液晶显示器)监视器或OLED显示器)以及用于向计算机提供输入的输入设备(例如键盘、鼠标或存在敏感显示器或其他表面)的计算机上实施。也可以使用其他类型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;并且可以以任何形式接收来自用户的输入,包括声音、语音或触觉输入。此外,计算机可以通过向用户使用的设备发送资源和从用户使用的设备接收资源来与用户交互;例如,通过响应于从web浏览器接收的请求,将网页发送到用户客户端设备上的web浏览器。

[0051] 本说明书中描述的主题的实施例可以在计算系统中实施,该计算系统包括后端组件,例如数据服务器,或者包括中间件组件,例如应用服务器,或者包括前端组件,例如具有图形用户界面或Web浏览器的客户端计算机,用户可以通过该界面或Web浏览器与本说明书中描述的主题的实施交互,或者一个或多个这样的后端、中间件或前端组件的任何组合。系统的组件可以由任何形式或介质的数字数据通信(例如通信网络)互连。通信网络的示例包括局域网(local area network,“LAN”)和广域网(wide area network,“WAN”),例如因特网。

[0052] 计算系统可以包括客户端和服务器。客户端和服务器通常彼此远离,并且典型地通过通信网络进行交互。客户端和服务器的关系由于在各自的计算机上运行并且彼此具有客户端-服务器关系的计算机程序而产生。

[0053] 虽然本说明书包含许多具体的实施细节,但是这些不应该被解释为对任何发明或可能要求保护的范围的限制,而是作为特定于特定发明的特定实施例的特征的描述。本说明书中在单独实施例的上下文中描述的某些特征也可以在单个实施例中组合实施。反过来,在单个实施例的上下文中描述的各种特征也可以在多个实施例中单独地或以任何合适的子组合来实施。此外,尽管特征可以在上文中被描述为以某些组合起作用,甚至最初这样要求保护,但是在某些情况下,来自要求保护的组合的一个或多个特征可以从该组合中删除,并且要求保护的组合可以指向子组合或子组合的变体。

[0054] 类似地,虽然在附图中以特定顺序描绘了操作,但是这不应该被理解为要求以所示的特定顺序或按顺序执行这些操作,或者执行所有示出的操作,以实现期望的结果。在某些情况下,多任务处理和并行处理可能是有利的。此外,上述实施例中的各种系统模块和组件的分离不应被理解为在所有实施例中都需要这种分离,并且应当理解,所描述的程序组件和系统通常可以集成在单个软件产品中或被封装到多个软件产品中。

[0055] 已经描述了主题的特定实施例。其他实施例在所附权利要求的范围内。例如,权利要求中记载的动作可以以不同的顺序执行并且仍然实现期望的结果。作为一个示例,附图中描绘的过程不一定需要所示的特定顺序或序列顺序来实现期望的结果。在某些实施中,多任务和并行处理可能是有利的。

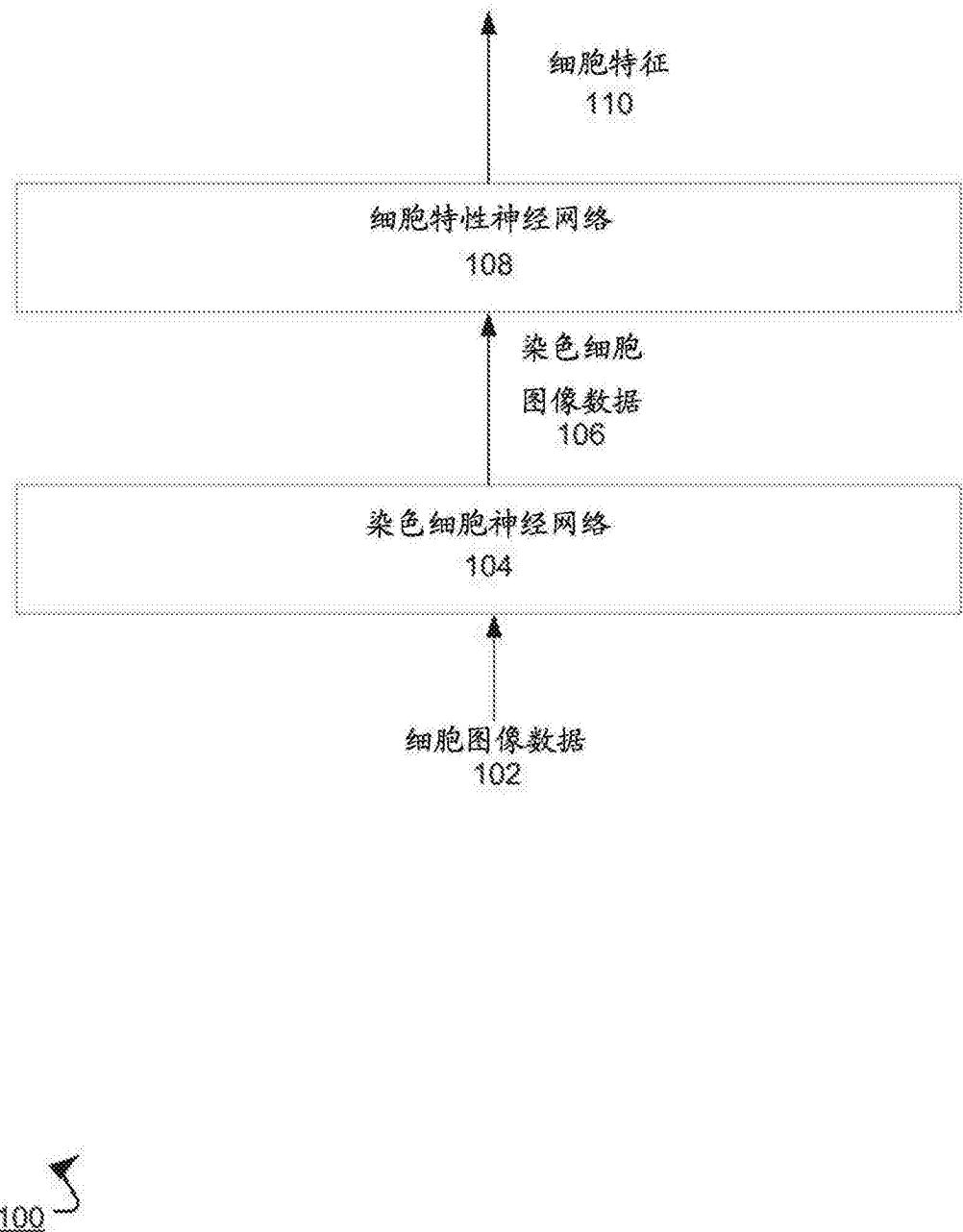


图1



图2

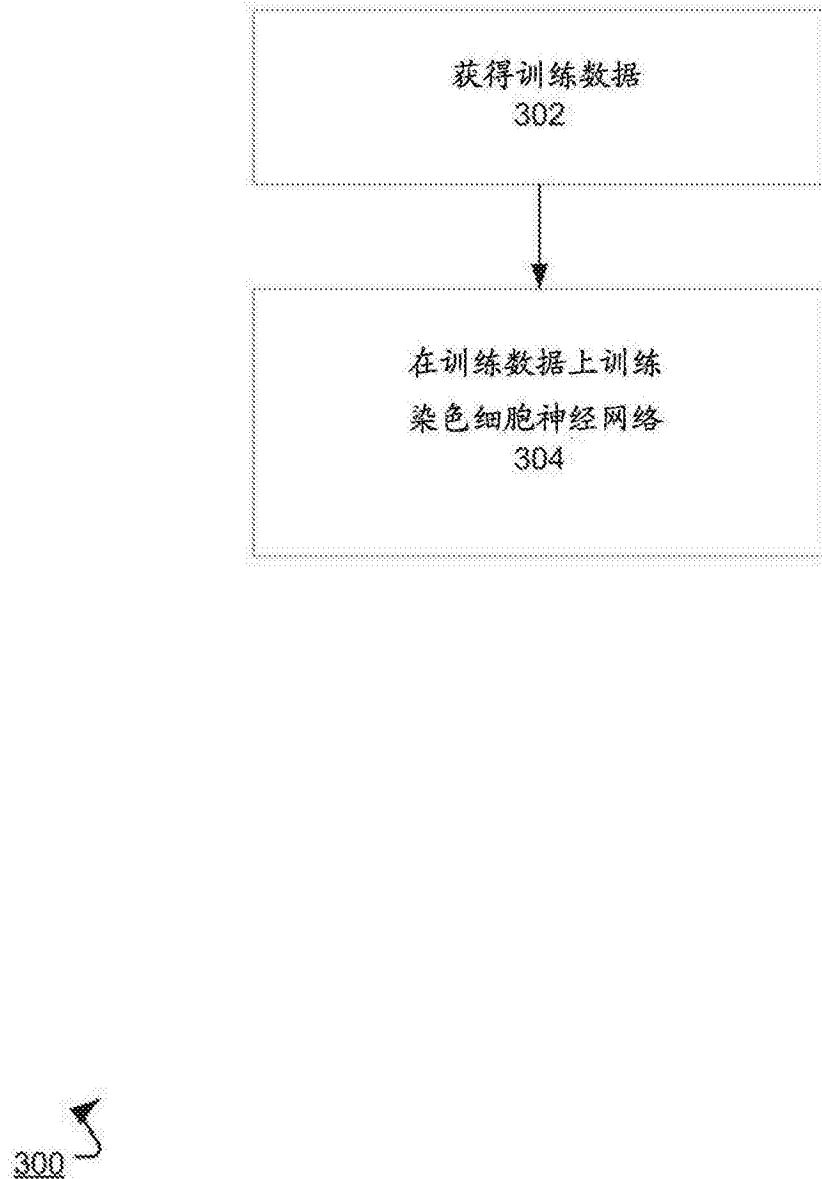
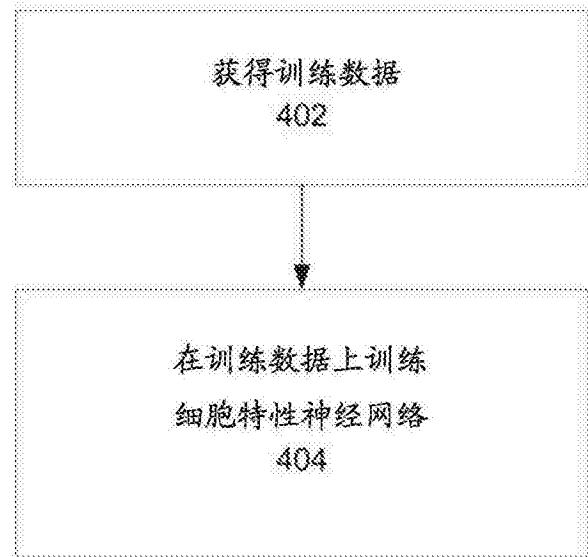


图3



400 ↗

图4