



(12) 发明专利申请

(10) 申请公布号 CN 113573633 A

(43) 申请公布日 2021. 10. 29

(21) 申请号 202080020504.3

(22) 申请日 2020.01.08

(30) 优先权数据

62/790,747 2019.01.10 US

16/736,467 2020.01.07 US

(85) PCT国际申请进入国家阶段日
2021.09.10

(86) PCT国际申请的申请数据
PCT/US2020/012740 2020.01.08

(87) PCT国际申请的公布数据
W02020/146501 EN 2020.07.16

(71) 申请人 威里利生命科学有限责任公司
地址 美国加利福尼亚州

(72) 发明人 X.金 J.巴拉尔

(74) 专利代理机构 北京市柳沈律师事务所
11105

代理人 金玉洁

(51) Int.Cl.
A61B 5/06 (2006.01)
G06K 9/00 (2006.01)
G06Q 50/10 (2006.01)

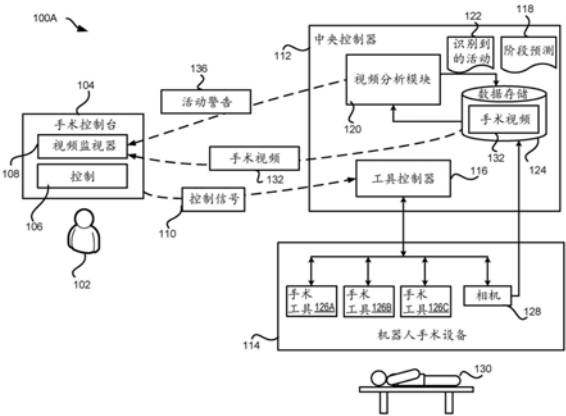
权利要求书3页 说明书17页 附图12页

(54) 发明名称

基于手术视频的手术工作流和活动检测

(57) 摘要

用于经过视频处理检测手术过程的阶段的一个示例方法包括：访问手术过程的视频并将视频划分成一个或多个块，每个块包含一个或多个视频帧。对于每个块，本方法包括对各个块的视频帧应用预测模型，以对每个视频帧获得阶段预测。预测模型被配置为对输入视频帧预测手术过程的多个阶段之一。本方法进一步包括通过聚合视频帧的阶段预测来对各个块生成聚合阶段预测，以及基于聚合阶段预测，修改手术过程的视频，以包括对各个块的预测阶段的指示。



1. 一种方法,包括:
访问手术过程的视频,手术过程包括多个阶段;
将视频划分成一个或多个块,一个或多个块中的每一个包括一个或多个视频帧;
对于每个块:
对各个块的一个或多个视频帧应用预测模型,以对一个或多个视频帧中的每一个获得阶段预测,预测模型被配置为对输入视频帧预测手术过程的多个阶段之一;
通过对一个或多个视频帧聚合阶段预测,对各个块生成聚合阶段预测;以及
基于聚合阶段预测,修改手术过程的视频,以包括对各个块的预测阶段的指示。
2. 根据权利要求1所述的方法,其中,对视频帧的阶段预测包括概率向量,概率向量包括视频帧属于手术过程的各个阶段的概率。
3. 根据权利要求2所述的方法,其中,对各个块生成聚合阶段预测包括:
对一个或多个视频帧的概率向量求平均,以对块生成平均概率向量;
将对块的聚合阶段预测生成为对应于平均概率向量中的最高概率的阶段;以及
确定聚合阶段预测的置信水平为平均概率向量中的最高概率。
4. 根据权利要求1所述的方法,进一步包括:
响应于确定聚合阶段预测的置信水平高于阈值;以及
确定块的预测阶段为聚合阶段预测。
5. 根据权利要求4所述的方法,进一步包括:
响应于确定聚合阶段预测的置信水平低于阈值,确定块的预测阶段为先前块的预测阶段。
6. 根据权利要求4所述的方法,进一步包括:
响应于确定块的预测阶段包括早于先前块的预测阶段的阶段或晚于后续块的预测阶段的阶段,将块的预测阶段修改为先前块的阶段预测。
7. 根据权利要求1所述的方法,进一步包括:训练预测模型以找到用于预测模型的一组参数,使得使用该组参数的损失函数的值小于使用另一组参数的损失函数的值,其中,训练是基于多个训练帧和训练帧的各个标签来执行的。
8. 根据权利要求7所述的方法,其中,训练预测模型包括:
接收多个训练帧之一是代表帧的指示;以及
将更高的权重分配给损失函数中的与代表帧相关联的项。
9. 根据权利要求7所述的方法,其中,训练预测模型包括:在多个阶段之间的逻辑关系不能被违反的约束下训练预测模型。
10. 根据权利要求7所述的方法,其中,训练预测模型基于多个不带标签的训练视频来进一步执行。
11. 一种方法,包括:
访问手术过程的视频,手术过程包括多个阶段;
将视频划分成一个或多个块,一个或多个块中的每一个包括一个或多个视频帧;
对于每个块:
对各个块中的一个或多个视频帧生成特征向量;
对特征向量应用预测模型,以对各个块生成阶段预测,阶段预测指示手术过程的阶段;

以及

修改手术过程的视频,以包括对各个块的阶段预测的指示。

12.根据权利要求11所述的方法,进一步包括:响应于确定阶段预测的置信水平低于阈值,将对各个块的阶段预测生成为先前块的阶段预测。

13.根据权利要求11所述的方法,其中,对块中的一个或多个视频帧生成特征向量包括:

将第二预测模型应用于一个或多个视频帧,以对每个视频帧生成预测向量;以及对一个或多个视频帧聚合预测向量,以生成特征向量。

14.根据权利要求13所述的方法,其中,聚合预测向量通过对一个或多个视频帧的预测向量求平均来执行。

15.根据权利要求11所述的方法,其中,预测模型包括门控循环单元(“GRU”)神经网络或长短期记忆(“LSTM”)神经网络中的一个或多个。

16.根据权利要求11所述的方法,还包括:

完善手术过程的视频的预测阶段;以及
使用完善的预测阶段修改手术过程的视频。

17.根据权利要求16所述的方法,其中,完善预测阶段包括:

将视频划分成多个区域;

对多个区域生成区域特征向量;

基于描述手术过程的视频的多个阶段之间的逻辑关系的规则,构建有向图;

对于每个区域:

基于有向图,识别区域的一个或多个相邻区域;

将区域和一个或多个相邻区域的区域特征向量进行组合,以生成组合的区域特征向量;以及

通过将组合的区域特征向量应用于被配置为对组合的区域特征向量预测完善的预测阶段的机器学习模型,生成完善的预测阶段。

18.根据权利要求16所述的方法,其中,完善预测阶段包括:基于视频的两个相邻预测阶段的组合特征向量来完善两个相邻预测阶段之间的边界。

19.一种方法,包括:

在手术过程期间访问手术过程的多个手术图像;

从多个手术图像生成第一特征向量;

从由多个手术图像导出的光流生成第二特征向量;

基于第一特征向量和第二特征向量,对多个手术图像确定识别到的活动;以及
使得用户界面被修改为在用户界面中呈现识别到的活动。

20.根据权利要求19所述的方法,其中,从多个手术图像生成第一特征向量包括:将三维卷积神经网络(3DCNN)模型应用于多个手术图像,3DCNN被配置为将一组图像分类为多个活动,并且第一特征向量包括各个活动存在于多个手术图像中的概率。

21.根据权利要求19所述的方法,其中,对多个手术图像确定识别到的活动包括:

将第一特征向量和第二特征向量进行组合,以生成组合特征向量;以及

将循环神经网络模型应用于组合特征向量,以确定识别到的活动。

22. 根据权利要求19所述的方法,进一步包括:

访问与手术过程的手术视频中的多个手术图像邻近的相邻手术图像;以及

通过将多个手术图像中的第一特征向量和第二特征向量中的至少一个与相邻手术图像的特征向量组合,并且通过将组合特征向量应用于被配置为预测识别到的活动的起点和终点的机器学习模型,调整识别到的活动的起点和终点。

23. 一种计算设备,包括:

处理器;以及

在其上存储有处理器可执行指令的非暂时性计算机可读介质,处理器可执行指令在由处理器执行时,使处理器:

将手术过程的视频划分成一个或多个块,手术过程包括多个阶段,并且一个或多个块中的每一个包括一个或多个手术图像;

对于每个块:

将预测模型应用于各个块的一个或多个手术图像,以对一个或多个手术图像中的每一个获得阶段预测,预测模型被配置为对输入手术图像预测手术过程的多个阶段之一;

通过对一个或多个手术图像聚合阶段预测,对各个块生成聚合阶段预测;以及

基于聚合阶段预测,修改手术过程的视频,以包括对各个块的预测阶段的指示。

24. 一种非暂时性计算机可读介质,包括处理器可执行指令,处理器可执行指令使处理器:

将手术过程的视频划分为一个或多个块,手术过程包括多个阶段,并且一个或多个块中的每一个包括一个或多个手术图像;

对于每个块:

将预测模型应用于各个块的一个或多个手术图像,以对一个或多个手术图像中的每一个获得阶段预测,预测模型被配置为对输入手术图像预测手术过程的多个阶段之一;

通过对一个或多个手术图像聚合阶段预测,对各个块生成聚合阶段预测;以及

基于聚合阶段预测,修改手术过程的视频,以包括对各个块的预测阶段的指示。

基于手术视频的手术工作流程和活动检测

技术领域

[0001] 本申请一般涉及到机器人手术和视频处理,更具体地,涉及到使用视频处理技术来检测手术视频中的手术工作流程阶段和活动。

背景技术

[0002] 随着机器人手术变得越来越流行,每天都有大量的手术视频被记录下来,特别是用于腹腔镜手术。这些视频包含有价值的信息,是用于诸如手术分析和新手术医生训练的任务的重要资源。然而,由于大量的手术视频,让作为稀有资源的高技能的医疗专业人员来检查长手术视频来识别手术的各个阶段并标记手术期间的所有有意义的事件是不可行的。此外,在手术过程期间分辨出手术阶段和异常活动是有益的,以便可以实时或接近实时地向手术医生提供信息和警告。

发明内容

[0003] 描述了用于经过视频处理检测手术工作流程阶段和活动的各种示例。一种示例方法包括访问手术过程的视频,手术过程包括多个阶段;将视频划分成一个或多个块,一个或多个块中的每一个都包括一个或多个视频帧;对于每个块:对各个块的一个或多个视频帧应用预测模型,以对一个或多个视频帧中的每一个获得阶段预测,预测模型被配置为对输入视频帧预测手术过程的多个阶段中的一个;通过对一个或多个视频帧聚合阶段预测,对各个块生成聚合阶段预测;以及,基于聚合阶段预测,修改手术过程的视频为包括对各个块的预测阶段的指示。

[0004] 另一种方法包括访问手术过程的视频,手术过程包括多个阶段;将视频划分成一个或多个块,一个或多个块中的每一个都包括一个或多个视频帧;对于每个块:对各个块中的一个或多个视频帧生成特征向量;对特征向量应用预测模型,以对各个块生成阶段预测,阶段预测指示手术过程的阶段;以及,修改手术过程的视频为包括对各个块的阶段预测的指示。

[0005] 另一种方法包括在手术过程期间访问手术过程的多个手术图像;从多个手术图像生成第一特征向量;从由多个手术图像导出的光流生成第二特征向量;基于第一特征向量和第二特征向量,对多个手术图像确定识别到的活动;以及,使用户界面被修改为在用户界面中呈现识别到的活动。

[0006] 一个示例计算设备包括处理器;以及在其上存储有处理器可执行指令的非暂时性计算机可读介质,当指令被处理器执行时,使处理器:将手术过程的视频划分成一个或多个块,手术过程包括多个阶段,并且一个或多个块中的每一个都包括一个或多个手术图像;对于每个块:对各个块的一个或多个手术图像应用预测模型,以对一个或多个手术图像中的每一个获得阶段预测,预测模型被配置为对输入手术图像预测手术过程的多个阶段中的一个;通过对一个或多个手术图像聚合阶段预测,对各个块生成聚合阶段预测;以及,基于聚合阶段预测,修改手术过程的视频以包括对各个块的预测阶段的指示。

[0007] 一个示例非暂时性计算机可读介质,其包括处理器可执行指令以使处理器:将手术过程的视频划分成一个或多个块,手术过程包括多个阶段,并且一个或多个块中的每一个都包括一个或多个手术图像;对于每个块:对各个块的一个或多个手术图像应用预测模型,以对一个或多个手术图像中的每一个获得阶段预测,预测模型被配置为对于输入手术图像预测手术过程的多个阶段中的一个;通过对一个或多个手术图像聚合阶段预测,对各个块生成聚合阶段预测;以及,基于聚合阶段预测,修改手术过程的视频以包括对各个块的预测阶段的指示。

[0008] 提及这些说明性示例不是为了限制或限定本公开的范围,而是为了提供示例以帮助理解本公开的范围。说明性示例在提供进一步的描述的详细描述中进行了讨论。通过查考本说明书,可以进一步理解各种示例提供的优点。

附图说明

[0009] 结合到本说明书中并构成本说明书一部分的附图示出了一个或多个特定示例,并且与示例的描述一起用于解释特定示例的原理和实现。

[0010] 图1A示出机器人手术系统的示例,其中,可以执行并利用本文描述的手术工作流程和活动检测,以在手术过程期间向手术医生提供反馈;

[0011] 图1B示出手术视频分析系统的示例,其中,可以执行并利用本文描述的手术工作流程和活动检测,以对一个或多个机器人手术系统提供分析结果;

[0012] 图2是示出被配置为基于手术视频来检测手术工作流程和活动的视频分析模块的各方面的框图;

[0013] 图3A示出手术视频、视频块和包含在手术视频中的视频帧之间的关系;

[0014] 图3B示出用于对视频块预测工作流程阶段的方法的示例;

[0015] 图3C示出用于对视频块预测工作流程阶段的另一方法的示例;

[0016] 图3D示出用于改进手术视频的阶段预测的方法的示例;

[0017] 图3E示出用于改进手术视频的阶段预测的方法中使用的有向图的示例;

[0018] 图4示出用于对手术视频识别手术活动的方法的示例;

[0019] 图5A示出手术工作流程阶段预测的输出的示例;

[0020] 图5B示出手术工作流程阶段预测和活动识别的输出的示例;

[0021] 图6示出用于检测手术视频中的手术工作流程阶段的过程的示例;

[0022] 图7A示出用于对视频块预测工作流程阶段的过程的示例;

[0023] 图7B示出用于对视频块预测工作流程阶段的过程的另一示例;

[0024] 图8示出用于检测手术视频中的手术活动的过程的过程的示例;以及

[0025] 图9示出适合于实现本文呈现的技巧和技术(techniques and technologies)的各方面的计算设备的示例。

具体实施方式

[0026] 本文在检测手术视频中的手术工作流程阶段和活动的背景下描述了示例。本领域普通技术人员将认识到,以下描述仅是说明性的,并且不旨在以任何方式进行限制。现在将详细参考附图中示出的示例的实施方式。在附图和下面的描述当中将使用相同的附图标记来

指代相同或类似的项目。

[0027] 为了清楚起见,不是本文描述的示例的所有常规特征都被示出和描述。当然,可以理解,在开发任何这样的实际实现中,为了实现开发人员的特定目标(诸如遵守与应用和业务相关的约束),必须做出许多特定于实现的决定,并且这些特定目标将因实施方式而异,也因开发人员而异。

[0028] 在检测手术视频中的手术 workflow 阶段和活动的说明性示例中,获得手术过程的视频。手术过程包括不同的阶段。为了检测手术视频中的不同阶段,视频被划分成几个块,每个块包含多个视频帧。块中的帧或帧的子集被提取并应用于阶段预测模型,以预测块的阶段。块的预测阶段通过预测每个帧的阶段并组合各个帧的预测阶段来生成。阶段预测模型包括机器学习模型,诸如神经网络模型。

[0029] 然后,检查并调整块的预测阶段,以消除具有低置信水平和/或违反手术过程的固有逻辑的预测。然后,利用预测阶段(诸如通过附加元数据或通过修改手术视频的可视内容)来注释手术视频。此外,预测阶段用于索引手术视频。

[0030] 除了检测手术过程之外,还在手术视频中检测活动,诸如手术任务或事件。为了检测活动,一组视频帧被一起分析以生成特征向量。该组视频帧还被分析以估计或提取光流,并且另一特征向量基于光流来生成。两个特征向量中的每一个或两者的组合被发送到活动识别模型,以识别帧中的活动。活动识别模型还包括一个机器学习模型,诸如神经网络模型。在手术过程期间,实时或接近实时地执行活动识别。向进行手术的手术医生提供警告,提醒他或她注意异常活动。

[0031] 本文呈现的技术改进了对大量手术视频的内容管理。使用本文呈现的技术,可以用检测到的阶段和活动来注释手术视频,从而允许更好地索引和组织手术视频以及更有效地检索相关内容。作为结果,可以增加对涉及手术过程或活动的搜索查询的响应速度,并且可以减少手术视频的相关部分的检索时间。此外,活动检测还可以通过向手术医生提供关于异常活动的实时或接近实时的警告来提高机器人手术系统的安全性。除了本文提及的技术优点以外的其他技术优点也可以从本文公开的技术的实现中认识到。

[0032] 给出该说明性示例是为了向读者介绍本文讨论的一般主题,并且本公开不限于该示例。以下部分描述了从手术视频检测手术 workflow 阶段和活动的各种附加非限制性和非详尽示例。

[0033] 现在参考图1A,图1A示出配置有手术 workflow 阶段和活动检测能力的机器人手术系统100A的示例。该示例机器人手术系统100a包括被配置为对患者130操作的机器人手术设备114和用于控制机器人手术设备114的中央控制器112。机器人手术系统100A还包括连接到中央控制器112和机器人手术设备114的手术医生控制台104。手术医生控制台104由手术医生102操作,以控制和监视使用机器人手术设备114执行的手术。除了这些组件之外,机器人手术系统100A还可以包括可由手术室中的其他人员使用的附加工作站(图1A中未示出),例如,用于观看从机器人手术设备114发送的手术信息、视频等。在该示例中,机器人手术设备114、中央控制器112、手术医生控制台104和其他工作站彼此直接连接,尽管在一些示例中,它们可以使用网络(诸如局域网(“LAN”)、广域网(“WAN”)或本领域公知的连接机器人手术系统100A中的各个工作站的任何其他网络拓扑(networking topology))来连接。

[0034] 机器人手术设备114可以是用于对患者执行手术过程的任何合适的机器人系统。

例如,机器人手术设备114可以具有连接到基座的一个或多个机器人手臂。机器人手臂可以由工具控制器116操纵,工具控制器116可以包括一个或多个用户界面设备,诸如操纵杆、旋钮、手柄或其他可旋转或可平移的设备,以影响一个或多个机器人手臂的移动。机器人手臂可以配备有一个或多个手术工具,以执行手术过程的各个方面。例如,机器人手臂可以配备有手术工具126A-126C。每个手术工具都可以由手术医生102通过手术医生控制台104和工具控制器116来控制。

[0035] 此外,机器人手术设备114配备有被配置为在手术期间提供操作现场的视图来引导手术医生102的一个或多个相机128,诸如内窥镜相机。在一些示例中,如图1A中所示,相机128可以附接到由工具控制器116控制的机器人手术设备114的机器人手臂中的一个。在其他示例中,相机128可以附接到与机器人手臂分开的机器人手术设备114的机械结构,诸如用于携带相机128的专用手臂。

[0036] 不同的机器人手术设备114可以被配置为用于特定类型的手术,诸如心血管手术、胃肠手术、妇科手术、移植手术、神经外科、肌肉骨骼手术等,而有些可以具有多种不同的用途。作为结果,根据不同的示例,可以采用不同类型的手术机器人,包括诸如用于内窥镜检查过程的没有机器人手臂的手术机器人。应该理解,虽然只描绘了一个机器人手术设备114,但是可以在机器人手术系统100A内采用任何合适的数量的机器人手术设备。

[0037] 在一些示例中,机器人手术设备(或各个控制器)可以被配置为在手术过程期间记录数据。例如,由机器人手术设备114执行的手术过程的图像和视频也可以被记录和存储,以供进一步使用。例如,存储服务器124可以由机器人手术设备114采用,以存储由相机128捕捉的手术过程的手术视频132。

[0038] 在图1A中示出的示例中,由相机128捕捉的机器人手术过程的手术视频132也可以被传输到手术医生控制台104,并且被实时地显示在视频监视器108上,使得手术医生102可以在手术工具126正在被用来对患者130操作的同时观看过程。在该示例中,手术医生102使用手术医生控制台104来控制手术工具126和相机128,并且使用手术医生控制台104上的控件106来通过向工具控制器116发送相应的控制信号110来操纵手术工具126和相机128。

[0039] 如图1A中所示,中央控制器112还包括视频分析模块120,以处理在手术过程期间捕捉的手术视频132。视频分析模块120分析手术视频132,以预测正在进行的手术过程的工作流阶段118或以识别在该过程期间发生的各种活动122。如果识别到的活动122包括异常事件(诸如出血),则中央控制器112生成一个或多个活动警告136,该一个或多个活动警告136被呈现在手术医生控制台104的视频监视器108上,以向手术医生102通知异常活动。在该示例中,阶段预测118(在本文也称为“预测阶段118”)和识别到的活动122与手术视频132一起被存储在存储服务器124中,以供未来使用,诸如存档、索引、手术后分析、训练新手术医生等。

[0040] 应当理解,尽管图1A示出在机器人手术系统100A的背景下呈现的手术工作流阶段和活动检测的技术,但是该技术可以在其他类型的系统和设置中实现。例如,该技术可以在与机器人手术系统100A分开的计算设备中实现和/或在手术过程完成之后离线执行。图1B示出计算环境100B的示例,其中,手术视频分析系统140被配置为执行本文描述的手术工作流和活动检测,以提供对一个或多个机器人手术系统的分析结果。

[0041] 在图1B中所示的示例中,手术视频分析系统140包括如上面参考图1A所述的视频

处理模块120。视频分析模块120被配置为对一个或多个机器人手术系统146A-146C(在本文可以单独称为机器人手术系统146或统称为机器人手术系统146)执行手术工作流和活动检测。在一个示例中,机器人手术系统146以与上面参考图1A讨论的机器人手术系统100A类似的方式被配置,只是机器人手术系统146不包括视频分析模块120。取而代之的是,机器人手术系统146向手术视频分析系统140发送记录的手术视频132,以供分析。在一些示例中,手术视频132通过网络148(诸如LAN、WAN或本领域公知的将机器人手术系统146连接到手术视频分析系统140的任何其他网络拓扑)来发送。

[0042] 视频分析模块120接收手术视频132,并且执行如上面简要描述并在下面更详细地描述的手术工作流和活动检测。在图1B中所示的示例中,识别到的活动122或阶段预测118被利用来注释手术视频132,以生成带注释的手术视频144。如下面将详细讨论的,注释可以通过生成指示阶段和活动的元数据并将该元数据附加到手术视频来执行。可替换或附加地,注释可以通过修改视频的内容以(诸如通过将指示阶段和活动的文本、图像、图标或徽标插入到视频帧中来)标记检测到的阶段和活动来执行。带注释的手术视频144被存储在手术视频分析系统140本地的数据存储142中。在另一示例中,带注释的手术视频144可以被存储在可由手术视频分析系统140、机器人手术系统146或任何其他授权系统通过网络访问的存储库中。

[0043] 图2是示出被配置为基于手术视频132来检测手术工作流和活动的视频分析模块120的各方面的框图。如图2中所示,由视频分析模块120执行的手术工作流和活动检测可以包括几个时期(stage):预测模型训练时期202、工作流阶段预测时期204和活动识别时期206。预测模型训练时期202构建并训练要在其他两个时期中使用的各种预测模型210A-210C(在本文可以单独称为预测模型210或统称为预测模型210)。例如,预测模型210可以包括用于对特定手术过程(诸如胆囊切除术、肾切除术、结肠切除术等)预测工作流阶段的模型。预测模型210还可以包括用于从手术视频132中分辨出或识别特定活动(诸如缝合、切除、烧灼、切割、冲洗和抽吸等手术任务或出血、胆汁渗漏等重要事件等)的模型。尽管如此,其他类型的预测模型可以在根据本公开的其他示例中采用。

[0044] 预测模型306可以是机器学习(“ML”)模型,诸如卷积神经网络(“CNN”) (例如,由来自加利福尼亚州山景城的Google LLC提供的初始神经网络、残差神经网络(“Resnet”)或NASNET)或者循环神经网络(例如,长短期记忆(“LSTM”)模型或门控循环单元(“GRU”)模型)。预测模型306还可以是任何其他合适的ML模型,可以被训练来对视频帧预测阶段或活动,诸如三维CNN(“3DCNN”)、动态时间扭曲(“DTW”)技术、隐马尔可夫模型(“HMM”)等或者一种或多种这样的技术的组合(例如,CNN-HMM或MCNN(多尺度卷积神经网络))。视频分析模块120可以采用相同类型的预测模型或不同类型的预测模型,以用于手术过程和活动检测。

[0045] 为了训练本示例中的各种预测模型210,生成了针对每个预测模型210的训练样本212。针对特定预测模型210的训练样本212可以包括输入视频帧(或视频帧的输入特征)和对应于输入视频帧(或输入特征)的标签。例如,对于被用来基于输入视频帧来识别出血事件的预测模型210,输入可以是视频帧本身或从帧中提取的特征,并且标签可以包括显示在输入帧中是否发生出血的标志。类似地,对于被用来对视频帧预测工作流阶段的预测模型210,输入可以包括视频帧或从视频帧中提取的特征,并且标签可以包括指示输入视频帧所属的阶段的数字或者指示视频帧属于不同阶段的概率的向量。

[0046] 训练过程包括找到最小化用于预测模型210的损失函数的用于预测模型210的一组参数的迭代操作。每次迭代可以涉及为预测模型210找到一组参数,使得使用该组参数的损失函数的值小于使用先前迭代中的另一组参数的损失函数的值。损失函数可以被构造为测量使用预测模型210预测的输出和训练样本212中包含的标签之间的差。一旦识别到该组参数,预测模型210就已经被训练,并且可以按设计用于预测。

[0047] 除了训练样本212之外,其他信息也可以被采用来完善预测模型210的训练过程。例如,在手术视频中,一些视频帧是手术过程的代表帧,诸如包含手术过程期间的代表动作或事件的帧。一旦识别到这些视频帧,这些视频帧就可以为接近代表帧的视频帧的手术过程提供线索。例如,在胆囊切除术中,可以在胆囊三角区域(calot triangle area)或胆囊和肝脏之间进行切除。描述这些区域中的活动的视频帧可以指示视频中正在发生的胆囊切除术的切除阶段。在另一示例中,描述胆囊从肝床完全脱离的视频帧可以是指手术过程结束的代表帧。在又一示例中,特定手术工具的存在也可以是手术 workflow 阶段的指示符。例如,从视频帧检测到的吻合器手术工具可以指示视频帧描述手术的袖状胃切除术胃吻合阶段。

[0048] 这些代表帧可以用指示其代表性的附加标签来标记。在训练阶段预测模型210期间,可以将更高的权重分配给对应于这些代表帧的损失函数的项(term)。作为结果,在预测 workflow 阶段时,训练的预测模型210可以向与代表帧相似的输入帧给予更多权重。

[0049] 此外,手术过程逻辑214可以被合并到预测模型训练时期202中,以确保由阶段预测模型210预测的阶段不违反手术过程逻辑214。手术过程一般在过程的不同阶段之间具有固有逻辑。例如,胆囊包装(gallbladder packaging)只能在胆囊切除术中切除胆囊之后才能进行,而胃空肠吻合术只能在胃旁路手术中切断空肠之后才能进行。手术过程的阶段之间的固有逻辑关系可以被利用来促进阶段预测。

[0050] 根据本文呈现的本公开的一些方面,手术过程的 workflow 阶段之间的逻辑关系可以被表示为对上面参考训练预测模型210讨论的优化问题的一个或多个约束条件。可以构建对违反约束条件进行惩罚的训练损失函数,以便训练可以将 workflow 阶段逻辑约束条件考虑在内。可替换或附加地,描述预测输出的当前特征和时间依赖的结构(诸如有向图)可以用于调整或完善预测模型210的特征和预测。在示例实现中,特征从当前视频帧中提取,并且与来自有向图中示出的先前帧和后续帧的特征组合。以这种方式生成的特征可以固有地合并训练样本212中的帧之间的时间关系,并因此可以固有地合并训练样本212中的逻辑关系。因此,使用这些特征训练的预测模型210可以捕捉手术 workflow 的各个阶段之间的逻辑关系。

[0051] 如上所述,手术视频通常较长,并且可以持续几个小时或更长时间。获得训练样本212中的标签需要医疗专业人员手动审阅这些视频的专业知识,因此是耗时的任务。作为结果,对于医疗专业人员来说,对所有的手术视频添加标签是不切实际的,因此大量的手术视频仍然是不带标签的。这些不带标签的手术视频(的取得)可能比带标签的手术视频更便宜,也可以被采用来训练预测模型210。例如,对于不带标签的训练视频,可以应用预测模型210来预测其阶段。如果预测阶段违反手术过程的固有逻辑,则可以通过在损失函数中引入项来惩罚该不带标签的视频。也就是说,可以利用其预测阶段违反手术过程的固有逻辑的那些不带标签的训练视频来重新定义训练损失函数。作为结果,训练损失函数可以是如上

面讨论的带标签的视频的损失和基于不带标签的视频的手术步骤逻辑的损失的组合。

[0052] 另一方面,如果使用预测模型210的对不带标签的训练视频的预测阶段不违反手术过程的固有逻辑,则损失函数可以保持不变。作为结果,不带标签的训练视频只有在违反手术过程的固有逻辑时才会对损失函数产生影响。相比之下,带标签的视频可以对损失函数产生影响,而不管它们是否违反手术过程的固有逻辑。

[0053] 应该理解,上述示例仅仅是说明性的。在预测模型训练时期202期间,可以以各种其他方式利用不带标签的视频。例如,不带标签的视频可以用作训练样本212,以例如包括无监督损失(诸如预测的平滑度),以及用于强制实施手术过程的固有逻辑。如此,即使不带标签的视频的预测阶段不违反手术过程的固有逻辑,不带标签的视频也可以在损失函数中具有相应的项。

[0054] 类似地,在活动识别模型210的训练期间,可以利用辅助信息。准备训练样本212可以涉及针对要识别的活动的类型,对输入视频手动添加标签。在长达数小时(hour-long)的手术视频中,对每一次手术活动的发生添加标签是具有挑战性并辛苦的。例如,抓取动作通常一次持续几秒钟,但是在手术过程中会发生多次。本文描述的训练机制允许医疗专业人员对可管理次数的这些类型的动作的发生添加标签,并且将其余的标记为“未知”。在预测模型210的训练期间,“未知”标签不被使用,并且被排除作为用于这些特定标签的训练损失函数的一部分。这可以防止不带标签的视频帧被视为负面示例,即,目标活动被识别为不存在于输入视频帧中,尽管稍后可以通过向经训练的模型提供视频以供分析来确定这些“未知”标签。可替换或附加地,可以生成经选择的正面示例和负面示例的集合,并且可以使用这些正面示例和负面示例来微调模型。

[0055] 此外,本文描述的训练机制还允许分层标签或多标签。手术任务和细粒度更高的(more fine-grained)子任务可以重叠,并且一个任务可以包含多个子任务。作为结果,可以对相同的视频帧标记多个标签。例如,多个解剖结构和多个手术器械可以出现在相同的视频帧中。因此,多个手术动作或任务可以在相同的视频中同时发生,并且可能伴随着事件。通过允许给定帧中的多个标签,包含在训练视频帧中的潜在知识可以由视频分析模块充分利用,以训练活动识别模型210。

[0056] 尽管上述训练机制主要集中于训练预测模型210,但是这些训练机制还可以用于微调由其他数据集训练的现有预测模型210。例如,在某些情况下,预测模型210可能已经使用非手术视频帧或图像进行预训练。在这些情况下,预测模型210可以使用包含如上面讨论的手术视频和其他辅助信息的训练样本212来重新训练。

[0057] 预测模型训练时期202输出包括经训练的阶段预测模型218和经训练的活动识别模型220的经训练的预测模型210。经训练的阶段预测模型218可以在工作流阶段预测时期204中用来对输入手术视频132生成阶段预测188。经训练的活动识别模型220可以用于识别手术视频132中的活动,以在活动识别时期206中生成识别到的活动122。

[0058] 工作流阶段预测时期204和活动识别时期206可以在具有分开的模型的一些示例中独立地运行。例如,工作流阶段预测时期204可以在不识别手术活动的情况下,对手术视频132应用经训练的阶段预测模型218,以生成阶段预测118。类似地,活动识别时期206可以在不识别工作流阶段的情况下,对手术视频132或手术视频132的一部分应用经训练的活动识别模型220,以识别输入视频中发生的活动,。

[0059] 或者, workflows阶段预测时期204和活动识别时期206可以使用另一个的输出作为输入, 从而按照以一个时期来顺序地操作。例如, 对于给定的手术视频132, 视频分析模块120可以首先执行 workflows阶段预测时期204, 以对手术视频132中的帧生成阶段预测。然后, 对于特定阶段, 视频分析模块120可以使用被预测在该特定阶段中的视频帧, 进入活动识别时期206中, 以识别在此阶段期间发生的活动122。在另一实施方式中, 视频分析模块120可以首先进入活动识别时期206中, 以对手术视频132或手术视频132的一部分生成识别到的活动122。然后, 可以在 workflows阶段预测时期204期间利用识别到的活动122, 以促进对手术视频132的阶段的预测。下面将参考图3-8, 提供关于 workflows阶段预测时期204和活动识别时期206的附加细节。

[0060] 图3A-3C示出用于对 workflows阶段预测时期204中的手术视频132预测 workflows阶段的方法的示例。图3A-3C将结合图5A来呈现, 其中, 示出了带有生成的阶段预测118的手术视频132。图3A示出手术视频132。为了对手术视频132生成阶段预测118, 手术视频132可以被划分成多个视频块302A-302N(其在本文可以单独称为视频块302或统称为视频块302)。视频块302中的每一个都可以包括多个视频帧或图像304A-304D(其在本文可以单独地称为视频帧或图像304或者统称为视频帧或图像304)。对于不同的视频块302而言, 视频块302的大小(即包含在视频块302中的视频帧304的数量)可以相同或不同。

[0061] 图3B示出用于对视频块302预测 workflows阶段的方法的一个示例。在图3B中, (诸如通过对视频块302的每一秒选择一个或两个帧,) 可以从视频块302提取多个视频帧304。视频帧304中的每一个都可以被输入到阶段预测模型306中, 阶段预测模型306可以是CNN, 诸如初始神经网络、“Resnet”或NASNET。

[0062] 阶段预测模型306可以对视频帧304中的每一个生成阶段预测308。阶段预测308可以包括指示由阶段预测模型306预测的视频帧304的 workflows阶段的单个值。在一些实施方式中, 阶段预测308可以包括包含对输入视频帧304预测的不同阶段的概率的向量。例如, 对于四阶段手术过程而言, 阶段预测308可以包括向量 $p = [p_1, p_2, p_3, p_4]$, 其中, p_i 代表输入视频帧304处于阶段i的概率。

[0063] 通过将多个视频帧304馈送到阶段预测模型306中, 可以生成多个阶段预测308。这些阶段预测308可以用作对各个帧的阶段预测。或者, 这些阶段预测308可以被聚合, 以对视频块302生成聚合阶段预测。聚合可以帮助降低噪声(诸如预测误差)对针对视频块302的阶段预测的影响。在一个实施方式中, 可以通过将针对输入视频帧304的阶段预测308进行按时间平均来执行聚合, 即:

$$[0064] \quad p_{blk} = \frac{1}{M} \sum_{i=1}^M p_m \quad (1)$$

[0065] 其中, $p_m = [p_{m1}, p_{m2}, p_{m3}, \dots, p_{mT}]$ 是针对T阶段手术过程中的 m^{th} 输入视频帧304的预测向量; M是被馈送到阶段预测模型306中的视频帧的数量; 以及, p_{blk} 是针对视频块302的聚合预测向量。

[0066] 然后, 可以分析聚合阶段预测310, 以确定与聚合阶段预测310相关联的置信水平是否足够高, 从而可以信任预测结果。在上述示例中, 预测向量 p_{blk} 包含输入视频块302属于各个 workflows阶段的概率。预测向量 p_{blk} 中的最高概率可以用于指示预测的置信水平。对应于

最高概率的工作流阶段可以被选择为对视频块302的预测阶段。

[0067] 例如,三个视频帧304可以被选择以用于视频块302,并且三个视频帧304中的每一个都可以被馈送到阶段预测模型306中,以对四阶段手术过程生成三个阶段预测308: $p_1 = [0.1, 0.5, 0.2, 0.2]$ 、 $p_2 = [0.1, 0.4, 0.3, 0.2]$ 和 $p_3 = [0.1, 0.8, 0.1, 0]$ 。这意味着针对第一视频帧304属于手术过程的第一阶段至第四阶段的概率分别为0.1、0.5、0.2和0.2。类似地,针对第二视频帧304属于手术过程的第一阶段至第四阶段的概率分别为0.1、0.4、0.3和0.2,并且针对第三视频帧304属于手术过程的第一阶段至第四阶段的概率分别为0.1、0.8、0.1和0。

[0068] 聚合阶段预测310可以被生成为如上面等式(1)中所讨论的三个阶段预测308的平均值,即 $p_{blk} = [0.1, 0.57, 0.2, 0.13]$ 。在该情况下,最高概率是0.57,因此对当前块的相应阶段预测是阶段2。在选择阶段2作为对当前块的最终阶段预测之前,最高概率0.57可以用作阶段预测的置信水平,并且与阈值进行比较,以确定阶段预测是否可靠。如果阈值设置为0.55,则阶段预测(即上述示例中的阶段2)可以被选择作为对块的阶段预测。另一方面,如果阈值设置为0.7,则当前阶段预测的置信水平不够可靠,不能用作对块的阶段预测。在后一种情况下,对先前块的阶段预测可以用作对当前块的阶段预测。这在图5A中示出,其中,示出了对视频块302中的每一个生成的阶段预测。在该图中,手术视频132的第三视频块具有低于置信阈值的置信水平,因此对此块的聚合阶段预测被分配为对第二块的阶段预测。

[0069] 应当理解,上述聚合仅用于说明,并且不应被解释为限制。对于视频块的多个帧而言,可以利用聚合阶段预测的各种其他方式。例如,可以通过应用各种线性或非线性的函数来聚合阶段预测308。在阶段预测308包括指示针对帧的预测工作流阶段的单个值的情况下,可以通过在多个阶段预测308之间进行多数投票来执行聚合阶段预测308。

[0070] 此外,可以对视频帧304中的每一个而不是视频块302中的每一个生成聚合阶段预测310。例如,大小为M的滑动窗口可以被采用并应用于阶段预测308。如等式(1)所示,对于视频帧304中的每一帧,其阶段预测308可以被更新为其相邻M个帧的阶段预测308的平均值。作为结果,对视频帧304的聚合阶段预测310通过使用其相邻的视频帧304,变为平滑版本的阶段预测308,从而消除阶段预测308中的随机噪声并提高其可靠性。类似于对视频块302的聚合阶段预测310,对帧的聚合阶段预测310的置信水平与置信阈值进行比较。如果置信水平高于置信阈值,则聚合阶段预测310用来确定对视频帧304的预测阶段。如果置信水平低于置信阈值,则对先前视频帧的预测阶段可以用作对当前视频帧的预测阶段。

[0071] 图3C示出用于对视频块302预测工作流阶段的方法的另一概述的示例。类似于图3B中所示的方法,(例如通过对视频块302的每一秒选择一个或两个帧,)可以从视频块302中提取多个视频帧304。选择的视频帧304中的每一个都可以被分析,以提取特征向量312。然后,用于视频帧304的特征向量可以聚合,以生成聚合特征向量,其随后被发送到阶段预测模型314,以对视频块302生成阶段预测。

[0072] 对视频帧304的特征提取可以以各种方式执行。在一些实施方式中,可以通过对视频帧304应用卷积神经网络来提取特征向量312。例如,特征向量312可以通过使用图3B中的阶段预测模型306来生成作为阶段预测向量。类似地,特征向量312可以以类似于如上面参考图3B描述的阶段预测聚合的方式来聚合。或者,聚合的特征向量可以通过级联(concatenating)特征向量312来聚合,以形成具有比单个特征向量更高的维度的向量。聚

合特征向量312可以帮助减少特征提取中的噪声影响,以及减小对阶段预测模型314的输入的大小。

[0073] 在该示例方法中使用的阶段预测模型314可以将聚合特征向量作为输入,并且输出用于当前块的阶段预测以及置信水平。考虑到手术过程的工作流阶段的顺序性质,循环神经网络可以用作阶段预测模型314,诸如LSTM模型或GRU模型。在循环神经网络中,节点之间的连接沿序列形成有向图,这允许神经网络展示对时间序列的时间动态行为。类似于图3B中所示的示例,由阶段预测模型314生成的阶段预测可以包括指示当前块属于各个阶段的概率的概率向量。阶段预测还可以与置信水平相关联,诸如概率向量中的最高概率。然后,如上面参考图3B所讨论的,置信水平可以与置信阈值进行比较,以确定由阶段预测模型314生成的阶段预测或先前块的阶段预测是否应该用于当前块。

[0074] 上面在图3B或图3C中讨论的阶段预测方法可以应用于手术视频132的视频块302中的每一个。工作流阶段预测时期204的输出在图5A中示出,其中,手术视频132的每个视频块302都可以被预测为与手术过程的工作流阶段之一相关联。

[0075] 除了上述方法之外,其他信息也可以用来完善或校正对视频块302的阶段预测。例如,如上所述,手术过程在各项工作流阶段之间具有固有逻辑关系。例如,胆囊包装只有在胆囊切除术中切除胆囊之后才能进行。换句话说,视频块302不可以具有晚于对后续视频块302的阶段预测的阶段预测。图5A示出该场景的示例。在图5A中,手术视频132的第5视频块被预测为处于阶段3中,但是下一个块(即第6视频块)被预测为处于阶段2中。因此,对第5视频块的阶段预测违反手术过程的固有逻辑。为了校正阶段预测中的不一致性,先前视频块302的阶段预测(即该示例中的第4视频块)可以用来替换对当前块的预测。类似地,如果视频块302被预测为处于其先前视频块302的阶段之前的阶段中,则可以用对其先前视频块302的阶段预测替换对当前视频块302的阶段预测。应该理解,该示例仅用于说明,并且不应解释为限制。利用手术过程的固有逻辑的各种其他方式可以被采用来修改阶段预测。

[0076] 通过使用工作流的固有逻辑,对视频的阶段预测可以进一步完善或改进。图3D和图3E示出用于改进手术视频的阶段预测的方法的示例。在图3D中所示的示例中,视频分析模块120将手术视频132划分成多个区域322A-322E(其在本文可以单独地称为区域322或统称为区域322)。在一个示例中,一个区域包括被预测为处于相同的阶段中的帧或视频块。例如,如图5A中所示,被预测为处于阶段1中的视频帧可以被包括在区域1 322A中;被预测为处于阶段2中的视频帧可以被包括在区域2 322B中,以此类推。

[0077] 对各个区域322生成区域特征向量324A-324E(其在本文可以单独称为区域特征向量324或统称为区域特征向量324)。在一个示例中,视频分析模块120通过组合对相应区域322中的帧的阶段预测308(如图3B中所示)或通过组合针对相应区域322中的帧的特征向量312(如图3C中所示)来生成区域特征向量324。该组合可以通过对针对区域322中的帧的特征向量312或阶段预测308求平均来执行。或者,该组合可以通过利用机器学习模型(诸如被配置为接受特征向量312或阶段预测308的序列并输出区域特征向量324的LSTM)来执行。

[0078] 应当理解,上述生成区域特征向量324的方法仅用于说明,并且不应被解释为限制。各种方式可以被采用来生成区域特征向量324,例如,除特征向量312或阶段预测308之外的特征可以从区域322的帧中提取,并且这些特征可以使用各种线性和非线性的函数来组合,以生成区域特征向量324。

[0079] 基于区域322,视频分析模块120构建有向图,以对区域322之间的时间关系进行建模。视频分析模块120基于各种规则(诸如反映工作流的阶段之间的逻辑关系的规则)来构建有向图。有向图300的示例在图3E中示出。有向图包括分别表示图3D中所示的区域322A-322E的多个节点326A-326E。有向图300的箭头328A-328F指示节点326之间的时间关系,并因此指示区域322之间的时间关系。例如,箭头328A指示节点326B跟随节点326A,而箭头328D指示节点326D跟随节点326B。由有向图300识别的时间关系还指示节点326之间的相邻关系,并且因此指示区域322之间的相邻关系。例如,基于有向图300,可以确定节点326C具有三个邻居:节点326A、节点326B和节点326D,因此区域322C具有三个邻居:区域322A、区域322B和区域322D。

[0080] 基于相邻关系,视频分析模块120完善对手术视频132的阶段预测。在一个示例中,视频分析模块120通过将区域322的区域特征向量324与其相邻区域322的区域特征向量324组合来更新与区域322相关联的区域特征向量324。在上述示例中,通过将区域特征向量324C与区域特征向量324A、区域特征向量324B和区域特征向量324D组合来更新区域特征向量324C。将当前区域的区域特征向量表示为 f_c ,将其相邻区域的区域特征向量表示为 $f_{n1}, f_{n2}, \dots, f_{nN}$,其中N是当前区域的相邻区域数。当前区域的更新后的区域特征向量 f'_c 可以表达为:

$$[0081] \quad f'_c = g(f_c, f_{n1}, f_{n2}, \dots, f_{nN}) \quad (2)$$

[0082] 其中 $g()$ 是用于组合这些区域特征向量的组合函数。在一个实施方式中, $g()$ 表示这些区域特征向量的加权和,并且分配给 f_c 的权重高于分配给 $f_{n1}, f_{n2}, \dots, f_{nN}$ 的权重。在其他实施方式中,组合函数 $g()$ 可以是配置为接受多个区域特征向量作为输入并输出的组合区域特征向量的机器学习模型,诸如图神经网络。

[0083] 在进一步的实施方式中,更新后的区域特征向量可以使用基于如上所述的有向图300的更新的区域特征向量再次更新。也就是说,

$$[0084] \quad f'_c = g(f'_c, f'_{n1}, f'_{n2}, \dots, f'_{nN}) \quad (3)$$

[0085] 如此,更新后的区域特征向量可以受到其最近邻居以及最近邻居的邻居的影响。然后,更新后的区域特征向量被发送到被配置为对输入区域特征向量预测阶段的机器学习模型。在一个示例中,机器学习模型是全连接的神经网络,其中,网络的输入层具有与更新后的区域特征向量的维度相同的维度,并且输出层具有等同于工作流的阶段的数量的维度。对于给定的输入特征向量,全连接的神经网络可以预测输出阶段之一作为针对输入的相应阶段。给定更新后的区域特征向量,其他类型的机器学习模型(诸如类似于阶段预测模型306或阶段预测模型314的机器学习模型)也可以用来预测阶段。

[0086] 对手术视频132的阶段预测的另一种完善涉及预测阶段的边界完善。在一些实施方式(诸如上面参考图3A-3C描述的基于块的方法)中,预测阶段的边界可能由于使用视频块作为阶段预测的单位而偏离阶段的实际边界。换句话说,视频块可能包含来自两个相邻阶段的视频帧,并且阶段的实际边界可能在块的中间,而不是块的边缘(border)。为了校正预测阶段边界,视频分析模块120将两个相邻阶段的特征向量组合在一起,并且将组合的特征馈送到被配置为预测两个相邻阶段之间的边界的机器学习模型中。

[0087] 例如,视频分析模块120可以使用上面参考图3D讨论的区域的区域特征向量作为阶段的特征,或者使用上面参考图3B讨论的相邻区域中的帧的特征向量。两个相邻阶段的

区域特征向量的组合可以通过对区域特征向量或对帧特征向量应用递归神经网络(诸如LSTM)来执行,或者通过对这些特征向量应用一维卷积神经网络来执行。然后,组合特征向量被输入到被训练为预测两个相邻阶段的边界的机器学习模型。机器学习模型可以是神经网络,诸如全连接的神经网络或任何其他类型的适当训练(properly training)的机器学习模型。

[0088] 现在参考图4,图4示出用于在活动识别时期206中从手术视频132中识别手术活动(诸如手术任务或事件)的方法的示例。图4将结合图5B来讨论,其中,示出了活动识别时期206的输出的示例。在图4所示的示例中,从视频帧302中提取的一组视频块304可以由特征提取模型402用来提取特征向量408。在一个实施方式中,特征提取模型402可以包括被训练为基于一组输入视频帧来生成特征向量的3D卷积神经网络(“3DCNN”)。此外,可以从一组视频帧304中提取光流406。光流是视频帧304中的对象、表面和边缘(edge)的外观运动的图案。可以使用本领域公知的任何光流提取方法从视频帧304中提取光流406。提取的光流406被馈送到光流特征提取模型404中,以生成特征向量410。类似地,光流特征提取模型404还可以包括被训练为基于光流输入来生成特征向量的3DCNN。

[0089] 在一些实施方式中,特征向量408和特征向量410中的每一个都可以包括包含具有不同活动的一组视频帧304的概率的概率向量。例如,特征向量408或特征向量410可以包括向量 $q=[q_1, q_2, q_3, q_4, q_5, q_6]$,其中, q_i 表示一组视频帧304包含活动i的概率。然后,特征向量408和特征向量410可以被组合,并且由在预测模型训练时期202期间训练的活动识别模型412用来生成识别到的活动122。特征向量408和特征向量410的组合可以包括但不限于对两个特征向量求平均、级联两个特征向量、对每个向量元素选择两个值中的较大值、或者对每个向量元素选择两个值中的较小值。活动识别模型412可以包括神经网络模型,诸如被训练为基于从视频帧和光流生成的特征向量来识别手术活动的循环神经网络。

[0090] 应当理解,虽然图4示出活动识别模型412用来输出识别到的活动122,但是可以通过组合特征向量408和特征向量410而不使用活动识别模型412来生成识别到的活动122。组合特征向量可以与阈值进行比较,并且对应于高于阈值的概率的活动可以被输出为识别到的活动122。

[0091] 还应当理解,虽然图4中所示的示例方法利用特征提取模型402和光流特征提取模型404,但是视频分析模块120可以在活动识别时期206中利用其他类型的模型和方法。例如,与图3B和图3C中所示的方法类似,类似于阶段预测模型306的预测模型可以被采用来逐帧地识别活动122。

[0092] 基于在手术过程期间检测到的事件或任务,可以生成警报,并且将警报呈现给手术医生102,以通知他或她检测到的事件或任务。例如,如果在活动识别时期206中使用的预测模型被训练为识别出血事件,则可以从在手术过程期间捕捉的手术视频132中识别类似的出血事件。警告消息或其他类型的反馈可以由视频分析模块120提供,以通知手术医生102发生出血。类似地,如果预测模型被进一步训练为识别不同类型的出血(诸如轻微出血和大出血),则视频分析模块120可以识别这些不同类型的出血,并且相应地通知手术医生102。

[0093] 图5B示出 workflow 阶段预测时期204和活动识别时期206的示例输出。如图5B中所示,除了对每个块302的阶段预测之外,包括手术任务和事件的识别到的活动122还可以用

来注释手术视频132的相应部分。识别到的活动122还可以相对于各个活动的起点和终点进一步完善。在一个示例中,视频分析模块120通过将活动的相邻视频帧考虑在内来执行对识别到的活动的完善。例如,可以通过考虑事件C之前和之后的视频帧(诸如先前块510中的视频帧和后续块512中的视频帧)来完善图5B中所示的阶段3中的事件C的起点和终点。应当理解,相邻帧可以通过除了基于视频块之外的多种方式(诸如通过选择活动之前和活动之后的预定数量的帧而不管视频块)来选择。

[0094] 类似于上面讨论的对检测到的 workflow 阶段的边界完善,视频分析模块通过应用递归神经网络(诸如LSTM或一维卷积神经网络)来组合包含检测到的活动的帧和相邻帧的特征向量,以生成组合特征向量。然后,组合特征向量被馈送到被配置为预测活动的起点和终点的机器学习模型(诸如全连接的神经网络或任何其他类型的适当训练的机器学习模型)中,以预测检测到的活动的起点和终点。类似的过程可以应用于其他检测到的活动,以完善各个活动的起点和终点。

[0095] 从图5B和上面的讨论可以看出,本文呈现的手术 workflow 阶段预测和活动识别可以自动注释长手术视频,以便可以出于诸如术后分析、培养新手术医生或用于安全检查的各种目的,将视频进行索引和存档。注释可以通过生成指示阶段和活动的元数据并将该元数据附加到手术视频来执行。可替换或附加地,注释可以通过修改视频的内容以标记检测到的阶段和活动(诸如通过将指示阶段和活动的文本、图像、图标或徽标插入到视频帧中)来执行。

[0096] 应该理解,虽然图5B示出阶段预测和活动识别两者的结果,但是没有必要对给定的手术视频132执行两种分析。换句话说,对于手术视频132,可以执行阶段预测,而无需进行活动识别。类似地,手术视频132的一部分可以用来识别活动122,而无需对该手术视频132的一部分执行阶段预测。

[0097] 现在参考图6,其中,呈现了用于检测手术视频中的手术 workflow 阶段的过程600的示例。将针对图1A中所示的示例系统100A和图1B中所示的示例系统100B来讨论示例过程600,但是示例过程600可以按照根据本公开的任何合适的系统而被采用。过程600包括两个时期(stage):包括操作602和操作604的训练时期和包括过程600的其余部分的预测时期。

[0098] 在操作602中,手术系统100A中的视频分析模块120或手术视频分析系统140选择要用于 workflow 阶段预测中的预测模型210。预测模型210可以包括用于对特定手术过程(诸如胆囊切除术、肾切除术、结肠切除术等)预测 workflow 阶段的阶段预测模型。预测模型210还可以包括用于从手术视频132中分辨出或识别特定活动(诸如缝合、切除、烧灼、切割、冲洗和抽吸等手术任务或出血、胆汁渗漏等重要事件)的活动识别模型。在一些实施方式中,阶段预测模型306可以是卷积神经网络(诸如初始神经网络、Resnet或NASNET)或循环神经网络(诸如LSTM模型、GRU模型或ConvLSTM)。用于活动检测的预测模型可以包括3DCNN和/或带有CNN的循环神经网络。

[0099] 在操作604中,视频分析模块120训练选择的预测模型。训练样本是针对各个预测模型收集和生成的。每个训练样本可以包括一组输入帧(诸如视频块或整个视频中的帧)或者特征和相应的标签。如上所述,训练预测模型可以包括为预测模型找到最小化预测模型的损失函数的一组参数的迭代操作。每次迭代可以涉及为预测模型找到一组参数,以便使用该组参数的损失函数的值小于使用先前迭代中的另一组参数的损失函数的值。损失函数

可以被构建来测量使用预测模型对输入特征预测的输出和包含在训练样本中的相应的标签之间的差。一旦识别到该组参数,预测模型就已经被训练,并且可以按设计用于预测。

[0100] 应当理解,虽然上面描述了视频分析模块120执行训练时期中的操作,但是可以实现手术系统100A或手术视频分析系统140中的其他模块或者手术系统100A或手术视频分析系统140以外的系统,以选择并训练预测模型。例如,可以实现独立于手术系统100A或手术视频分析系统140的一个或多个计算设备(诸如数据中心中的机器集群),以在训练时期期间训练预测模型。然后,可以将经训练的模型部署到手术系统100A,以执行预测时期中的操作。

[0101] 在预测时期的操作606中,过程600涉及获得要对其执行阶段预测的手术视频132。获得手术视频132可以在机器人手术正在进行时实时发生,或者可以在本地从数据存储获得手术视频132或从(例如,通过网络可从云服务器访问的)远程数据存储获得手术视频132。

[0102] 在操作608中,视频分析模块120将手术视频132划分成多个视频块302,每个块包含一个或多个视频帧304。视频分析模块120逐块地处理或预测手术视频132的阶段。在操作610中,视频分析模块120对块302之一预测 workflow 阶段。预测可以使用稍后将详细描述图7A中所示的处理或图7B中所示的处理来执行。在操作610中,如上面参考图3B和图3C讨论的,视频分析模块120对当前块生成阶段预测,并且生成与阶段预测相关联的置信水平。

[0103] 在操作612中,视频分析模块120确定与对当前视频块的阶段预测相关联的置信水平是否高于阈值。如果是,则在操作614中,视频分析模块120使用对当前视频块的预测阶段来注释手术视频132;否则,在操作616中,视频分析模块120使用先前块的阶段预测作为对当前块的预测阶段。在操作618中,视频分析模块120确定是否存在更多的视频块要处理。如果是,则如上所述,视频分析模块120在操作610中处理下一个视频块;否则,过程600在操作620中结束。

[0104] 现在参考图7A,其中,示出了用于对视频块预测 workflow 阶段的示例过程700A。将针对图1A中所示的示例系统100A或图1B中所示的手术视频分析系统140来讨论示例过程700A,但是示例过程700A可以按照根据本公开的任何合适的系统而被采用。在操作702中,视频分析模块120获得视频块302中的视频帧,以供处理。因为连续的帧通常彼此相似,所以处理块302中的所有帧一般效率较低。因此,块302中的帧的子集可以被选择为用于阶段预测。例如,可以为视频的每一秒选择一帧。作为结果,如果块包含5秒的视频,则将为该块选择5帧(诸如来自帧速率为每秒24帧的视频的每24帧中的一帧)。

[0105] 在操作704中,视频分析模块120对每个选择的帧生成阶段预测。例如,可以采用上面参考图3B讨论的方法,以对每个帧生成阶段预测。在操作706中,视频分析模块120对选择的帧聚合阶段预测。上面参考图3B讨论的聚合方法可以用来执行聚合,诸如对针对帧的阶段预测求平均,或者使用其他线性或非线性的函数来组合阶段预测。

[0106] 在操作708中,视频分析模块120对视频块生成聚合阶段预测,以及预测的置信水平。置信值可以基于阶段预测的聚合来计算。例如,如果对视频帧的阶段预测包括包含视频帧属于各个 workflow 阶段的概率的预测向量,则可以生成聚合阶段预测,以将指示视频块属于各个 workflow 阶段的聚合概率的预测向量也包括在内。在此情况下,具有最高聚合概率的 workflow 阶段可以被识别为对视频块的阶段预测,并且最高聚合概率可以用来指示该预测的

置信水平。

[0107] 图7B示出用于对视频块预测 workflow 阶段的另一示例过程700B。类似于过程700A的操作702,在操作722中,视频分析模块120获得视频块中的视频帧,以供处理。在操作724中,视频分析模块120为每个选择的视频帧生成特征向量。特征向量可以使用上面参考图3C讨论的方法(诸如经过卷积神经网络)来生成。然后,在操作726中,这些特征向量可以通过例如对特征向量求平均、级联或应用其他函数来聚合。

[0108] 在操作728中,视频分析模块120基于聚合特征向量,对视频块生成阶段预测。可以使用被训练为将聚合特征向量作为输入并输出对当前块的阶段预测的阶段预测模型来生成阶段预测。这里使用的阶段预测模型可以是循环神经网络,诸如LSTM模型或GRU模型。除了阶段预测之外,阶段预测模型还可以被训练并用来生成与预测相关联的置信水平。可替换或附加地,预测本身可以包括指示当前块属于各个阶段的概率的概率向量。预测阶段可以被确定为与最高概率相关联的阶段,并且这种预测的置信水平可以被确定为最高概率本身。

[0109] 图8示出用于检测手术视频中的手术活动的过程800的示例。将针对图1A中所示的示例系统100A或图1B中所示的手术视频分析系统140来讨论示例过程800,但是示例过程800可以按照根据本公开的任何合适的系统而被采用。在操作802中,视频分析模块120从手术视频132获得手术视频帧。在一些实施方式中,视频帧从手术视频132的视频块302获得。在其他实施方式中,这里使用的手术视频132不需要被划分成视频块,并且如果手术视频132已经被划分成块,则在操作802中获得的视频帧不一定是从相同的视频块获得的。换句话说,在操作802中获得的视频帧可以独立于手术视频132的视频块划分。

[0110] 在操作804中,视频分析模块120使用上面参考图4讨论的特征提取模型从获得的视频帧生成第一特征向量。特征提取模型可以是例如被训练为基于一组输入视频帧生成特征向量的3DCNN。在操作806中,视频分析模块120从视频帧中提取或估计光流。光流可以使用本领域公知的任何光流方法(诸如基于阶段相关性的方法、基于块的方法、差分方法或深度学习方法来提取。

[0111] 在操作808中,视频分析模块120随后可以基于估计的光流使用光流特征提取模型生成第二特征向量。类似于在操作804中使用的特征提取模型,光流特征提取模型404还可以包括被训练为基于光流输入生成特征向量的3DCNN或其他合适的模型。在操作810中,视频分析模块120组合第一特征向量和第二特征向量。两个特征向量的组合可以包括但不限于对两个特征向量求平均、级联两个特征向量或为每个向量元素选择两个值中的较大值。在操作812中,视频分析模块120使用组合特征向量确定对视频帧的活动预测。该预测可以通过使用活动识别模型(诸如被训练为基于从视频帧和光流生成的特征向量来识别手术活动的ML模型)做出。如上所述,识别到的活动可以用来注释视频以达到索引目的,或被传输给手术医生以在手术过程期间提供反馈。

[0112] 现在参考图9,图9示出适合于在用于经过视频处理来检测手术 workflow 阶段和活动的示例系统或方法中使用的示例计算设备900。示例计算设备900包括使用一个或多个通信总线902与存储器920和计算设备900的其他组件通信的处理器910。处理器910被配置为执行存储在存储器920中的处理器可执行指令,以根据不同的示例(诸如上面参考图6-8描述的示例过程600、示例过程700A、示例过程700B和示例过程800中的一部分或全部)来执行手

术 workflow 阶段和活动检测。在该示例中,计算设备还包括一个或多个用户输入设备 970 (诸如键盘、鼠标、触摸屏、麦克风等),以接受用户输入。计算设备 900 还包括显示器 960,以向用户提供视觉输出。

[0113] 计算设备 900 可以包括或连接到为计算设备 900 提供非易失性存储的一个或多个存储设备 930。存储设备 930 可以存储由计算设备 900 利用的系统或应用程序和数据,诸如实现由视频分析模块 120 提供的功能的模块。存储设备 930 还可以存储在本文中未具体识别的其他程序和数据。

[0114] 计算设备 900 还包括通信接口 940。在一些示例中,通信接口 940 可以使用一个或多个网络 (包括局域网 (“LAN”); 广域网 (“WAN”), 诸如互联网; 城域网 (“MAN”); 点对点连接或对等连接等) 来实现通信。可以使用任何合适的网络协议来实现与其他设备的通信。例如,一种合适的网络协议可以包括互联网协议 (“IP”)、传输控制协议 (“TCP”)、用户数据报协议 (“UDP”) 或其组合,诸如 TCP/IP 或 UDP/IP。

[0115] 虽然本文的方法和系统的一些示例是根据在各种机器上执行的软件来描述的,但是这些方法和系统也可以被实现为特定配置的硬件,诸如特定于执行各种方法的现场可编程门阵列 (FPGA)。例如,示例可以在数字电子电路中实现,或者在计算机硬件、固件、软件中或在其组合中实现。在一个示例中,设备可以包括一个或多个处理器。处理器包括计算机可读介质,诸如耦接到处理器的随机存取存储器 (RAM)。处理器执行存储在存储器中的计算机可执行程序指令,诸如执行一个或多个计算机程序。这种处理器可以包括微处理器、数字信号处理器 (DSP)、专用集成电路 (ASIC)、现场可编程门阵列 (FPGA) 和状态机。这种处理器还可以包括可编程电子设备,诸如 PLC、可编程中断控制器 (PIC)、可编程逻辑设备 (PLD)、可编程只读存储器 (PROM)、电子可编程只读存储器 (EPROM 或 EEPROM) 或其他类似设备。

[0116] 这种处理器可以包括介质或可以与介质通信,该介质是例如可以存储指令的非暂时性计算机可读存储介质,该指令在由处理器执行时,可以使处理器执行由处理器执行或协助的本文描述的步骤。非暂时性计算机可读介质的示例可以包括但不限于电子存储设备、光学存储设备、磁性存储设备或能够向处理器 (诸如网络服务器中的处理器) 提供计算机可读指令的其他存储设备。介质的其他示例包括但不限于软盘、CD-ROM、磁盘、存储芯片、ROM、RAM、ASIC、经配置的处理器、所有光学介质、所有磁带或其他磁性介质,或计算机处理器可以从其读取的任何其他介质。所描述的处理器和处理可以在一个或多个结构中,并且可以通过一个或多个结构分散。处理器可以包括用于执行本文中描述的一个或多个方法 (或方法的部分) 的代码。

[0117] 某些示例的前述描述仅为了说明和描述的目的而呈现,并不旨在作为详尽的或将公开内容限制为所公开的精确形式。在不背离本公开的精神和范围的情况下,对本领域技术人员而言,其多种修改和适配将是明显的。

[0118] 本文对示例或实施方式的参考意味着结合示例描述的特定特征、结构、操作或其他特性可以被包括在本公开的至少一个实施方式中。本公开不限于这样描述的特定示例或实现。短语“在一个示例中”、“在示例中”、“在一个实施方式中”或“在实施方式中”或其在本说明书中不同地方的变体的出现不一定指相同的示例或实现。本说明书中描述的与一个示例或实施方式相关的任何特定特征、结构、操作或其他特征可以与参考任何其他示例或实施方式描述的其他特征、结构、操作或其他特征组合。

[0119] 本文中“或”一词的使用旨在涵盖包含和异或条件。换句话说,A或B或C包括适用于特定用途的任何或所有以下替代组合:只有A;只有B;只有C;仅A和B;仅A和C;仅B和C;以及A和B和C。

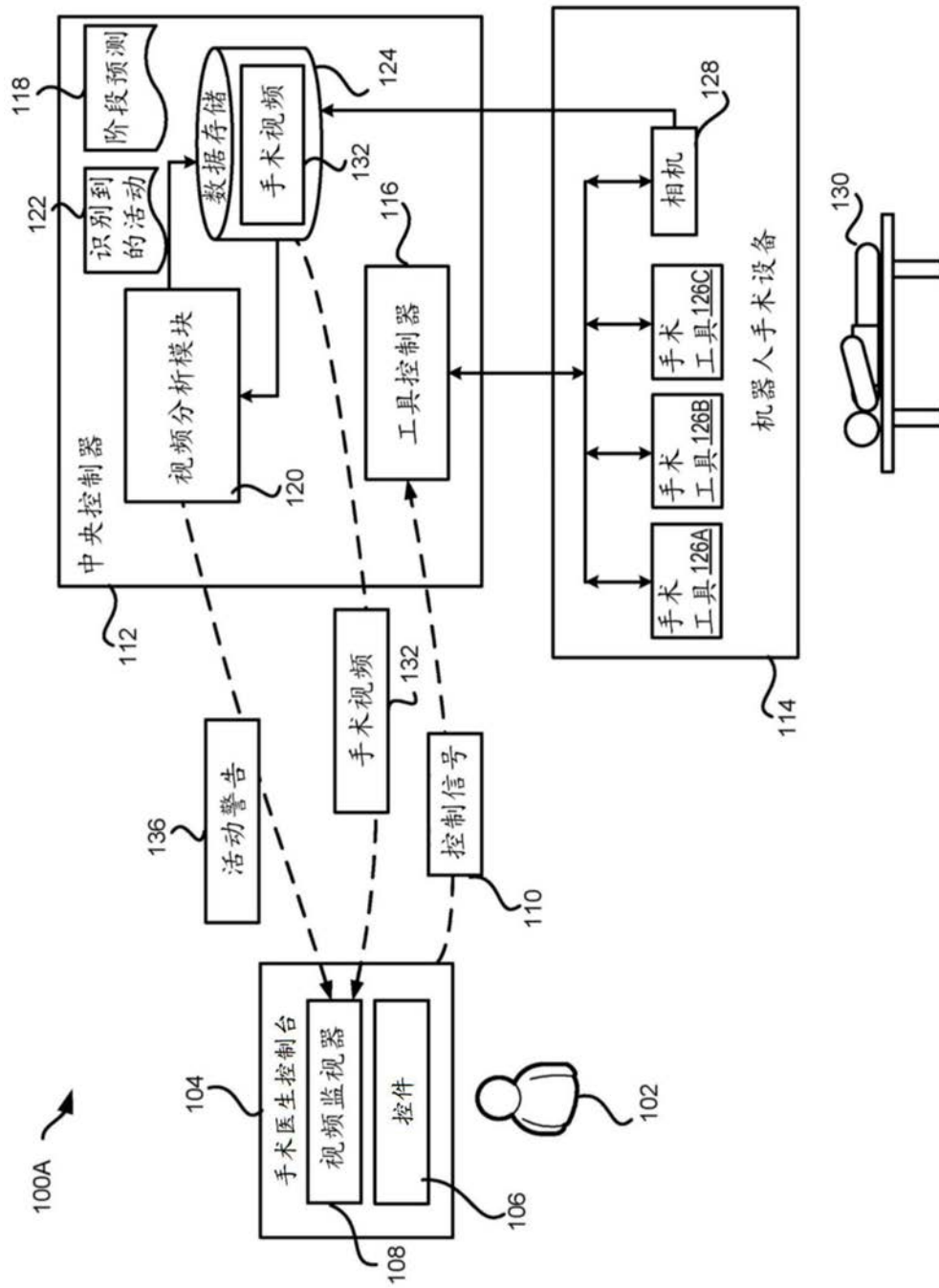


图1A

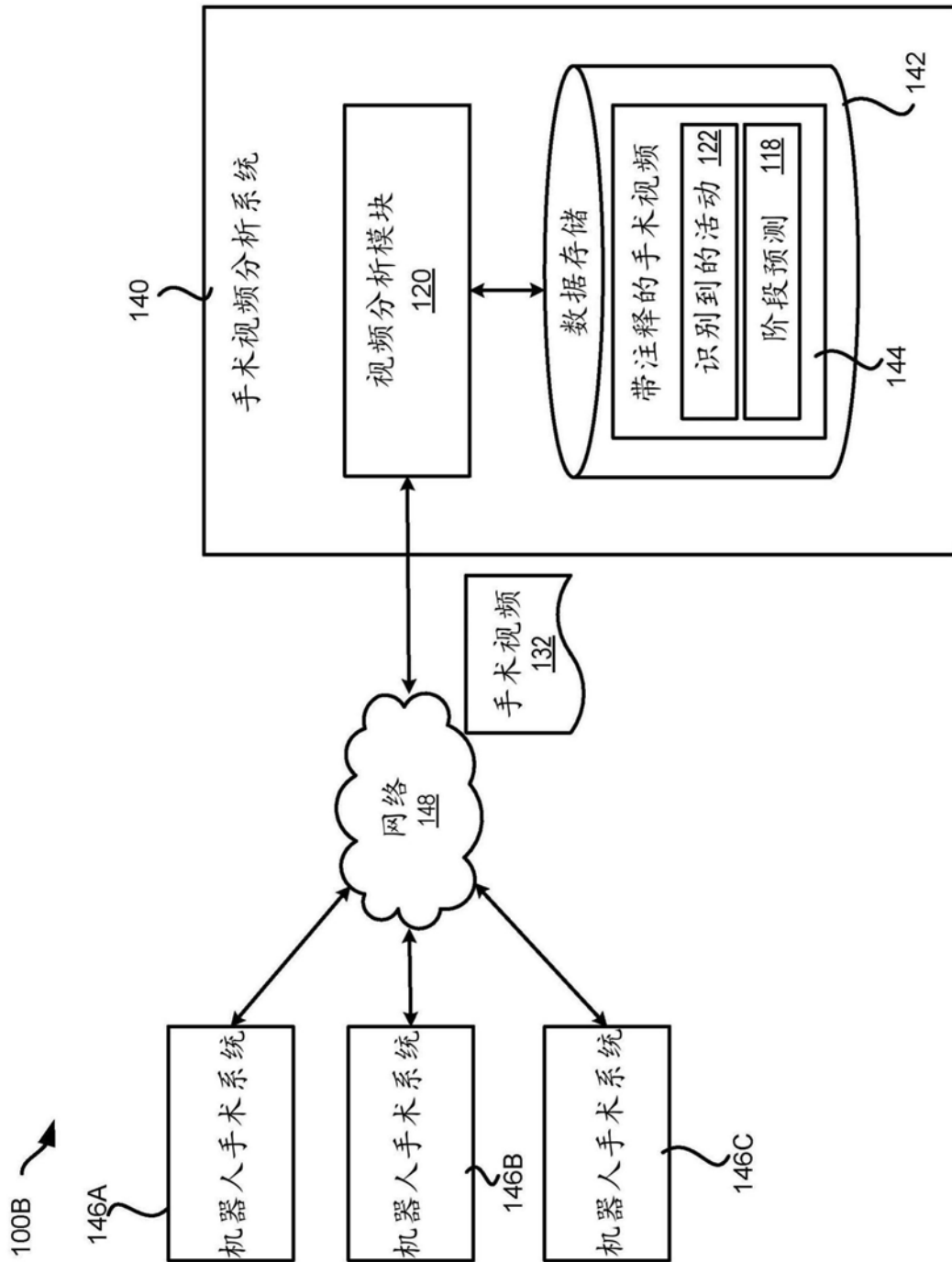


图1B

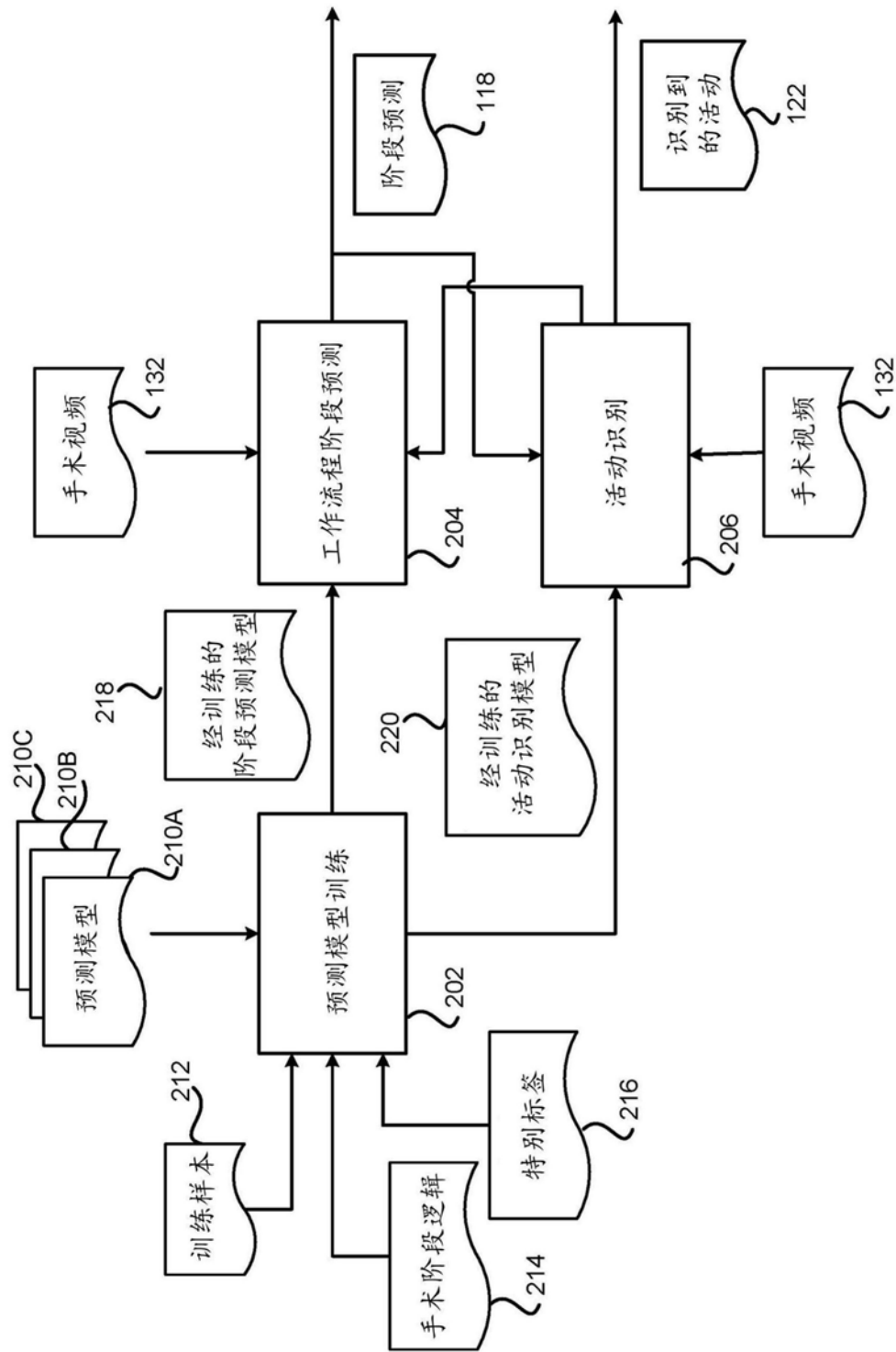


图2

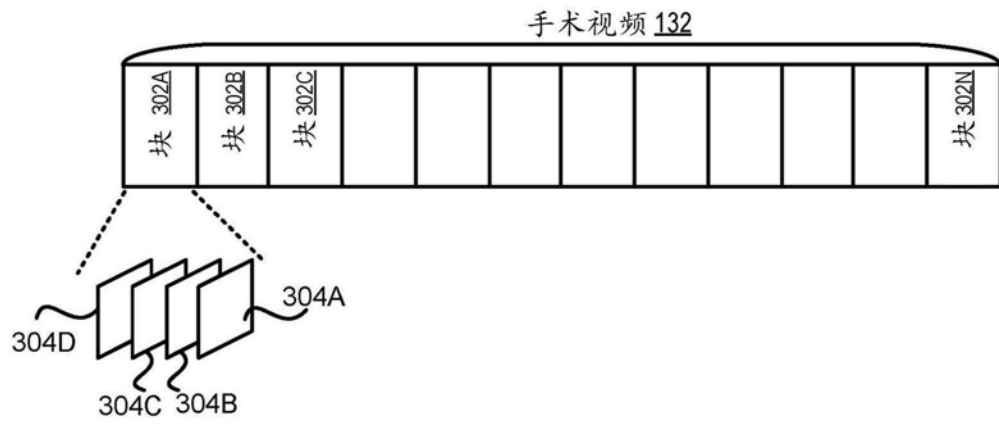


图3A

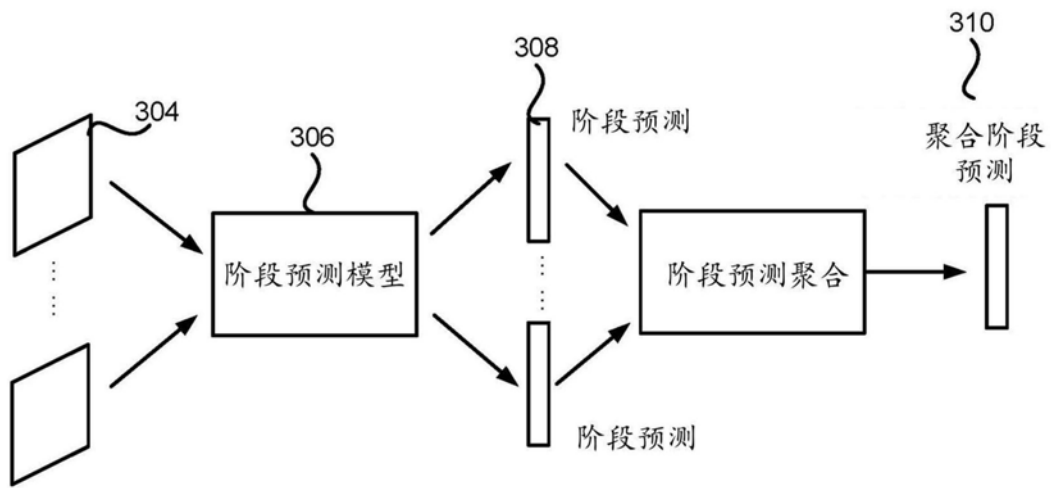


图3B

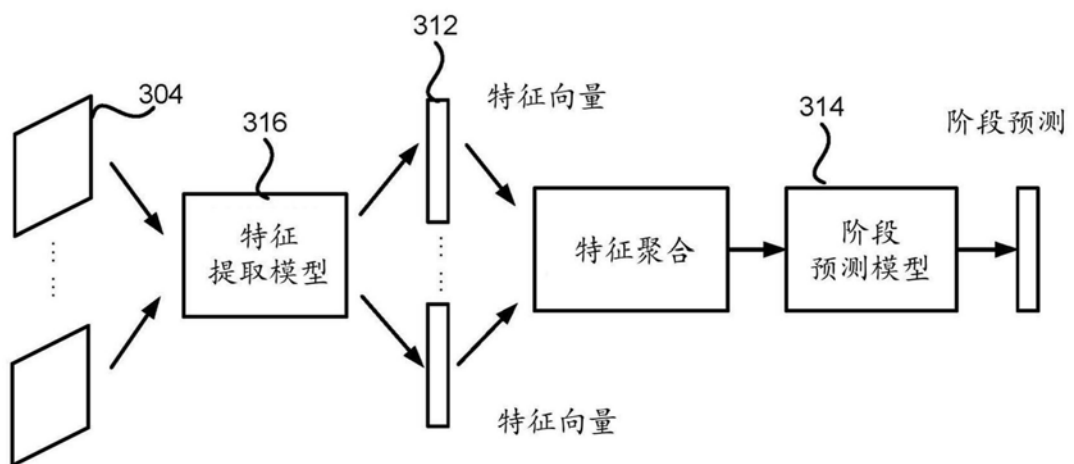


图3C

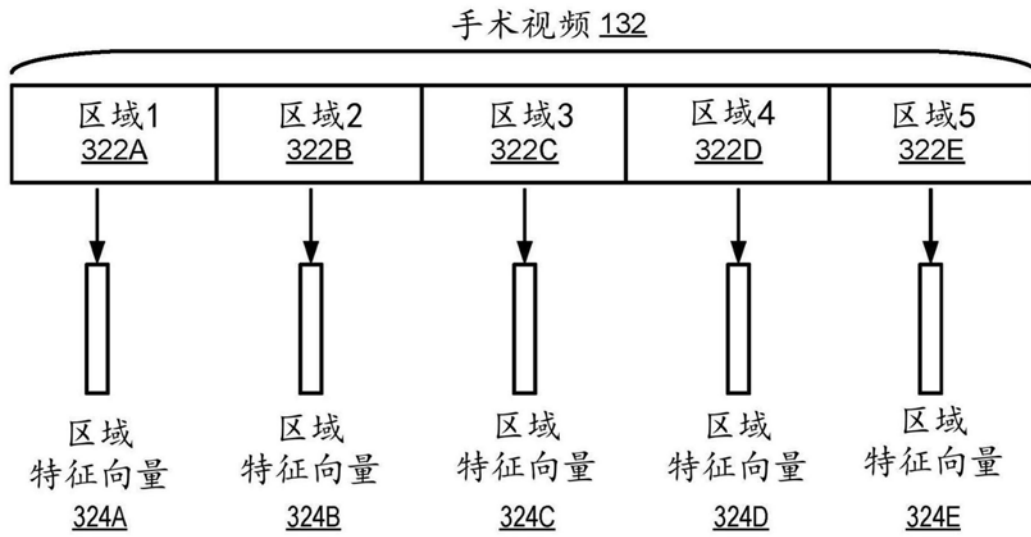


图3D

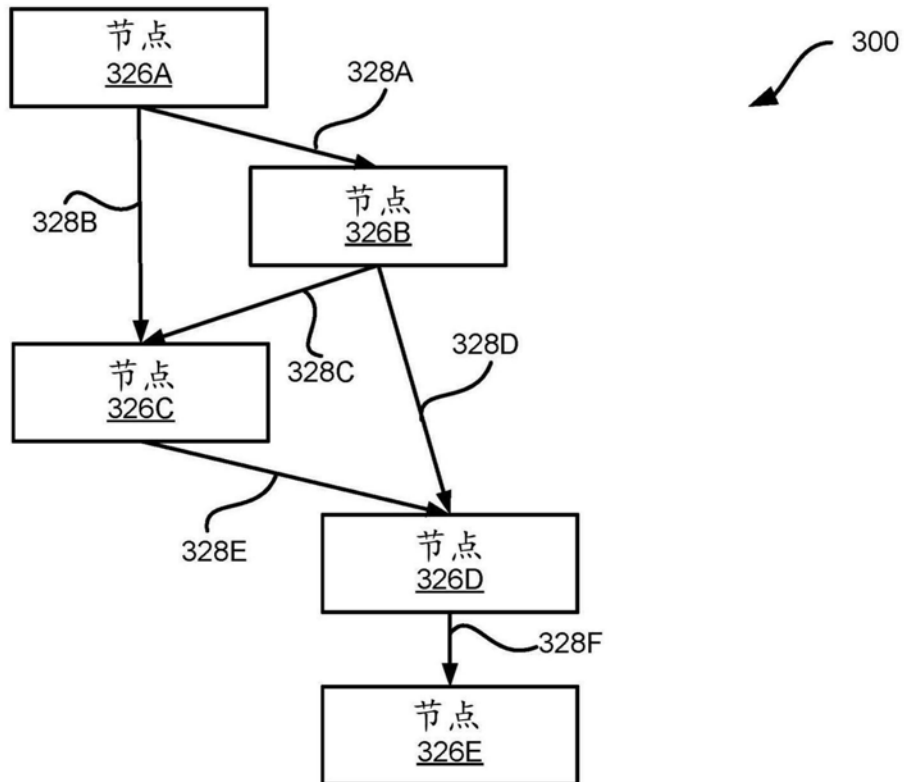


图3E

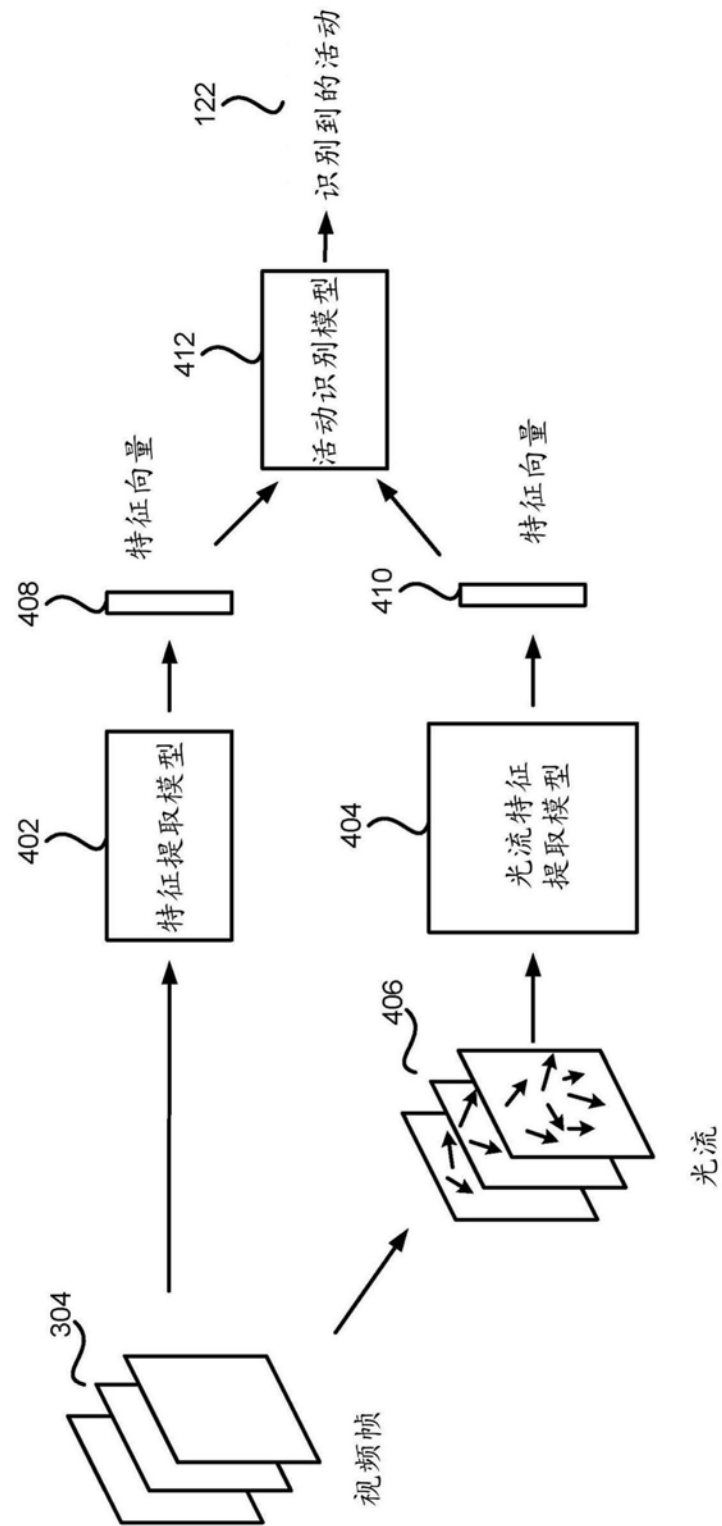


图4

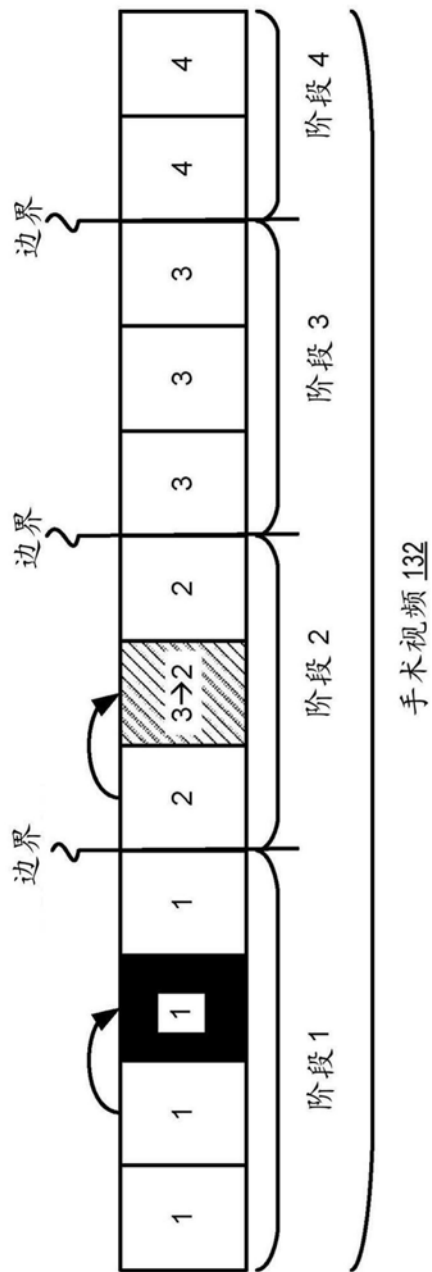


图5A

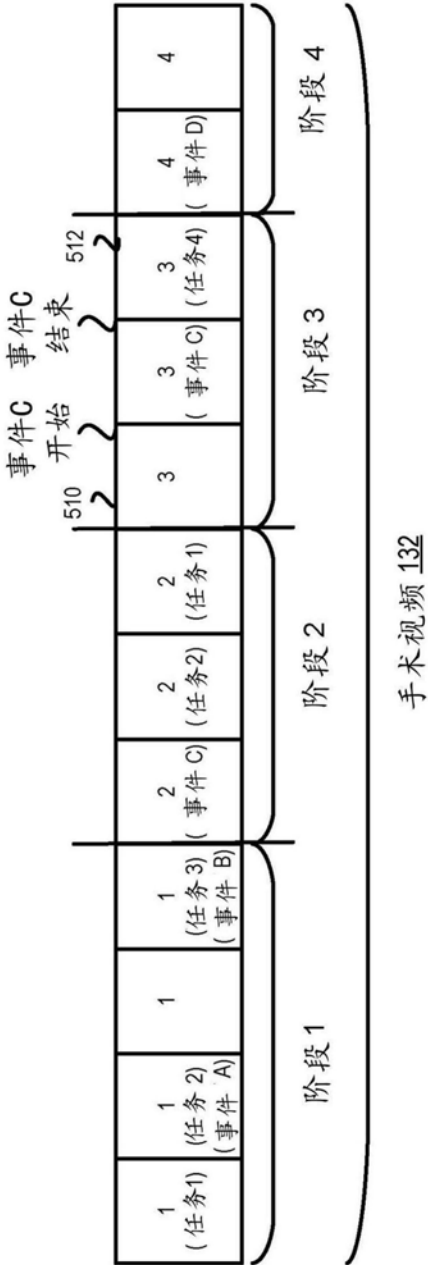


图5B

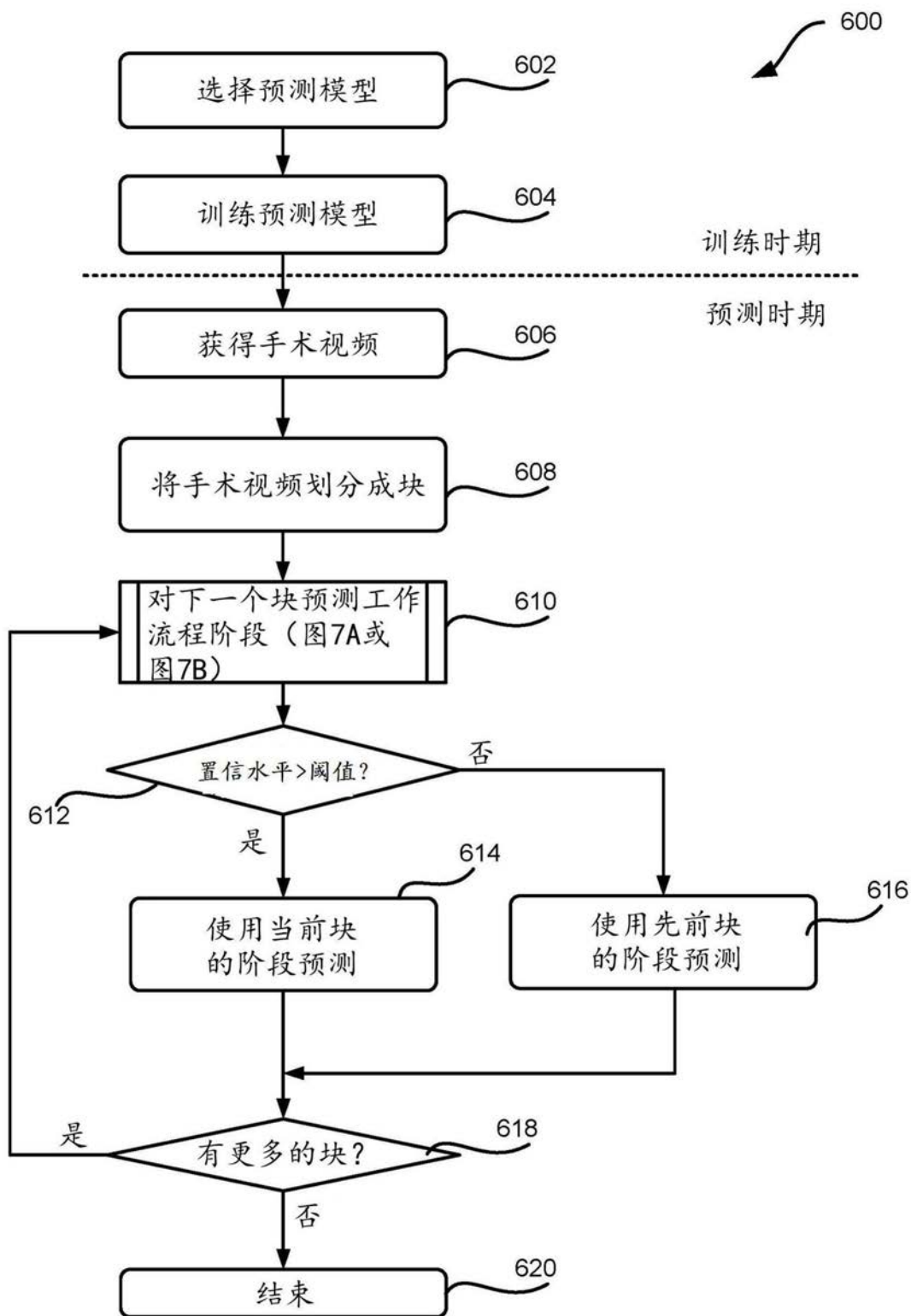


图6

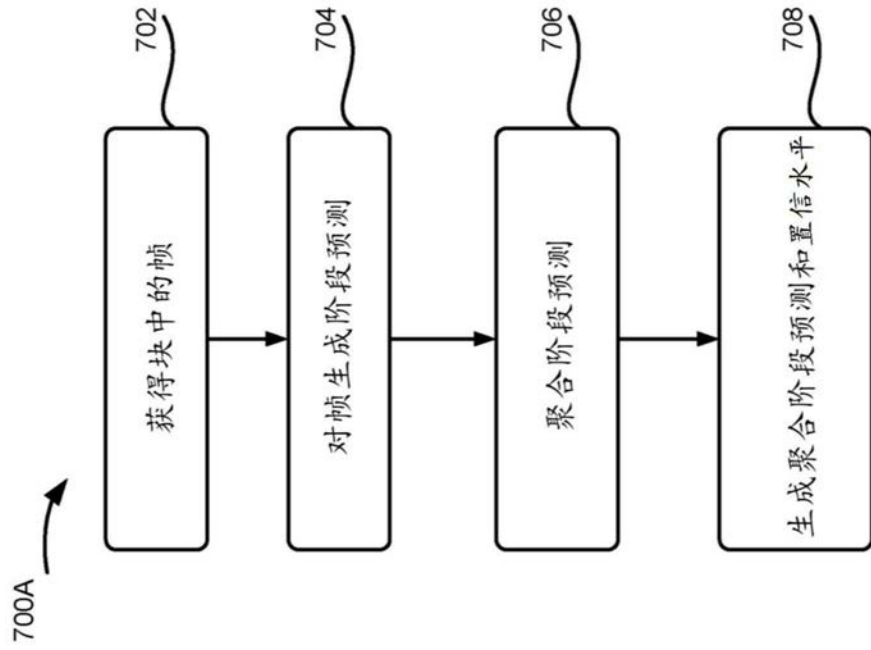


图7A

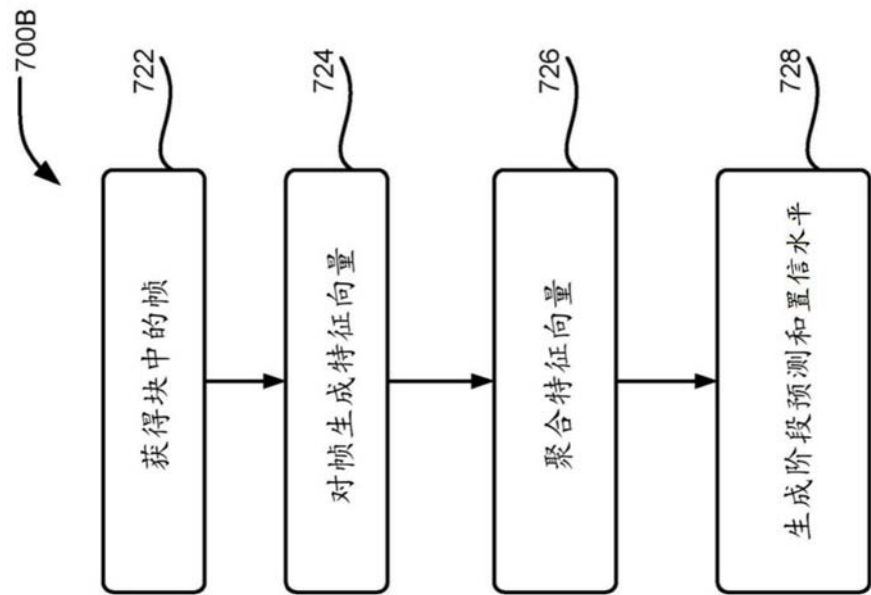


图7B

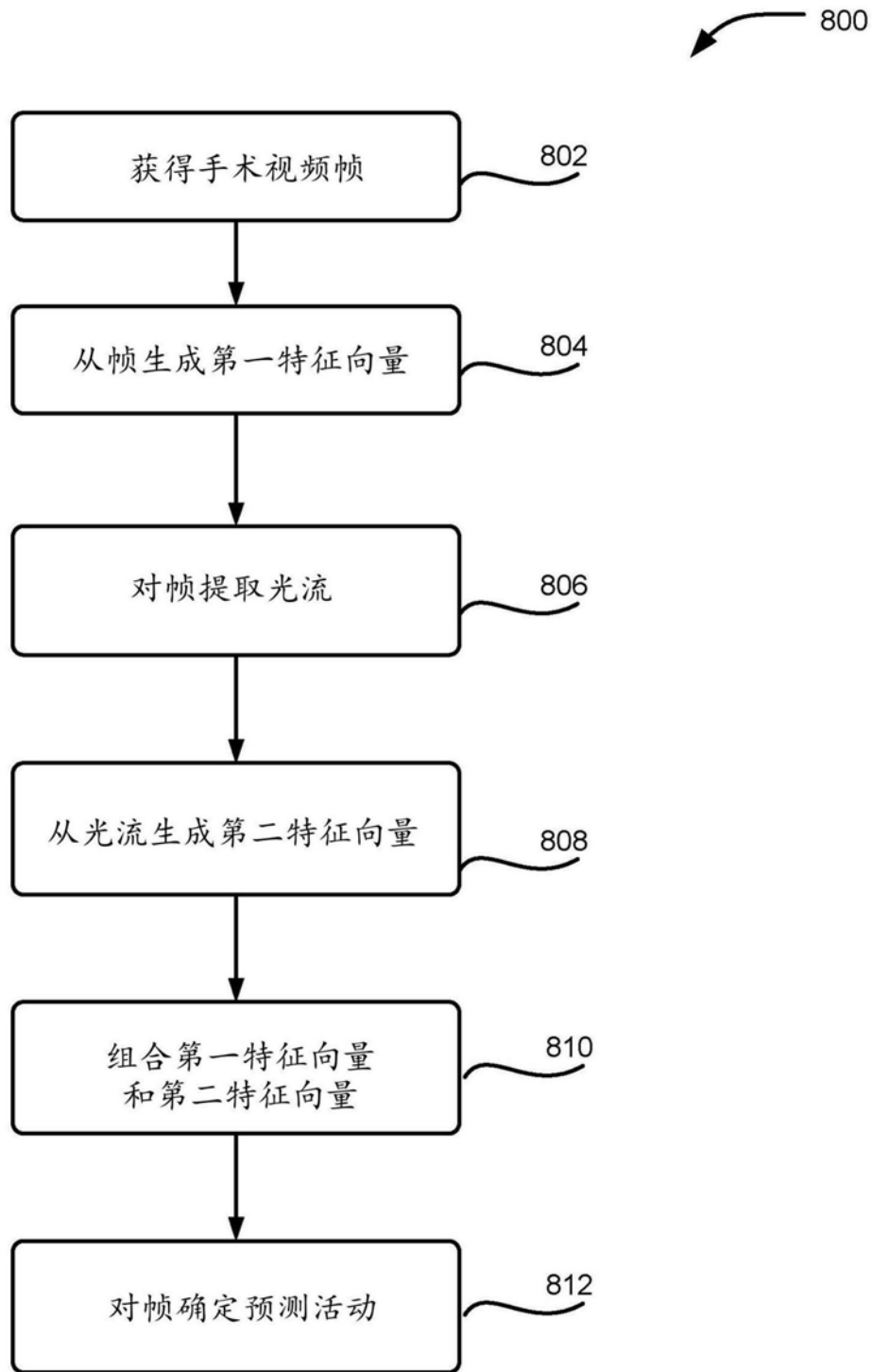


图8

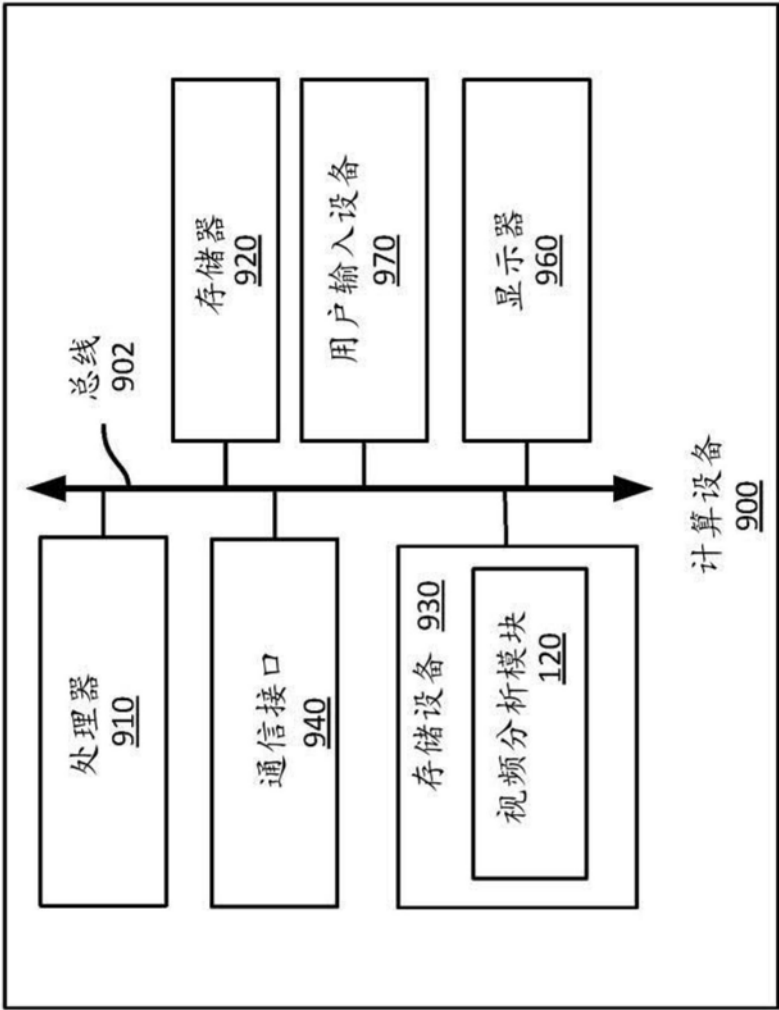


图9