

(12) 按照专利合作条约所公布的国际申请

(19) 世界知识产权组织
国际局

(43) 国际公布日
2021 年 6 月 17 日 (17.06.2021)



(10) 国际公布号
WO 2021/114892 A1

(51) 国际专利分类号:
G06K 9/00 (2006.01)

(21) 国际申请号: PCT/CN2020/123214

(22) 国际申请日: 2020 年 10 月 23 日 (23.10.2020)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
202010475795.7 2020年5月29日 (29.05.2020) CN

(71) 申请人: 平安科技 (深圳) 有限公司 (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) [CN/CN];
中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼时梦如, Guangdong 518000 (CN)。

(72) 发明人: 冯颖龙 (FENG, Yinglong); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼时梦如, Guangdong 518000 (CN)。 付佐毅 (FU, Zuoyi); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼时梦如, Guangdong 518000 (CN)。 周宸 (ZHOU, Chen); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼时梦如, Guangdong 518000 (CN)。 周宝 (ZHOU, Bao); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼时梦如, Guangdong 518000 (CN)。 陈远旭 (CHEN, Yuanxu); 中国广东省深圳市福田区福田街道福安社区益田路 5033 号平安金融中心 23 楼时梦如, Guangdong 518000 (CN)。

(54) Title: ENVIRONMENTAL SEMANTIC UNDERSTANDING-BASED BODY MOVEMENT RECOGNITION METHOD, APPARATUS, DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 基于环境语义理解的人体行为识别方法、装置、设备及存储介质

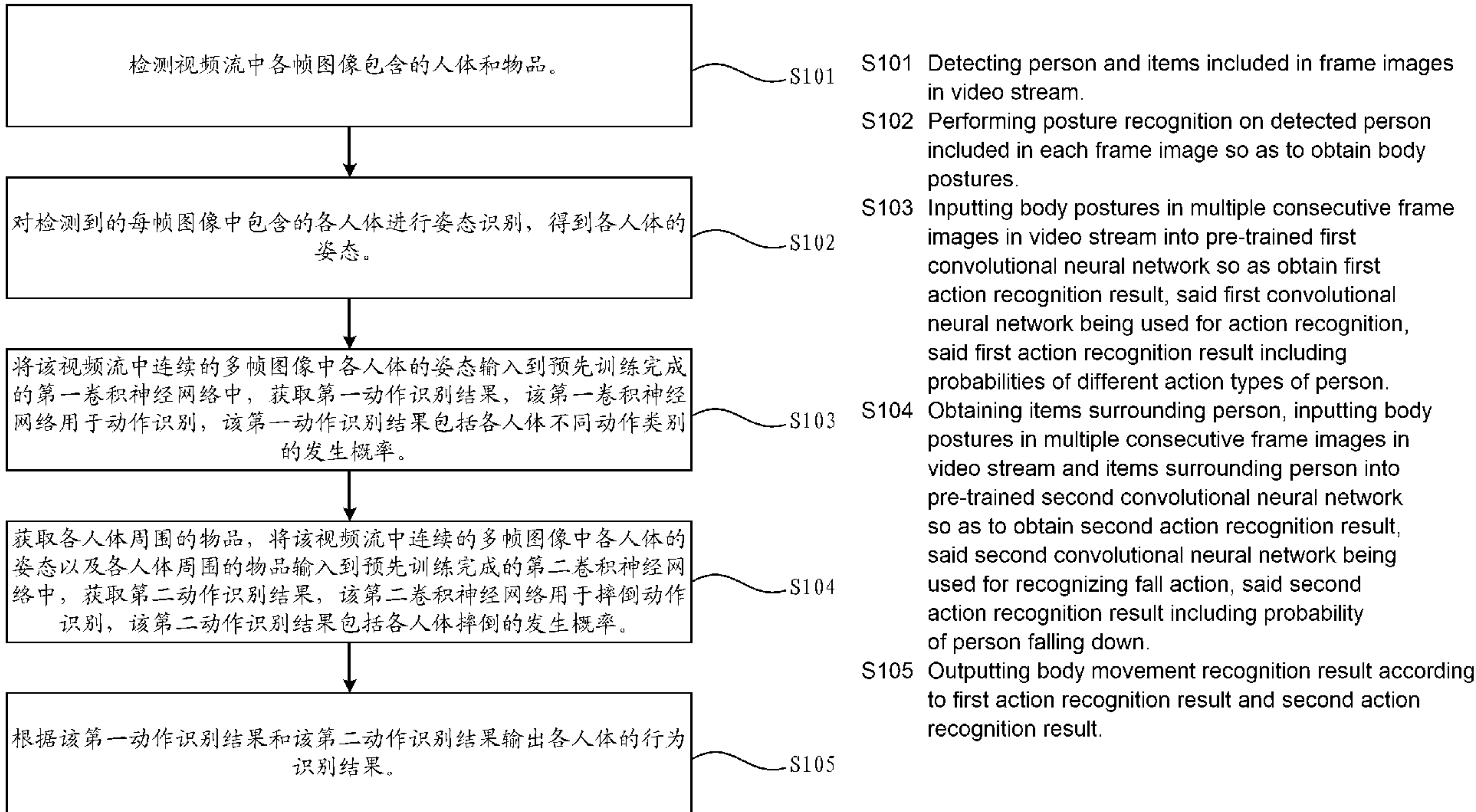


图 1

(57) Abstract: The present application relates to the technical field of video image processing and artificial intelligence, and specifically relates to an environmental semantic understanding-based body movement recognition method, an apparatus, a device, and a storage medium. The method comprises: detecting a person and items included in the frame images in a video stream; performing posture recognition on the detected person included in each frame image so as to obtain the body postures; inputting the body postures into a first

(74) 代理人：深圳国新南方知识产权代理有限公司 (SHENZHEN CHINA INNOVATION SOUTH INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国广东省深圳市南山区深圳湾科技生态园 11 栋 A 座 501C/ 姜宇, Guangdong 518000 (CN)。

(81) 指定国 (除另有指明, 要求每一种可提供的国家保护): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, IT, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW。

(84) 指定国 (除另有指明, 要求每一种可提供的地区保护): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), 欧亚 (AM, AZ, BY, KG, KZ, RU, TJ, TM), 欧洲 (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG)。

本国际公布:

- 包括国际检索报告 (条约第21条 (3))。
- 在修改权利要求的期限届满之前进行, 在收到该修改后将重新公布 (细则48.2 (h))。
- 根据申请人的请求, 在条约第21条 (2) (a) 所规定的期限届满之前进行。

convolutional neural network to obtain the probabilities of different action types of the person; inputting the body postures and the items surrounding the person into a second convolutional neural network so as to obtain the probability of person falling down; outputting the movement recognition result. The present method prevents mis-recognizing an item as a person during posture recognition, and enhances the accuracy and time relevancy of body posture recognition. The second convolutional neural network uses the body postures and surrounding items to recognize a fall, thereby enhancing the accuracy of action detection, and providing good robustness to the unstable process of body posture recognition.

(57) 摘要: 本申请涉及视频图像处理及人工智能技术领域, 尤其涉及一种基于环境语义理解的人体行为识别方法、装置、设备及存储介质。该方法包括: 检测视频流中各帧图像包含的人体和物品; 对检测到的每帧图像中包含的各人体进行姿态识别, 得到各人体的姿态; 将人体的姿态输入第一卷积神经网络获取各人体不同动作类别的发生概率; 将人体的姿态和人体周围的物品输入第二卷积神经网络获取各人体摔倒的发生概率; 输出行为识别结果。通过上述方式, 避免了姿态识别过程中将物品作为人体进行误识别, 提高了人体姿态识别的准确性和实时性; 第二卷积神经网络利用人体的姿态和周围的物品进行摔倒识别, 提高了检测动作的准确度, 对于不稳定的人体的姿态识别具有良好的鲁棒性。

基于环境语义理解的人体行为识别方法、装置、设备及存储介质

本申请要求于 2020 年 05 月 29 日提交中国专利局、申请号为 202010475795.7，发明名称为“基于环境语义理解的人体行为识别方法、装置及存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

【技术领域】

本申请涉及视频图像处理技术领域，还涉及人工智能领域，尤其涉及一种基于环境语义理解的人体行为识别方法、装置、设备及存储介质。

【背景技术】

现有技术中人体姿态识别的主流方案是采用自上而下和自下而上的算法，发明人意识到，采用自下而上的算法识别人体姿态时，会有很大的概率产生误识别，例如，误将椅子或仓库摆放的机器人等物品作为人体并从中预测出人体的姿态，误识别会严重影响到算法的识别准确度以及使用场景，并且模型的不稳定性会给算法的应用增加较大的不确定性；同时，自下而上的算法增加了计算的时间复杂度和空间复杂度；另外，自上而下的算法对多人存在的复杂场景中估计位姿的准确率以及速度都较低。

估计出人体位姿后，需要根据人体位姿进行动作分类以识别人体行为，现有技术中大多采用基于端到端的算法模型进行动作分类，该算法模型对输入的人体位姿的准确度要求很高，也对标注的数据的质量要求很高，从而使得端到端的动作识别容易产生较大的偏差，识别的准确率较低。

因此，有必要提供一种新的人体行为识别方法以解决上述技术问题。

【发明内容】

本申请的目的在于提供一种基于环境语义理解的人体行为识别方法、装置、设备及存储介质，能够解决现有技术中人体姿态识别的准确性低以及检测动作的准确度低的问题。

为解决上述技术问题，本申请采用的一个技术方案是：提供一种基于环境语义理解的人体行为识别方法，包括：

检测视频流中各帧图像包含的人体和物品；

对检测到的每帧图像中包含的各人体进行姿态识别，得到各人体的姿态；

将所述视频流中连续的多帧图像中各人体的姿态输入到预先训练完成的第一卷积神经网络中，获取第一动作识别结果，所述第一卷积神经网络用于动作识别，所述第一动作识别结果包括各人体不同动作类别的发生概率；

获取各人体周围的物品，将所述视频流中连续的多帧图像中各人体的姿态以及各人体周围的物品输入到预先训练完成的第二卷积神经网络中，获取第二动作识别结果，所述人体周

围的物品为每帧图像中与所述人体的距离小于或等于预设阈值的物品，所述第二卷积神经网络用于摔倒动作识别，所述第二动作识别结果包括各人体摔倒的发生概率；

根据所述第一动作识别结果和所述第二动作识别结果输出各人体的行为识别结果。

为解决上述技术问题，本申请采用的另一个技术方案是：提供一种基于环境语义理解的人体行为识别装置，包括：

目标检测模块，用于检测视频流中各帧图像包含的人体和物品；

姿态识别模块，用于对检测到的每帧图像中包含的各人体进行姿态识别，得到各人体的姿态；

通用动作分类模块，用于将所述视频流中连续的多帧图像中各人体的姿态输入到预先训练完成的第一卷积神经网络中，获取第一动作识别结果，所述第一卷积神经网络用于动作识别，所述第一动作识别结果包括各人体不同动作类别的发生概率；

摔倒动作识别模块，用于获取各人体周围的物品，将所述视频流中连续的多帧图像中各人体的姿态以及各人体周围的物品输入到预先训练完成的第二卷积神经网络中，获取第二动作识别结果，所述人体周围的物品为每帧图像中与所述人体的距离小于或等于预设阈值的物品，所述第二卷积神经网络用于摔倒动作识别，所述第二动作识别结果包括各人体摔倒的发生概率；

输出模块，用于根据所述第一动作识别结果和所述第二动作识别结果输出各人体的行为识别结果。

为解决上述技术问题，本申请采用的另一个技术方案是：提供一种电子设备，包括处理器、以及与所述处理器耦接的存储器，所述存储器存储有可被所述处理器执行的程序指令；所述处理器执行所述存储器存储的所述程序指令时实现以下步骤：

检测视频流中各帧图像包含的人体和物品；

对检测到的每帧图像中包含的各人体进行姿态识别，得到各人体的姿态；

将所述视频流中连续的多帧图像中各人体的姿态输入到预先训练完成的第一卷积神经网络中，获取第一动作识别结果，所述第一卷积神经网络用于动作识别，所述第一动作识别结果包括各人体不同动作类别的发生概率；

获取各人体周围的物品，将所述视频流中连续的多帧图像中各人体的姿态以及各人体周围的物品输入到预先训练完成的第二卷积神经网络中，获取第二动作识别结果，所述人体周围的物品为每帧图像中与所述人体的距离小于或等于预设阈值的物品，所述第二卷积神经网络用于摔倒动作识别，所述第二动作识别结果包括各人体摔倒的发生概率；

根据所述第一动作识别结果和所述第二动作识别结果输出各人体的行为识别结果。

为解决上述技术问题，本申请采用的另一个技术方案是：提供一种存储介质，所述存储介质内存储有程序指令，所述程序指令被处理器执行时实现以下步骤：

检测视频流中各帧图像包含的人体和物品;

对检测到的每帧图像中包含的各人体进行姿态识别,得到各人体的姿态;

将所述视频流中连续的多帧图像中各人体的姿态输入到预先训练完成的第一卷积神经网络中,获取第一动作识别结果,所述第一卷积神经网络用于动作识别,所述第一动作识别结果包括各人体不同动作类别的发生概率;

获取各人体周围的物品,将所述视频流中连续的多帧图像中各人体的姿态以及各人体周围的物品输入到预先训练完成的第二卷积神经网络中,获取第二动作识别结果,所述人体周围的物品为每帧图像中与所述人体的距离小于或等于预设阈值的物品,所述第二卷积神经网络用于摔倒动作识别,所述第二动作识别结果包括各人体摔倒的发生概率;

根据所述第一动作识别结果和所述第二动作识别结果输出各人体的行为识别结果。

本申请的有益效果在于:本申请的基于环境语义理解的人体行为识别方法、装置及存储介质,先检测视频流中各帧图像包含的人体和物品,再对检测到的每帧图像中包含的各人体进行姿态识别,得到各人体的姿态;将人体的姿态输入第一卷积神经网络获取各人体不同动作类别的发生概率,将人体的姿态和人体周围的物品输入第二卷积神经网络获取各人体摔倒的发生概率,再根据人体不同动作类别的发生概率和人体摔倒的发生概率输出行为识别结果,通过上述方式,避免了姿态识别过程中将物品作为人体进行误识别,提高了人体姿态识别的准确性和实时性;第一卷积神经网络进行通用动作识别,第二卷积神经网络利用人体的姿态和周围的物品进行摔倒识别,提高了检测动作的准确度,对于不稳定的人体的姿态识别具有良好的鲁棒性。

【附图说明】

图1为本申请第一实施例的基于环境语义理解的人体行为识别方法的流程图;

图2为本申请第二实施例的基于环境语义理解的人体行为识别方法的流程图;

图3为本申请第三实施例的基于环境语义理解的人体行为识别装置的结构示意图;

图4为本申请第四实施例的基于环境语义理解的电子设备的结构示意图;

图5为本申请实施例的存储介质的结构示意图。

【具体实施方式】

下面将结合本申请实施例中的附图,对本申请实施例中的技术方案进行清楚、完整地描述,显然,所描述的实施例仅是本申请的一部分实施例,而不是全部的实施例。基于本申请中的实施例,本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例,都属于本申请保护的范围。

本申请中的术语“第一”、“第二”、“第三”仅用于描述目的,而不能理解为指示或暗示相对重要性或者隐含指明所指示的技术特征的数量。由此,限定有“第一”、“第二”、“第三”的特征可以明示或者隐含地包括至少一个该特征。本申请的描述中,“多个”的含义是至少

两个，例如两个，三个等，除非另有明确具体的限定。本申请实施例中所有方向性指示(诸如上、下、左、右、前、后……)仅用于解释在某一特定姿态(如附图所示)下各部件之间的相对位置关系、运动情况等，如果该特定姿态发生改变时，则该方向性指示也相应地随之改变。此外，术语“包括”和“具有”以及它们任何变形，意图在于覆盖不排他的包含。例如包含了一系列步骤或单元的过程、方法、系统、产品或设备没有限定于已列出的步骤或单元，而是可选地还包括没有列出的步骤或单元，或可选地还包括对于这些过程、方法、产品或设备固有的其它步骤或单元。

在本文中提及“实施例”意味着，结合实施例描述的特定特征、结构或特性可以包含在本申请的至少一个实施例中。在说明书中的各个位置出现该短语并不一定均是指相同的实施例，也不是与其它实施例互斥的独立的或备选的实施例。本领域技术人员显式地和隐式地理解的是，本文所描述的实施例可以与其它实施例相结合。

图1是本申请第一实施例的基于环境语义理解的人体行为识别方法的流程示意图。需要注意的是，若有实质上相同的结果，本申请的方法并不以图1所示的流程顺序为限。如图1所示，该基于环境语义理解的人体行为识别方法包括步骤：

S101，检测视频流中各帧图像包含的人体和物品。

在步骤S101中，视频流包括机器人拍摄的连续多个视频帧，或者视频流包括机器人拍摄的连续多个视频帧中的任意几个视频帧。

在步骤S101中，基于环境语义信息理解，检测出环境中的人体和物品，将视频流输入到预先训练完成的深度学习网络中，获取该视频流中各帧图像包含的人体和物品，该深度学习网络用于目标预测，该目标包括人体和物品，该端到端的深度学习网络包括多层卷积神经网络、多层最大池化层以及全连接层，例如为23层卷积神经网络、5层最大池化层以及最后采用全连接层进行分类和回归，具体地，按照预设的划分方式，将该视频流中各帧图像划分为多个网格；在每个网格中，通过预先设置的不同类型的检测框进行目标预测，针对每个检测框，获取该检测框预测的目标的坐标参数(x,y)、检测框的宽度和高度(w,h)以及检测框的置信度(Ptr)，将置信度最高的检测框作为检测结果，该预测结果包括目标、检测框、该目标的坐标参数以及该目标的类别，该检测框为框选出所述目标的外接区域，该目标的类别包括人体和物品；根据所述预测结果确定所述视频流中各帧图像包含的人体和物品。

其中，可以将每帧图像分成 $s \times s$ 个网格，在每个网格中，按照不同类型的检测框进行目标预测，实现目标的位置和类别的预测，对于每个网格共有n个类型的检测框，目标预测的类别共有m种，包括人体、床、桌子、椅子、机器人、瑜伽垫等m种。对于不同类型的检测框，检测结果包括坐标参数(x,y)、宽度和高度(w,h)以及置信度(Ptr)，共有5个参数，则参数数量为 $(s \times s \times n \times (m + 5))$ 。

为了预测图像中目标的类别和位置，对该深度学习网络进行训练，具体过程如下：针对

样本图像集合中的每个样本图像，采用矩形的检测框标注目标；利用深度学习网络对该样本图像中目标的位置和类别进行预测，根据预测结果和目标的标注信息确定该深度学习网络的误差，该误差采用深度学习网络的损失函数进行确定，深度学习网络的损失函数包括坐标预测损失函数、置信度损失函数以及类别损失函数，分别如下所示：

(1) 坐标预测损失函数：

$$\sum \sum P_{ij} [(u_i - \hat{u}_i)^2 + (v_i - \hat{v}_i)^2] + \sum \sum P_{ij} [(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (h_i - \sqrt{\hat{h}_i})^2]$$

其中， P_{ij} 表示预测的第 j 个检测框内目标的中心点是否位于第 i 个网格内， u_i 为预测的目标的中心点在第 i 个网格的横坐标， $\hat{u}_i$ 为标注的目标的中心点在第 i 个网格的横坐标， v_i 为预测的目标的中心点在第 i 个网格的纵坐标， $\hat{v}_i$ 为标注的目标的中心点在第 i 个网格的纵坐标， w_i 为预测的中心点在第 i 个网格内的目标所在检测框的宽度， $\hat{w}_i$ 为标注的中心点在第 i 个网格内的目标所在检测框的宽度， h_i 为预测的中心点在第 i 个网格内的目标所在检测框的高度， $\hat{h}_i$ 为标注的中心点在第 i 个网格内的目标所在检测框的高度；

(2) 置信度损失函数：

$$\sum \sum P_{ij} [(Conf_i - \widehat{Conf}_i)^2]$$

其中， P_{ij} 表示预测的第 j 个检测框内目标的中心点是否位于第 i 个网格内， $Conf_i$ 为预测的置信度， $\widehat{Conf}_i$ 为标注的置信度；

(3) 类别损失函数：

$$P_i \sum [(p_i(m) - \hat{p}_i(m))^2]$$

其中， P_i 表示预测的第 i 个网络是否存在目标的中心点， $p_i(m)$ 为预测的第 i 个网格内的目标归属于类别 m 的概率， $\hat{p}_i(m)$ 为标注的第 i 个网格内的目标归属于类别 m 的概率。

S102，对检测到的每帧图像中包含的各人体进行姿态识别，得到各人体的姿态。

在本实施例中，人体的姿态包括关节点的位置和关节点之间的连线，关节点包括头、左肩关节点、右肩关节点、颈关节点、腰关节点、左膝关节点、右膝关节点、左手腕关节点、右手腕关节点、左手肘关节点、右手肘关节点、左脚腕关节点、右脚腕关节点。

在本实施例中，将每帧图像中包含的各人体输入到预先训练完成的人体姿态检测网络中，获取人体的姿态，具体地，人体姿态检测网络包括用于提取高维特征的前馈神经网络、关节点位置预测网络以及关节点关系预测网络，其中，前馈神经网络包括 10 层卷积网络和 2 层池化网络，对每帧图像中包含的人体进行高维特征提取；关节点位置预测网络包括 5 层卷积网络，输出结果为每帧图像中第 K 个人体第 j 种关节点的置信度 $S_{j,k}^*(p)$ ，用于根据该高维特征确定该人体的关节点的位置；关节点关系预测网络用于估计两个关节点之间的连接方向，根

据该关节点的位置确定关节点之间的连线，将属于同一人体的关节点的位置以及该关节点之间的连线作为该人体的姿态。

在对多个人体关节点进行连线时，由于多个关节点之间是可以建立多种连线方式的，但是符合人体的构造，保证连线是可以表示人体的某个结构的连线方式只有一种，例如，只有将腕关节点和肘关节点连线才可以表示人体的胳膊，因此，按照人体构造对多个人体关节点进行连线只会有一种方式，这样，在连线后，便可以基于人体关节点和连线表现出人体的姿态。具体地，根据该关节点的位置确定关节点之间的连线的步骤包括：

第一步，针对每两个关节点，根据该两个关节点的位置获取该两个关节点的方向向量，并将该两个关节点的方向向量分解为平行方向向量和垂直方向向量。

具体地，判断第一关节点（位置为 a_1 ）和第二关节点（位置为 a_2 ）是否为第一关节（例如为左臂或右臂）的两端，第一关节点和第二关节点的方向向量 $\vec{v}$ ，将方向向量 $\vec{v}$ 分解为平行方向向量 $\vec{v}_{//}$ 和垂直方向向量 $\vec{v}_{\perp}$ ，其中， $\vec{v} = \frac{a_1 - a_2}{\|a_1 - a_2\|}$ ， $\vec{v} = \vec{v}_{//} + \vec{v}_{\perp}$ 。

第二步，针对该两个关节点之间的每个像素点，根据该像素点的位置以及该两个关节点的方向向量判断该像素点是否位于第一关节上。

具体地，第一关节的长度为 L ，第一关节的宽度为 w ，第一关节点（ a_1 ）和第二关节点（ a_2 ）之间的像素点 p ， p 为该像素点的位置，当像素点 p 满足 $\|\vec{v}\| \geq \vec{v}(p - a_1) \geq 0$ 且 $\|\vec{v}_{\perp}(p - a_1)\| \leq w$ 时，像素点 p 位于该第一关节上，第一关节点（ a_1 ）和第二关节点（ a_2 ）具有相关性。

第三步，若该像素点位于该第一关节上，根据相关性函数计算该两个关节点与该第一关节的相关度，将相关度最高的该两个关节点作为该第一关节的两端，生成该两个关节点之间的连线。

具体地，相关性函数为 $E = \int_{u=0}^{u=1} L(p(u)) \cdot \frac{a_1 - a_2}{\|a_1 - a_2\|} du$ ，其中， $p(u)$ 为对从第一关节点（ a_1 ）和第二关节点（ a_2 ）之间的像素进行采样， $p(u) = (1 - u)a_1 + ua_2$ 。

S103，将该视频流中连续的多帧图像中各人体的姿态输入到预先训练完成的第一卷积神经网络中，获取第一动作识别结果，该第一卷积神经网络用于动作识别，该第一动作识别结果包括各人体不同动作类别的发生概率。

在本实施例中，第一卷积神经网络针对通用动作进行分类，第一卷积神经网络为图卷积神经网络，步骤 S103 具体包括如下步骤：

对该视频流中连续的多帧图像中各人体的姿态进行归一化处理；

利用注意力网络从该视频流的每帧图像中提取感兴趣区域；

对该视频流的每帧图像中各人体的不同关节点进行图卷积操作；

对该视频流中连续的多帧图像中各人体的相同关节点进行时间卷积操作；

利用全连接层根据图卷积操作输出的特征和时间卷积操作输出的特征进行动作分类，获取各人体不同动作类别的出现概率。

具体地，动作分类的公式如下：

$$g_{out}(x) = \sum_{h=1}^K \sum_{w=1}^K f_{in}(\vec{p}(x, h, w)) \cdot \vec{w}(h, w)$$

其中： g_{out} 为分类结果； f_{in} 为特征图； $\vec{p}$ 为采样函数， $\vec{p}(x, h, w) = \vec{p}(v_{ti}, v_{tj}) = v_{tj}$ ，即当前关节点 v_{ti} 距离最近的关节点 v_{tj} ； x 为关节点位置； w 为权重； $\vec{w}$ 为加权函数； K 为卷积核大小；空间域中， $\vec{w}(h, w) = \vec{w}(l_{ti}(v_{tj}))$ ；时间域中， $\vec{w}(h, w) = \vec{w}(l_{ti}(v_{tj}) + (q - t + \Gamma/2))$

$$l_{ti}(v_{tj}) = \begin{cases} 0 & \text{if } r_j = r_i \\ 1 & \text{if } r_j < r_i \\ 2 & \text{if } r_j > r_i \end{cases}$$

其中， r_i 为当前关节点 v_{ti} 到人体中心的距离； r_j 为相邻关节点 v_{tj} 到人体中心的距离； Γ 为采样时间窗口大小； q 为采样时间； t 为当前时间。

S104，获取各人体周围的物品，将该视频流中连续的多帧图像中各人体的姿态以及各人体周围的物品输入到预先训练完成的第二卷积神经网络中，获取第二动作识别结果，该第二卷积神经网络用于摔倒动作识别，该第二动作识别结果包括各人体摔倒的发生概率。

在本实施例中，该人体周围的物品为每帧图像中与该人体的距离小于或等于预设阈值的物品。

在本实施例中，根据人体的姿态、该人体周围的物品以及该物品相对于该人体的位置进行摔倒识别，例如识别出要摔倒的人体以及周围的桌子，椅子的语义信息和位置信息，然后判断，如果该摔倒的人距离桌椅很近，很可能没有摔倒，如果距离桌椅很远，则有较大的概率可能摔倒。如果检测到摔倒的人体的下方是床铺或者瑜伽垫，此时可以判断出该行人没有摔倒，只是躺下或者做一些运动，配合周围环境的语义信息，大大提高了检测动作的准确度。

在本实施例中，该第二卷积神经网络是以发生摔倒时人体的姿态、该人体周围的物品以及该物品相对于该人体的位置作为样本集训练得到的。具体地，在本实施例中，第二卷积神经网络的训练过程包括：

S1041，分别获取包含发生摔倒动作的人体的第一样本图像以及包含发生类摔倒动作的人体的第二样本图像，分别检测第一样本图像包含的人体和物品以及第二样本图像包含的人体和物品；

S1042，分别对检测到的第一样本图像及第二样本图像中包含的人体进行姿态识别，得到所述人体的姿态；

S1043，在第一样本图像中，获取与所述人体的距离小于或等于所述预设阈值的物品，作为所述人体周围的物品，根据所述人体的位置以及所述人体周围的物品的位置确定所述物品

相对于所述人体的位置；将所述人体的姿态、所述人体周围的物品以及所述物品相对于所述人体的位置作为摔倒训练特征在所述第一样本图像中进行标注，得到第一标注样本图像；

S1044，在第二样本图像中，获取与所述人体的距离小于或等于所述预设阈值的物品，作为所述人体周围的物品，根据所述人体的位置以及所述人体周围的物品的位置确定所述物品相对于所述人体的位置；将所述人体的姿态、所述人体周围的物品以及所述物品相对于所述人体的位置作为非摔倒训练特征在所述第二样本图像中进行标注，得到第二标注样本图像；

S1045，将所述第一标注样本图像和所述第二样本标注图像输入预设的初始神经网络中进行训练，以获得第二卷积神经网络。

S105，根据该第一动作识别结果和该第二动作识别结果输出各人体的行为识别结果。

在本实施例中，通过分别对第一动作识别结果和第二动作识别结果设置对应权重，根据第一动作识别结果中人体不同动作类别的发生概率与第一动作识别结果的权重以及第二识别结果中人体摔倒的发生概率与第二识别结果的权重计算人体不同动作类别的调整概率以及人体摔倒的调整概率，将调整概率最大的动作类别作为人体的行为识别结果输出。

图2是本申请第二实施例的基于环境语义理解的人体行为识别方法的流程示意图。需要注意的是，若有实质上相同的结果，本申请的方法并不以图2所示的流程顺序为限。如图2所示，该基于环境语义理解的人体行为识别方法包括步骤：

S201，检测视频流中各帧图像包含的人体和物品。

S202，对检测到的每帧图像中包含的各人体进行姿态识别，得到各人体的姿态。

S203，对识别到的每帧图像中包含的各人体的姿态进行去遮挡操作。

S204，将该视频流中连续的多帧图像中各人体的姿态输入到预先训练完成的第一卷积神经网络中，获取第一动作识别结果，该第一卷积神经网络用于动作识别，该第一动作识别结果包括各人体不同动作类别的发生概率。

S205，获取各人体周围的物品，将该视频流中连续的多帧图像中各人体的姿态以及各人体周围的物品输入到预先训练完成的第二卷积神经网络中，获取第二动作识别结果，该第二卷积神经网络用于摔倒动作识别，该第二动作识别结果包括各人体摔倒的发生概率。

S206，根据该第一动作识别结果和该第二动作识别结果输出各人体的行为识别结果。

步骤S201、步骤S202、步骤S204至S206分别参见第一实施例的步骤S101至步骤S105，在此不进行一一赘述。

在步骤S203中，针对每个检测框，当所述检测框内包含多个人体时，基于位于所述检测框内的各人体的姿态，获取该检测框中的多组关节点组，每组关节点组包括属于同一人体的多个关节点，该人体的检测框为框选出每帧图像中包含的该人体的外接区域；从多组关节点组中获取左肩关节点和右肩关节点位于该检测框内的关节点组；从左肩关节点和右肩关节点位于该检测框内的关节点组中选择关节点数量最多的关节点组标记为目标关节点组，将该检

测框内除目标关节点组外的关节点组标记为被遮挡关节点组。在本实施例中，每组关节点组对应一个人体，当检测框中存在多个人体时，通过步骤 S203 的去遮挡操作去除被遮挡人体的关节点组，将该目标关节点组对应的人体的姿态作为动作识别的对象，在后续步骤 S204 和步骤 S205 中，根据目标关节点组对应的人体的姿态进行动作分类。在步骤 S205 中，在本实施例中，该人体周围的物品为每帧图像中与该人体的距离小于或等于预设阈值的物品。

在本实施例中，多人重叠场景设计算法来去遮挡，从而进一步避免利用被遮挡的人的位置信息识别未被遮挡人的行为动作，增加了算法的可靠性，提高了算法的准确率，从而使之可以在实际复杂场景之中应用。

在一个可选的实施方式中，步骤 S206 之后还包括如下步骤：

所述人体行为识别方法还包括：将所述各人体的姿态及各人体的行为识别结果上传至区块链中，以使得所述区块链对所述各人体的姿态及各人体的行为识别结果进行加密存储。

具体地，基于各人体的姿态或各人体的行为识别结果得到对应的摘要信息，具体来说，摘要信息由各人体的姿态或各人体的行为识别结果进行散列处理得到，比如利用 sha256s 算法处理得到。将摘要信息上传至区块链可保证其安全性和对用户的公正透明性。用户设备可以从区块链中下载得该摘要信息，以便查证人体的行为识别结果是否被篡改。本示例所指区块链是分布式数据存储、点对点传输、共识机制、加密算法等计算机技术的新型应用模式。区块链(Blockchain)，本质上是一个去中心化的数据库，是一串使用密码学方法相关联产生的数据块，每一个数据块中包含了一批次网络交易的信息，用于验证其信息的有效性(防伪)和生成下一个区块。区块链可以包括区块链底层平台、平台产品服务层以及应用服务层等。

图 3 是本申请第三实施例的基于环境语义理解的人体行为识别装置的结构示意图。如图 3 所示，该装置 30 包括目标检测模块 301、姿态识别模块 302、通用动作分类模块 303、摔倒动作识别模块 304 和输出模块 305。

其中，目标检测模块 301 用于检测视频流中各帧图像包含的人体和物品；姿态识别模块 302 用于对检测到的每帧图像中包含的各人体进行姿态识别，得到各人体的姿态；通用动作分类模块 303 用于将该视频流中连续的多帧图像中各人体的姿态输入到预先训练完成的第一卷积神经网络中，获取第一动作识别结果，该第一卷积神经网络用于动作识别，该第一动作识别结果包括各人体不同动作类别的发生概率；摔倒动作识别模块 304 用于获取各人体周围的物品，将该视频流中连续的多帧图像中各人体的姿态以及各人体周围的物品输入到预先训练完成的第二卷积神经网络中，获取第二动作识别结果，人体周围的物品为每帧图像中与人体的距离小于或等于预设阈值的物品，该第二卷积神经网络用于摔倒动作识别，该第二动作识别结果包括各人体摔倒的发生概率；输出模块 305 用于根据该第一动作识别结果和该第二动作识别结果输出各人体的行为识别结果。

图 4 是本申请第四实施例的电子设备的结构示意图。如图 4 所示，该电子设备 40 包括处

理器 41 及和处理器 41 耦接的存储器 42。

存储器 42 存储有用于实现上述任一实施例该基于环境语义理解的人体行为识别的程序指令。

处理器 41 用于执行存储器 42 存储的程序指令以进行基于环境语义理解的人体行为识别。

其中，处理器 41 还可以称为 CPU (Central Processing Unit, 中央处理单元)。处理器 41 可能是一种集成电路芯片，具有信号的处理能力。处理器 41 还可以是通用处理器、数字信号处理器 (DSP)、专用集成电路 (ASIC)、现场可编程门阵列 (FPGA) 或者其他可编程逻辑器件、分立门或者晶体管逻辑器件、分立硬件组件。通用处理器可以是微处理器或者该处理器也可以是任何常规的处理器等。

参阅图 5，图 5 为本申请实施例的存储介质的结构示意图。本申请实施例的存储介质存储有能够实现上述所有方法的程序指令 51，所述存储介质可以是非易失性，也可以是易失性。其中，该程序指令 51 可以以软件产品的形式存储在上述存储介质中，包括若干指令用以使得一台计算机设备 (可以是个人计算机，服务器，或者网络设备等) 或处理器 (processor) 执行本申请各个实施方式所述方法的全部或部分步骤。而前述的存储装置包括：U 盘、移动硬盘、只读存储器 (ROM, Read-Only Memory)、随机存取存储器 (RAM, Random Access Memory)、磁碟或者光盘等各种可以存储程序代码的介质，或者是计算机、服务器、手机、平板等终端设备。

在本申请所提供的几个实施例中，应该理解到，所揭露的系统，装置和方法，可以通过其它的方式实现。例如，以上所描述的装置实施例仅仅是示意性的，例如，单元的划分，仅仅为一种逻辑功能划分，实际实现时可以有另外的划分方式，例如多个单元或组件可以结合或者可以集成到另一个系统，或一些特征可以忽略，或不执行。另一点，所显示或讨论的相互之间的耦合或直接耦合或通信连接可以是通过一些接口，装置或单元的间接耦合或通信连接，可以是电性，机械或其它的形式。

另外，在本申请各个实施例中的各功能单元可以集成在一个处理单元中，也可以是各个单元单独物理存在，也可以两个或两个以上单元集成在一个单元中。上述集成的单元既可以采用硬件的形式实现，也可以采用软件功能单元的形式实现。以上仅为本申请的实施方式，并非因此限制本申请的专利范围，凡是利用本申请说明书及附图内容所作的等效结构或等效流程变换，或直接或间接运用在其他相关的技术领域，均同理包括在本申请的专利保护范围。

以上所述的仅是本申请的实施方式，在此应当指出，对于本领域的普通技术人员来说，在不脱离本申请创造构思的前提下，还可以做出改进，但这些均属于本申请的保护范围。

权 利 要 求 书

1、一种基于环境语义理解的人体行为识别方法，其中，所述方法包括：

检测视频流中各帧图像包含的人体和物品；

对检测到的每帧图像中包含的各人体进行姿态识别，得到各人体的姿态；

将所述视频流中连续的多帧图像中各人体的姿态输入到预先训练完成的第一卷积神经网络中，获取第一动作识别结果，所述第一卷积神经网络用于动作识别，所述第一动作识别结果包括各人体不同动作类别的发生概率；

获取各人体周围的物品，将所述视频流中连续的多帧图像中各人体的姿态以及各人体周围的物品输入到预先训练完成的第二卷积神经网络中，获取第二动作识别结果，所述人体周围的物品为每帧图像中与所述人体的距离小于或等于预设阈值的物品，所述第二卷积神经网络用于摔倒动作识别，所述第二动作识别结果包括各人体摔倒的发生概率；

根据所述第一动作识别结果和所述第二动作识别结果输出各人体的行为识别结果。

2、根据权利要求1所述的人体行为识别方法，其中，所述检测视频流中各帧图像包含的人体和物品，包括：

按照预设的划分方式，将所述视频流中各帧图像划分为多个网格；

在每个网格中，通过预先设置的不同类型的检测框进行目标预测，针对每个检测框，获取所述检测框预测的目标的坐标参数、检测框的宽度和高度以及检测框的置信度，将置信度最高的检测框作为预测结果，所述预测结果包括所述目标、所述检测框、所述目标的坐标参数以及所述目标的类别，所述检测框为框选出所述目标的外接区域，所述目标的类别包括人体和物品；

根据所述预测结果确定所述视频流中各帧图像包含的人体和物品。

3、根据权利要求1所述的人体行为识别方法，其中，所述人体的姿态包括关节点的位置和关节点之间的连线；所述对检测到的每帧图像中包含的各人体进行姿态识别，得到各人体的姿态，包括：

对每帧图像中包含的人体进行高维特征提取；

根据所述高维特征确定所述人体的关节点的位置；

根据所述关节点的位置确定关节点之间的连线，将所述关节点的位置以及所述关节点之间的连线作为所述人体的姿态。

4、根据权利要求3所述的人体行为识别方法，其中，所述根据所述关节点的位置确定关节点之间的连线，包括：

针对每两个关节点，根据所述两个关节点的位置获取所述两个关节点的方向向量，将所述两个关节点的方向向量分解为平行方向向量和垂直方向向量；

针对所述两个关节点之间的每个像素点，根据所述像素点的位置以及所述两个关节点的方向向量判断所述像素点是否位于第一关节上；

若所述像素点位于所述第一关节上，根据相关性函数计算所述两个关节点的相关度，将相关度最高的所述两个关节点作为所述第一关节的两端，生成所述两个关节点之间的连线。

5、根据权利要求1所述的人体行为识别方法，其中，所述对检测到的每帧图像中包含的各人体进行姿态识别，得到各人体的姿态之后，还包括：

针对每个人体的检测框，当所述检测框内包含多个人体时，基于位于所述检测框内的各人体的姿态，获取所述检测框中的多组关节点组，每组所述关节点组包括属于同一人体的多个关节点，所述人体的检测框为框选出每帧图像中包含的所述人体的外接区域；

从多组关节点组中获取左肩关节点和右肩关节点位于所述检测框内的关节点组；

从左肩关节点和右肩关节点位于所述检测框内的关节点组中选择关节点数量最多的关节点组标记为目标关节点组，将所述检测框内除目标关节点组外的关节点组标记为被遮挡关节点组，将所述目标关节点组对应的人体的姿态作为动作识别的对象。

6、根据权利要求1所述的人体行为识别方法，其中，所述将所述视频流中连续的多帧图像中各人体的姿态输入到预先训练完成的第一卷积神经网络中，获取所述人体的第一动作识别结果，包括：

利用注意力网络从所述视频流的每帧图像中提取感兴趣区域；

对所述视频流的每帧图像中各人体的不同关节点进行图卷积操作；

对所述视频流中连续的多帧图像中各人体的相同关节点进行时间卷积操作；

利用全连接层根据图卷积操作输出的特征和时间卷积操作输出的特征进行动作分类，获取各人体不同动作类别的出现概率。

7、根据权利要求6所述的人体行为识别方法，其中，所述人体行为识别方法还包括：将所述各人体的姿态及各人体的行为识别结果上传至区块链中，以使得所述区块链对所述各人体的姿态及各人体的行为识别结果进行加密存储；

所述利用注意力网络从所述视频流的每帧图像中提取感兴趣区域之前，还包括：对所述视频流中连续的多帧图像中各人体的姿态进行归一化处理。

8、根据权利要求1所述的人体行为识别方法，其中，所述第二卷积神经网络的训练过程包括：

获取包含发生摔倒动作的人体的第一样本图像，检测所述第一样本图像包含的人体和物品；

对检测到的所述第一样本图像中包含的人体进行姿态识别，得到所述人体的姿态；

获取与所述人体的距离小于或等于所述预设阈值的物品，作为所述人体周围的物品，根据所述人体的位置以及所述人体周围的物品的位置确定所述物品相对于所述人体的位置；

将所述人体的姿态、所述人体周围的物品以及所述物品相对于所述人体的位置作为摔倒训练特征在所述第一样本图像中进行标注，得到第一标注样本图像；

将所述第一标注样本图像输入预设的初始神经网络中进行训练，以获得第二卷积神经网络。

9、一种基于环境语义理解的人体行为识别装置，其中，所述装置包括：

目标检测模块，用于检测视频流中各帧图像包含的人体和物品；

姿态识别模块，用于对检测到的每帧图像中包含的各人体进行姿态识别，得到各人体的姿态；

通用动作分类模块，用于将所述视频流中连续的多帧图像中各人体的姿态输入到预先训练完成的第一卷积神经网络中，获取第一动作识别结果，所述第一卷积神经网络用于动作识别，所述第一动作识别结果包括各人体不同动作类别的发生概率；

摔倒动作识别模块，用于获取各人体周围的物品，将所述视频流中连续的多帧图像中各人体的姿态以及各人体周围的物品输入到预先训练完成的第二卷积神经网络中，获取第二动作识别结果，所述人体周围的物品为每帧图像中与所述人体的距离小于或等于预设阈值的物品，所述第二卷积神经网络用于摔倒动作识别，所述第二动作识别结果包括各人体摔倒的发生概率；

输出模块，用于根据所述第一动作识别结果和所述第二动作识别结果输出各人体的行为识别结果。

10、一种电子设备，其中，所述电子设备包括处理器、以及与所述处理器耦接的存储器，所述存储器存储有可被所述处理器执行的程序指令；所述处理器执行所述存储器存储的所述程序指令时实现以下步骤：

检测视频流中各帧图像包含的人体和物品；

对检测到的每帧图像中包含的各人体进行姿态识别，得到各人体的姿态；

将所述视频流中连续的多帧图像中各人体的姿态输入到预先训练完成的第一卷积神经网络中，获取第一动作识别结果，所述第一卷积神经网络用于动作识别，所述第一动作识别结果包括各人体不同动作类别的发生概率；

获取各人体周围的物品，将所述视频流中连续的多帧图像中各人体的姿态以及各人体周围的物品输入到预先训练完成的第二卷积神经网络中，获取第二动作识别结果，所述人体周围的物品为每帧图像中与所述人体的距离小于或等于预设阈值的物品，所述第二卷积神经网络用于摔倒动作识别，所述第二动作识别结果包括各人体摔倒的发生概率；

根据所述第一动作识别结果和所述第二动作识别结果输出各人体的行为识别结果。

11、根据权利要求10所述的电子设备，其中，所述检测视频流中各帧图像包含的人体和物品，包括：

按照预设的划分方式，将所述视频流中各帧图像划分为多个网格；

在每个网格中，通过预先设置的不同类型的检测框进行目标预测，针对每个检测框，获取所述检测框预测的目标的坐标参数、检测框的宽度和高度以及检测框的置信度，将置信度最高的检测框作为预测结果，所述预测结果包括所述目标、所述检测框、所述目标的坐标参数以及所述目标的类别，所述检测框为框选出所述目标的外接区域，所述目标的类别包括人体和物品；

根据所述预测结果确定所述视频流中各帧图像包含的人体和物品。

12、根据权利要求 10 所述的电子设备，其中，所述人体的姿态包括关节点的位置和关节点之间的连线；所述对检测到的每帧图像中包含的各人体进行姿态识别，得到各人体的姿态，包括：

对每帧图像中包含的人体进行高维特征提取；

根据所述高维特征确定所述人体的关节点的位置；

根据所述关节点的位置确定关节点之间的连线，将所述关节点的位置以及所述关节点之间的连线作为所述人体的姿态。

13、根据权利要求 10 所述的电子设备，其中，所述对检测到的每帧图像中包含的各人体进行姿态识别，得到各人体的姿态之后，还包括：

针对每个人体的检测框，当所述检测框内包含多个人体时，基于位于所述检测框内的各人体的姿态，获取所述检测框中的多组关节点组，每组所述关节点组包括属于同一人体的多个关节点，所述人体的检测框为框选出每帧图像中包含的所述人体的外接区域；

从多组关节点组中获取左肩关节点和右肩关节点位于所述检测框内的关节点组；

从左肩关节点和右肩关节点位于所述检测框内的关节点组中选择关节点数量最多的关节点组标记为目标关节点组，将所述检测框内除目标关节点组外的关节点组标记为被遮挡关节点组，将所述目标关节点组对应的人体的姿态作为动作识别的对象。

14、根据权利要求 10 所述的电子设备，其中，所述将所述视频流中连续的多帧图像中各人体的姿态输入到预先训练完成的第一卷积神经网络中，获取所述人体的第一动作识别结果，包括：

利用注意力网络从所述视频流的每帧图像中提取感兴趣区域；

对所述视频流的每帧图像中各人体的不同关节点进行图卷积操作；

对所述视频流中连续的多帧图像中各人体的相同关节点进行时间卷积操作；

利用全连接层根据图卷积操作输出的特征和时间卷积操作输出的特征进行动作分类，获取各人体不同动作类别的出现概率。

15、根据权利要求 10 所述的电子设备，其中，所述第二卷积神经网络的训练过程包括：

获取包含发生摔倒动作的人体的第一样本图像，检测所述第一样本图像包含的人体和物

品;

对检测到的所述第一样本图像中包含的人体进行姿态识别,得到所述人体的姿态;

获取与所述人体的距离小于或等于所述预设阈值的物品,作为所述人体周围的物品,根据所述人体的位置以及所述人体周围的物品的位置确定所述物品相对于所述人体的位置;

将所述人体的姿态、所述人体周围的物品以及所述物品相对于所述人体的位置作为摔倒训练特征在所述第一样本图像中进行标注,得到第一标注样本图像;

将所述第一标注样本图像输入预设的初始神经网络中进行训练,以获得第二卷积神经网络。

16、一种存储介质,其中,所述存储介质内存储有程序指令,所述程序指令被处理器执行时实现以下步骤:

检测视频流中各帧图像包含的人体和物品;

对检测到的每帧图像中包含的各人体进行姿态识别,得到各人体的姿态;

将所述视频流中连续的多帧图像中各人体的姿态输入到预先训练完成的第一卷积神经网络中,获取第一动作识别结果,所述第一卷积神经网络用于动作识别,所述第一动作识别结果包括各人体不同动作类别的发生概率;

获取各人体周围的物品,将所述视频流中连续的多帧图像中各人体的姿态以及各人体周围的物品输入到预先训练完成的第二卷积神经网络中,获取第二动作识别结果,所述人体周围的物品为每帧图像中与所述人体的距离小于或等于预设阈值的物品,所述第二卷积神经网络用于摔倒动作识别,所述第二动作识别结果包括各人体摔倒的发生概率;

根据所述第一动作识别结果和所述第二动作识别结果输出各人体的行为识别结果。

17、根据权利要求16所述的存储介质,其中,所述检测视频流中各帧图像包含的人体和物品,包括:

按照预设的划分方式,将所述视频流中各帧图像划分为多个网格;

在每个网格中,通过预先设置的不同类型的检测框进行目标预测,针对每个检测框,获取所述检测框预测的目标的坐标参数、检测框的宽度和高度以及检测框的置信度,将置信度最高的检测框作为预测结果,所述预测结果包括所述目标、所述检测框、所述目标的坐标参数以及所述目标的类别,所述检测框为框选出所述目标的外接区域,所述目标的类别包括人体和物品;

根据所述预测结果确定所述视频流中各帧图像包含的人体和物品。

18、根据权利要求16所述的存储介质,其中,所述对检测到的每帧图像中包含的各人体进行姿态识别,得到各人体的姿态之后,还包括:

针对每个人体的检测框,当所述检测框内包含多个人体时,基于位于所述检测框内的各人体的姿态,获取所述检测框中的多组关节点组,每组所述关节点组包括属于同一人体的多

个关节点，所述人体的检测框为框选出每帧图像中包含的所述人体的外接区域；

从多组关节点组中获取左肩关节点和右肩关节点位于所述检测框内的关节点组；

从左肩关节点和右肩关节点位于所述检测框内的关节点组中选择关节点数量最多的关节点组标记为目标关节点组，将所述检测框内除目标关节点组外的关节点组标记为被遮挡关节点组，将所述目标关节点组对应的人体的姿态作为动作识别的对象。

19、根据权利要求 16 所述的存储介质，其中，所述将所述视频流中连续的多帧图像中各人体的姿态输入到预先训练完成的第一卷积神经网络中，获取所述人体的第一动作识别结果，包括：

利用注意力网络从所述视频流的每帧图像中提取感兴趣区域；

对所述视频流的每帧图像中各人体的不同关节点进行图卷积操作；

对所述视频流中连续的多帧图像中各人体的相同关节点进行时间卷积操作；

利用全连接层根据图卷积操作输出的特征和时间卷积操作输出的特征进行动作分类，获取各人体不同动作类别的出现概率。20、根据权利要求 16 所述的存储介质，其中，所述第二卷积神经网络的训练过程包括：

获取包含发生摔倒动作的人体的第一样本图像，检测所述第一样本图像包含的人体和物品；

对检测到的所述第一样本图像中包含的人体进行姿态识别，得到所述人体的姿态；

获取与所述人体的距离小于或等于所述预设阈值的物品，作为所述人体周围的物品，根据所述人体的位置以及所述人体周围的物品的位置确定所述物品相对于所述人体的位置；

将所述人体的姿态、所述人体周围的物品以及所述物品相对于所述人体的位置作为摔倒训练特征在所述第一样本图像中进行标注，得到第一标注样本图像；

将所述第一标注样本图像输入预设的初始神经网络中进行训练，以获得第二卷积神经网络。

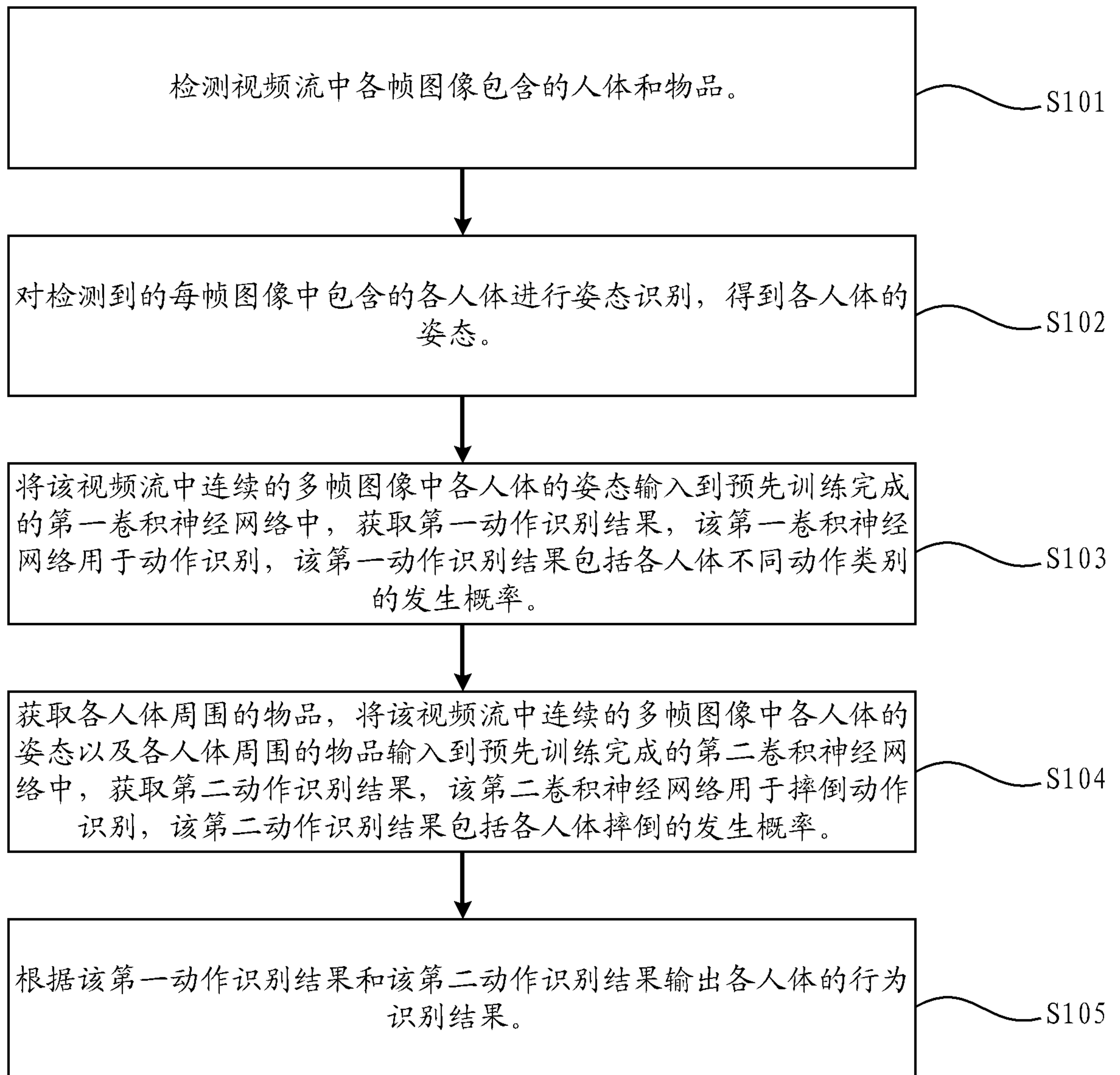


图 1

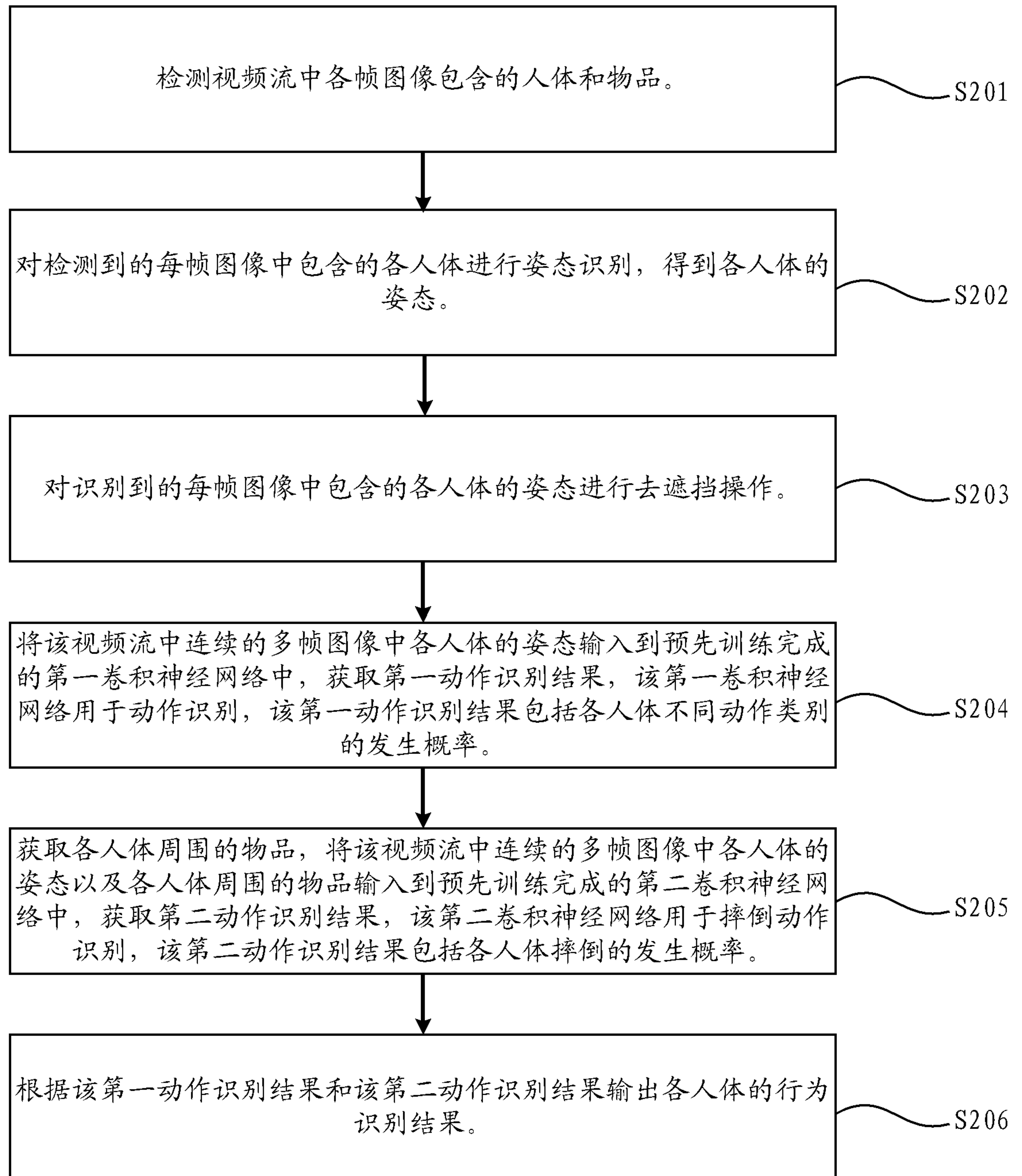


图 2

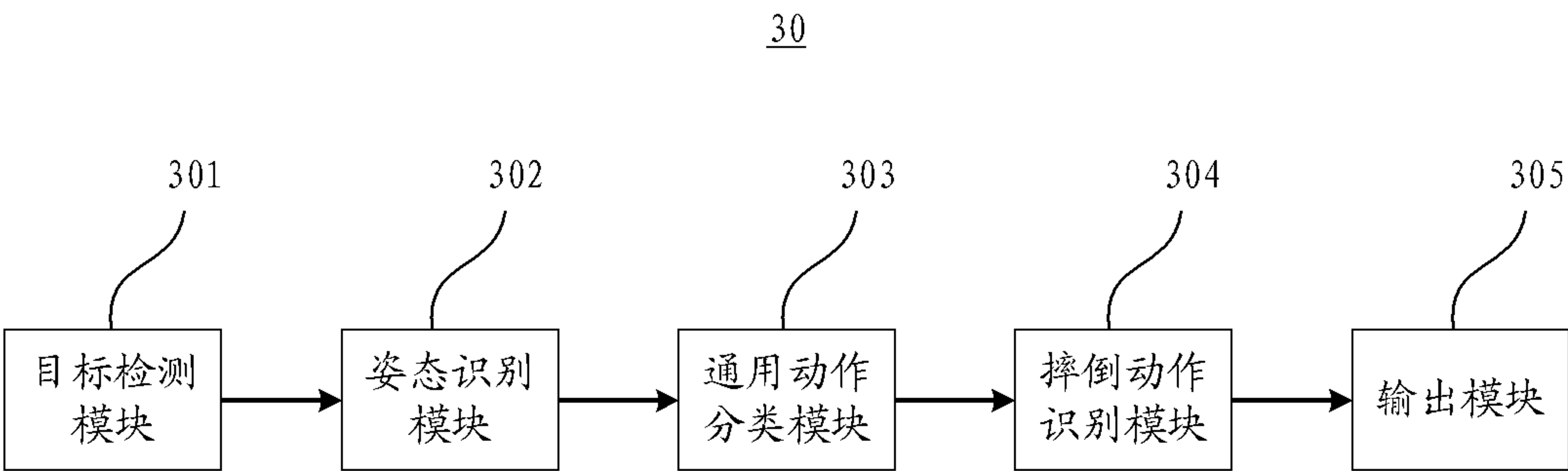


图 3

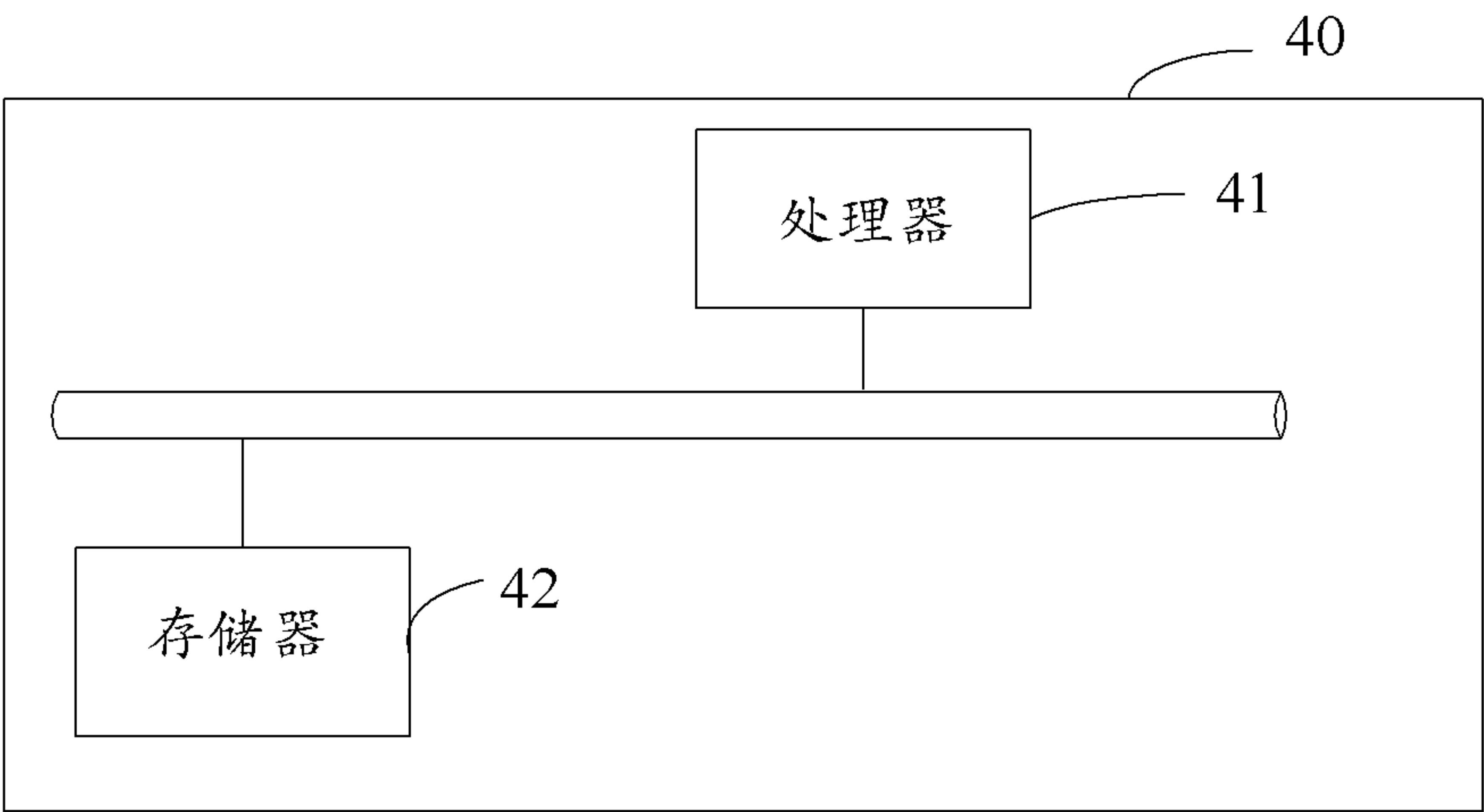


图 4

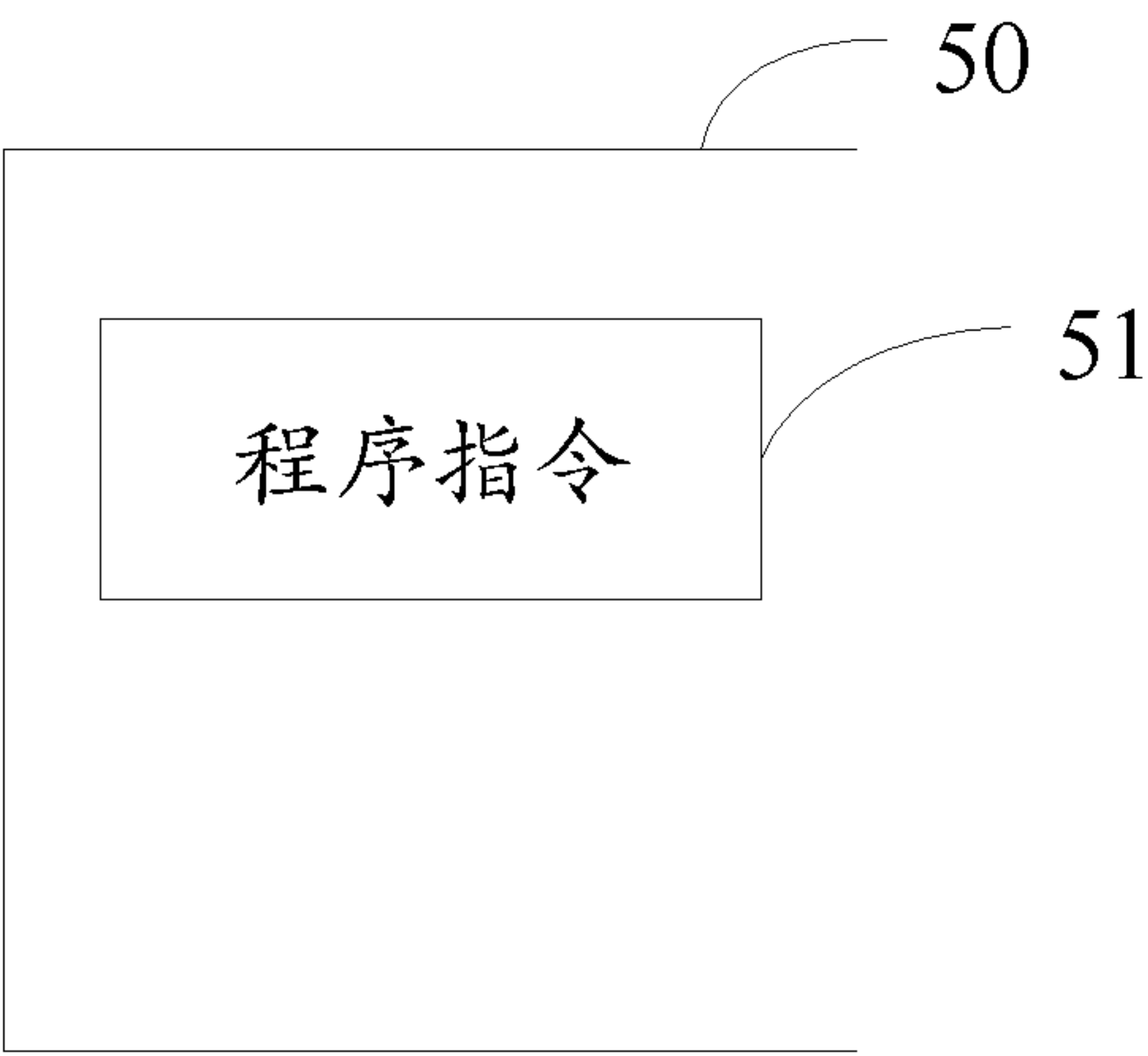


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2020/123214

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/00(2006.01)i

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

CNABS; CNTXT; VEN; USTXT; EPTXT; WOTXT; CNKI: 环境, 周围, 周边, 语义, 位置, 距离, 物体, 物品, 行为, 动作, 姿势, 姿态, 识别, 检测, circumstance, surrounding, semantic, position, distance, object, behavior, motion, pose, recognition, detection

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
PX	CN 111666857 A (PING AN TECHNOLOGY (SHENZHEN) CO., LTD.) 15 September 2020 (2020-09-15) claims 1-10, description paragraphs 54-122	1-20
Y	周以重 (ZHOU, Yichong). "人与物体交互行为算法的研究与应用 (Investigation and Application of Human-Object Interaction Detection Algorithm)" 中国优秀硕士学位论文全文数据库信息科技辑 (Information & Technology, China Master's Theses Full-Text Database), No. 1, 15 January 2020 (2020-01-15), ISSN: 1674-0246, chapter 4	1-20
Y	YAN, Sijie et al. "Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition" https://arxiv.org/pdf/1801.07455v2 , 25 January 2018 (2018-01-25), sections 1-4	1-20
Y	REDMON, Joseph et al. "You Only Look Once: Unified, Real-Time Object Detection" https://arxiv.org/pdf/1506.02640v5 , 09 May 2016 (2016-05-09), sections 1-2	2, 11, 17

☒ Further documents are listed in the continuation of Box C.

☒ See patent family annex.

* Special categories of cited documents:	"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A" document defining the general state of the art which is not considered to be of particular relevance	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"E" earlier application or patent but published on or after the international filing date	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"&" document member of the same patent family
"O" document referring to an oral disclosure, use, exhibition or other means	
"P" document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

21 February 2021

Date of mailing of the international search report

03 March 2021

Name and mailing address of the ISA/CN

China National Intellectual Property Administration (ISA/
CN)
No. 6, Xitucheng Road, Jimenqiao, Haidian District, Beijing
100088
China

Facsimile No. (86-10)62019451

Authorized officer

Telephone No.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	CAO, Zhe et al. "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields" <i>https://arxiv.org/pdf/1611.08050v2</i> , 14 April 2017 (2017-04-14), sections 1-2	3-5, 12-13, 18
A	CN 110610154 A (BEIJING MEGVII TECHNOLOGY CO., LTD.) 24 December 2019 (2019-12-24) entire document	1-20
A	US 2014334668 A1 (PALO ALTO RESEARCH CENTER INCORPORATED) 13 November 2014 (2014-11-13) entire document	1-20

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/CN2020/123214

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	111666857	A	15 September 2020	None			
CN	110610154	A	24 December 2019	None			
US	2014334668	A1	13 November 2014	US	9025825	B2	05 May 2015

国际检索报告		国际申请号 PCT/CN2020/123214
A. 主题的分类 G06K 9/00 (2006.01) i 按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类		
B. 检索领域 检索的最低限度文献 (标明分类系统和分类号) G06K 包含在检索领域中的除最低限度文献以外的检索文献 在国际检索时查阅的电子数据库 (数据库的名称, 和使用的检索词 (如使用)) CNABS; CNTXT; VEN; USTXT; EPTXT; WOTXT; CNKI: 环境, 周围, 周边, 语义, 位置, 距离, 物体, 物品, 行为, 动作, 姿势, 姿态, 识别, 检测, circumstance, surrounding, semantic, position, distance, object, behavior, motion, pose, recognition, detection		
C. 相关文件		
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求
PX	CN 111666857 A (平安科技深圳有限公司) 2020年 9月 15日 (2020 - 09 - 15) 权利要求1-10, 说明书第54-122段	1-20
Y	周以重. “人与物体交互行为算法的研究与应用” 中国优秀硕士学位论文全文数据库信息科技辑, 第1期, 2020年 1月 15日 (2020 - 01 - 15), ISSN: 1674-0246, 第4章	1-20
Y	YAN, Sijie 等. “Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition” https://arxiv.org/pdf/1801.07455v2 , 2018年 1月 25日 (2018 - 01 - 25), 第1-4节	1-20
Y	REDMON, Joseph 等. “You Only Look Once: Unified, Real-Time Object Detection” https://arxiv.org/pdf/1506.02640v5 , 2016年 5月 9日 (2016 - 05 - 09), 第1-2节	2、11、17
☒ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。		
* 引用文件的具体类型: “A” 认为不特别相关的表示了现有技术一般状态的文件 “E” 在国际申请日的当天或之后公布的在先申请或专利 “L” 可能对优先权要求构成怀疑的文件, 或为确定另一篇引用文件的公布日而引用的或者因其他特殊理由而引用的文件 (如具体说明的) “O” 涉及口头公开、使用、展览或其他方式公开的文件 “P” 公布日先于国际申请日但迟于所要求的优先权日的文件 “T” 在申请日或优先权日之后公布, 与申请不相抵触, 但为了解发明之理论或原理的在后文件 “X” 特别相关的文件, 单独考虑该文件, 认定要求保护的发明不是新颖的或不具有创造性 “Y” 特别相关的文件, 当该文件与另一篇或者多篇该类文件结合并且这种结合对于本领域技术人员为显而易见时, 要求保护的发明不具有创造性 “&” 同族专利的文件		
国际检索实际完成的日期 2021年 2月 21日		国际检索报告邮寄日期 2021年 3月 3日
ISA/CN的名称和邮寄地址 中国国家知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		授权官员 马婷婷 电话号码 86-(20)-28958507

C. 相关文件		
类 型*	引用文件，必要时，指明相关段落	相关的权利要求
Y	CAO, Zhe 等. "Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields" https://arxiv.org/pdf/1611.08050v2, 2017年 4月 14日 (2017 - 04 - 14), 第1-2节	3-5、12-13、18
A	CN 110610154 A (北京迈格威科技有限公司) 2019年 12月 24日 (2019 - 12 - 24) 全文	1-20
A	US 2014334668 A1 (PALO ALTO RESEARCH CENTER INCORPORATED) 2014年 11月 13日 (2014 - 11 - 13) 全文	1-20

国际申请号
PCT/CN2020/123214

检索报告引用的专利文件			公布日 (年/月/日)	同族专利	公布日 (年/月/日)
CN	111666857	A	2020年 9月 15日	无	
CN	110610154	A	2019年 12月 24日	无	
US	2014334668	A1	2014年 11月 13日	US	9025825 B2 2015年 5月 5日