

对象状态识别方法、装置、计算机设备和存储介质

本申请要求于 2024 年 7 月 11 日提交中国专利局，申请号为 2024109290569，申请名称为“对象状态识别方法、装置、计算机设备和存储介质”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

本申请涉及计算机技术领域，特别是涉及一种对象状态识别方法、装置、计算机设备、存储介质和计算机程序产品。

背景技术

随着计算机技术的发展，出现了物联网技术。物联网是指通过信息传感设备，按约定的协议，将任何物体与网络相连接，物体通过信息传播媒介进行信息交换和通信，以实现智能化识别、定位、跟随、监管等功能。

目前，通过物联网技术可以识别用户状态。主要是通过穿戴式设备中的加速度计、陀螺仪、磁力计等传感器，来测量穿戴式设备在空间中的运动变化和运动方向，从而确定穿戴式设备的佩戴者身体的姿态和运动方向，进而识别佩戴者的对象状态。然而，有些对象状态在运动学信号的表现中本身就存在相似性，较难区分，目前的识别方式存在准确性较低的问题。

发明内容

本申请实施例提供一种对象状态识别方法、装置、计算机设备、计算机可读存储介质和计算机程序产品。

本申请提供了一种对象状态识别方法，由计算机设备执行，包括：

获取目标视频；

获取所述目标视频对应的背景帧，从所述目标视频中确定目标对象的目标动作所对应的目标帧；

对所述背景帧和所述目标帧进行图像分割，得到所述背景帧中各个物体分别对应的物体区域和所述目标帧中所述目标对象对应的目标区域；

基于所述目标区域和所述物体区域的位置信息、所述物体区域对应的物体、所述目标动作，生成交互信息组，得到所述目标对象分别和所述各个物体之间的交互信息组；

基于各个交互信息组，确定所述目标对象在所述目标帧对应的目标对象状态。

本申请还提供了一种对象状态识别装置，包括：

视频获取模块，用于获取目标视频；

图像确定模块，用于获取所述目标视频对应的背景帧，从所述目标视频中确定目标对象的目标动作所对应的目标帧；

图像分割模块，用于对所述背景帧和所述目标帧进行图像分割，得到所述背景帧中各个物体分别对应的物体区域和所述目标帧中所述目标对象对应的目标区域；

信息组生成模块，用于基于所述目标区域和所述物体区域的位置信息、所述物体区域对应的物体、所述目标动作，生成交互信息组，得到所述目标对象分别和所述各个物体之间的交互信息组；

状态识别模块，用于基于各个交互信息组，确定所述目标对象在所述目标帧对应的目标对象状态。

本申请还提供了一种计算机设备，包括存储器和处理器，所述存储器存储有计算机可读指令，所述处理器执行所述计算机可读指令时实现上述对象状态识别方法所述的步骤。

本申请还提供了一种计算机可读存储介质，其上存储有计算机可读指令，所述计算机可读指令被处理器执行时实现上述对象状态识别方法所述的步骤。

本申请还提供了一种计算机程序产品，包括计算机可读指令，所述计算机可读指令被处理器执行时实现上述对象状态识别方法所述的步骤。

本申请的一个或多个实施例的细节在下面的附图和描述中提出。本申请的其它特征、目的和优点将从说明书、附图以及权利要求书变得明显。

附图说明

为了更清楚地说明本申请实施例或相关技术中的技术方案，下面将对实施例或相关技术描述中所需要使用的附图作简单地介绍，显而易见地，下面描述中的附图仅仅是本申请的一些实施例，对于本领域普通技术人员来讲，在不付出创造性劳动的前提下，还可以根据这些附图获得其他的附图。

图 1 为一个实施例中对象状态识别方法的应用环境图；

图 2 为一个实施例中对象状态识别方法的流程示意图；

图 3 为一个实施例中从目标视频中确定背景帧的示意图；

图 4 为一个实施例中从目标视频中确定目标帧的流程示意图；

图 5 为一个实施例中识别目标帧的示意图；

图 6 为一个实施例中根据图像分割结果进行交互分析的示意图；

图 7 为一个实施例中从参考对象状态中确定目标对象状态的流程示意图；

图 8 为一个实施例中通过交互模型确定目标对象状态的流程示意图；

图 9 为一个实施例中跌倒检测方法的示意图；

图 10 为另一个实施例中跌倒检测方法的示意图；

图 11 为一个实施例中对象状态识别装置的结构框图；

图 12 为一个实施例中计算机设备的内部结构图；

图 13 为另一个实施例中计算机设备的内部结构图。

具体实施方式

为了使本申请的目的、技术方案及优点更加清楚明白，以下结合附图及实施例，对本申请进行进一步详细说明。应当理解，此处描述的具体实施例仅仅用以解释本申请，并不用于限定本申请。

本申请实施例提供的对象状态识别方法，可以应用于如图 1 所示的应用环境中。其中，终端 102 通过网络与服务器 104 进行通信。数据存储系统可以存储服务器 104 需要处理的数据。数据存储系统可以单独设置，可以集成在服务器 104 上，也可以放在云上或其他设备上。其中，终端 102 可以但不限于各种个人计算机、笔记本电脑、智能手机、平板电脑、物联网设备和便携式可穿戴设备，物联网设备可为智能音箱、智能电视、智能空调、智能车载设备等。便携式可穿戴设备可为智能手表、智能手环、头戴设备等。服务器 104 可以用独立的服务器或者是多个服务器组成的服务器集群或者云服务器来实现。

可以理解，终端和服务器均可单独用于执行本申请实施例中提供的对象状态识别方法。终端和服务器也可协同用于执行本申请实施例中提供的对象状态识别方法。

例如，服务器获取目标视频，获取目标视频对应的背景帧，从目标视频中确定目标对象的目标动作所对应的目标帧。服务器对背景帧和目标帧进行图像分割，得到背景帧中各个物体分别对应的物体区域和目标帧中目标对象对应的目标区域。服务器基于目标区域和物体区域的位置信息、物体区域对应的物体、目标动作，生成交互信息组，得到目标对象分别和各个物体之间的交互信息组，基于各个交互信息组，确定目标对象在目标帧对应的目标对象状态。

在一个实施例中，如图 2 所示，提供了一种对象状态识别方法，以该方法应用于计算机设备来举例说明，计算机设备可以是终端或服务器。可以理解，该方法可以由终端或服务器自身单独执行，也可以通过终端和服务器之间的交互来实现。其中：

步骤 S202，获取目标视频。

其中，目标视频是指待分析的视频。可以理解，目标视频可以是预先采集好的视频。例如，目标视频可以是用于演示的安全录像。目标视频也可以是实时采集得到的视频，例如，目标视频可以是实时的看护视频。

具体地，计算机设备可以在本地或从其他设备获取目标视频，对目标视频进行分析，识别目标视频中的目标对象的对象状态。

步骤 S204，获取目标视频对应的背景帧，从目标视频中确定目标对象的目标动作所对应的目标帧。

其中，背景帧是背景图像，用于表示场景中静态或相对稳定不变的背景部分。背景帧主要用于提供场景的空间感和环境氛围，帮助识别场景所处的环境。目标视频对应的背景帧是代表目标视频中场景的背景的图像。可以理解，目标视频对应的背景帧可以是预先采集的背景图像，例如，目标视频是针对房间 A 的房间录像，在通过摄像头启动录像之前，通过同一摄像头拍摄房间 A 在无活物情况下的图像作为目标视频对应的背景帧。目标视频对应的背景帧可以是目标视频中确定的背景图像，例如，从目标视频中获取无活物的视频帧作为背景帧。

目标帧是代表目标对象处于目标动作的图像。在目标帧中，目标对象处于目标动作。目标对象是待分析的对象，目标对象具体可以根据需要进行设置。目标对象通常是活物，例如，目标对象可以是人；目标对象可以是动物；等等。目标动作是待分析的动作，目标动作具体可以根据需要进行设置。例如，目标动作可以是倒下动作；目标动作可以是吃饭动作；等等。举例说明，若目标对象为人，目标动作作为倒下动作，在目标帧中，人可以呈现躺姿。

可以理解，背景帧中不包括目标对象，目标帧中包括目标对象。

具体地，在确定目标视频后，计算机设备可以获取目标视频对应的背景帧，从目标视频中确定目标对象的目标动作所对应的目标帧，根据背景帧和目标帧，识别目标对象在目标帧对应的目标对象状态。

针对目标帧，计算机设备可以对目标视频中的视频帧进行目标对象识别，对识别出目标对象的视频帧进一步进行目标动作识别，将识别出目标动作的视频帧作为目标帧。

在其中一个实施例中，对目标视频的视频帧进行目标对象识别，将识别出目标对象的视频帧作为候选帧；获取目标时间段内的各个候选帧组成候选帧序列；将候选帧序列输入动作识别模型，得到候选帧序列对应的预测动作；在预测动作为目标动作时，从候选帧序列中确定目标对象的目标动作所对应的目标帧。

步骤 S206，对背景帧和目标帧进行图像分割，得到背景帧中各个物体分别对应的物体区域和目标帧中目标对象对应的目标区域。

其中，图像分割是指将图像划分为若干个不同区域，同一区域内的像素具有相似的特征，不同区域间特征差异明显。

对背景帧进行图像分割，得到背景帧中各个物体分别对应的物体区域。物体对应的区域是指物体在图像中所处区域，用于表示物体的位置。例如，若背景帧为房间 A 的图像，则对背景帧进行图像分割，可以得到房间 A 中各个设施分别所在区域，以便了解房间 A 中设施的分布情况。对目标帧进行图像分割，得到目标帧中目标对象对应的目标区域。目标区域是指目标对象在图像中所处的区域，用于表示目标对象的位置。

可以理解，针对背景帧和针对目标帧可以采用相同的图像分割算法，也可以采用不同的图像分割算法。

具体地，计算机设备可以对背景帧进行图像分割，得到背景帧中各个物体分别对应的物体区域，可以对目标帧进行图像分割，得到目标帧中目标对象对应的目标区域。例如，将背景帧输入图像分割模型，图像分割模型输出背景帧的图像分割结果，图像分割结果可以是在背景帧中通过定位框圈出物体，定位框可以是规则的边框，也可以是不规则的边框，

物体对应的定位框圈出的图像区域即为物体对应的物体区域。又例如，将目标帧输入目标对象识别模型，目标对象识别模型输出目标帧中目标对象对应的目标区域，目标对象识别模型只关心目标对象所在区域。

步骤 S208，基于目标区域和物体区域的位置信息、物体区域对应的物体、目标动作，生成交互信息组，得到目标对象分别和各个物体之间的交互信息组。

其中，一个物体对应一个交互信息组。交互信息组包括背景帧中的物体的相关信息和目标帧中的目标对象的相关信息，用于分析目标对象和物体的交互关系。物体的信息包括物体对应的物体标识、物体区域的位置信息中至少一者。物体标识用于标识物体，例如，物体标识可以是物体名称、物体编号等。物体区域的位置信息用于指示物体区域在图像中所处的位置。目标对象的相关信息包括目标对象的目标动作对应的动作标识、目标区域的位置信息、目标对象对应的对象标识中至少一者。动作标识用于标识动作，例如，动作标识可以是动作名称、动作编号等。目标区域的位置信息用于指示目标区域在图像中所处的位置。对象标识用于标识对象，例如，对象标识可以是对象编号。

具体地，针对背景帧中任意一个物体，计算机设备可以基于该物体对应的物体区域的位置信息、该物体、目标区域对应的位置信息、目标动作，建立交互信息组。例如，交互信息组包括物体区域和目标区域之间的区域重合度、物体的物体标识、目标动作的动作标识；交互信息组包括物体区域的位置信息、目标区域的位置信息、物体的物体标识、目标动作的动作标识；等等。最终，计算机设备可以得到目标帧中的目标对象分别和背景帧中各个物体之间的交互信息组。

步骤 S210，基于各个交互信息组，确定目标对象在目标帧对应的目标对象状态。

其中，对象状态是指对象因执行动作而处于的状态。可以理解，同一行为在不同情境下可以对应不同的对象状态。例如，人在执行倒下动作时，可能处于休息状态，也可以处于跌倒状态。在倒下动作和床、沙发、躺椅等家具产生高度重叠时，人通常是处于休息状态，而倒下动作和地板产生高度重叠时，人通常是处于跌倒状态。

具体地，计算机设备可以基于各个交互信息组分析目标对象和物体之间的交互关系，从而确定目标对象在目标帧对应的目标对象状态。例如，将目标帧和交互信息组输入已训练的交互模型，已训练的交互模型输出目标对象在目标帧对应的目标对象状态。已训练的交互模型是基于训练帧、训练信息组和训练帧对应的训练标签，对待训练的交互模型进行有监督训练得到的。又例如，基于各个交互信息组确定目标区域分别和各个物体区域之间的区域重合度，获取区域重合度大于重合度阈值的物体区域所对应的物体作为目标物体，基于目标物体和目标动作确定目标对象在目标帧对应的目标对象状态。

上述对象状态识别方法中，获取目标视频；获取目标视频对应的背景帧，从目标视频中确定目标对象的目标动作所对应的目标帧；对背景帧和目标帧进行图像分割，得到背景帧中各个物体分别对应的物体区域和目标帧中目标对象对应的目标区域；基于目标区域和物体区域的位置信息、物体区域对应的物体、目标动作，生成交互信息组，得到目标对象分别和各个物体之间的交互信息组；基于各个交互信息组，确定目标对象在目标帧对应的目标对象状态。这样，根据背景帧和目标帧对目标对象在目标帧中的目标动作进行分析，以识别目标对象在目标帧对应的目标对象状态。背景帧可以提供物体的信息，目标帧可以提供目标对象的信息，根据物体的信息和目标对象的信息建立交互信息组，通过交互信息组分析目标对象和物体之间的交互关系，能够有效提高对象状态的识别准确性。在识别对象状态的时候，考虑目标对象和环境的交互关系，能够有效区分因同一动作容易混淆的对象状态，提高对象状态的识别准确性。

在一个实施例中，获取目标视频对应的背景帧，包括：

对目标视频中的视频帧进行目标对象识别，每发现连续目标数量个视频帧均未识别出目标对象，从连续目标数量个视频帧中获取最后一个视频帧作为背景帧。

其中，目标对象识别是指识别图像中是否存在目标对象。例如，若目标对象为人，可以通过各种人体检测算法进行人体识别。可以理解，目标数量可以根据实际需要进行设置。目标数量可以是预先设置的数量。目标数量也可以是随着场景灵活变化的数量。

具体地，计算机设备可以从目标视频中确定背景帧，以提高背景帧的准确性。计算机设备可以对目标视频中的视频帧依次进行目标对象识别，以识别视频帧中是否存在目标对象。计算机设备可以统计连续未识别到目标对象的视频帧的数量，当连续目标数量个视频帧都没有检测到目标对象，则从未检测到目标对象的连续目标数量个视频帧中获取最后一个视频帧作为背景帧，以保障背景帧是最新的背景图像。

可以理解，在确定背景帧后，计算机设备可以将统计结果清空，重新开始统计连续未识别到目标对象的视频帧的数量，当再次发现连续目标数量个视频帧都没有检测到目标对象，则从新发现的连续目标数量个视频帧中获取最后一个视频帧作为新的背景帧，以进一步保障背景帧始终是最新的背景图像。

在一个具体的应用中，参考图 3，将目标视频逐帧进行人体检测，确定是否检测到人体（即 N_{object} 是否等于 0），统计连续未检测到人体的帧数 N_{empty} ，当连续 T_0 帧都没有检测到目标对象，则将当前帧 f_i 的图像更新为背景帧 f_{bg} 。在进行人体检测时，可以采用但不限于 YOLO 系列目标检测器、SSD (Single Shot MultiBox Detector)、CornerNet、CenterNet 等目标检测模型。

上述实施例中，对目标视频中的视频帧进行目标对象识别，每发现连续目标数量个视频帧均未识别出目标对象，从连续目标数量个视频帧中获取最后一个视频帧作为背景帧，能够保障将目标视频中最新、最准确的表示背景的视频帧作为背景帧，这样的背景帧有助于提高对象状态的识别准确性。

在一个实施例中，目标数量和采用的目标对象识别算法对应的识别准确度呈负相关。

其中，目标对象识别算法是指用于进行目标对象识别的算法，用于识别目标对象的算法。目标对象识别算法对应的识别准确度是指目标对象识别算法对应的识别准确性。

具体地，目标数量可以是动态变化的数量，具体可以是根据目标对象识别算法确定的数量。目标数量与目标对象识别算法的识别准确度有关，目标对象识别算法的识别准确度越高，则目标数量越小。可以理解，在连续目标数量个视频帧均未识别出目标对象时，获取其中最后一帧作为背景帧，可以有效避免背景帧中包含目标对象。若用于识别目标对象的目标对象识别算法的识别准确度较差，可以通过增大目标数量来提高背景帧的准确性。在目标对象识别算法的识别准确度较差时，需要连续识别到更多的视频帧中没有目标对象，才确定背景帧。

在一个实施例中，目标对象识别算法是根据目标对象的保护优先级确定的，目标对象的保护优先级越高，采用识别准确度越高的目标对象识别算法。可以理解，目标对象的重要度越高，目标对象的保护优先级越高。例如，人的保护优先级高于猫、狗的保护优先级。又例如，人的年龄越高，人的保护优先级越高；病患的保护优先级高于普通人的保护优先级。

此外，可以先通过通用对象识别算法进行对象识别，将首次识别到的活体对象作为目标对象，再根据目标对象的保护优先级确定目标对象识别算法，基于目标对象识别算法进行后续视频帧的目标对象识别，以提高目标对象识别的准确性。例如，先通过活体识别算法识别视频帧中是否存在活体，若先识别到视频帧中存在人体，则后续通过人体识别算法准确识别后续视频帧中是否存在人体，以提高人体识别准确性。

上述实施例中，目标数量和采用的目标对象识别算法对应的识别准确度呈负相关，在目标对象识别算法的识别准确度较差时，需要连续识别到更多的视频帧中没有目标对象，才确定背景帧，能够保障背景帧的准确性。

在一个实施例中，如图 4 所示，从目标视频中确定目标对象的目标动作所对应的目标

帧, 包括:

步骤 S402, 对目标视频的视频帧进行目标对象识别, 将识别出目标对象的视频帧作为候选帧。

具体地, 计算机设备可以对目标视频的视频帧进行目标对象识别, 将识别出目标对象的视频帧作为候选帧, 再进一步对候选帧进行目标动作识别, 将识别出目标动作的候选帧作为目标帧。

在一个实施例中, 在居家监测场景或看护场景下, 对目标视频的视频帧进行目标对象识别, 将识别出单个目标对象的视频帧作为候选帧。可以理解, 在居家监测场景或看护场景下, 目标对象通常是人, 若目标对象多于一个, 若有人出现异常状态, 则可由在场其他人员施救, 可以无需自动告警。

步骤 S404, 获取目标时间段内的各个候选帧组成候选帧序列。

其中, 目标对象执行动作会引起目标对象的姿态变化, 因此, 通过检测目标对象的姿态变换可以识别相应的动作。而姿态的变换是与时间相关的, 因此, 可以对一定时间内的多个视频帧进行分析来识别是否出现目标动作。例如, 倒下动作代表了对象从站姿到躺姿的变化, 坐下动作代表了对象从站姿到坐姿的变化。目标时间段是待分析的时间段。目标时间段可以根据需要进行设置。目标时间段的时间长度通常大于或等于一个动作的实现时间。例如, 目标时间段可以是 1 秒, 在时间跨度为 1 秒的各个候选帧中确定目标对象是否出现目标动作。候选帧序列包括时间跨度在目标时间段内的多个候选帧。

具体地, 计算机设备可以将在一个目标时间段内的各个候选帧按照时间顺序组成候选帧序列, 进而对候选帧序列进行目标动作识别, 从存在目标动作的候选帧序列中确定目标帧。

步骤 S406, 对候选帧序列中的候选帧进行目标对象骨架点识别, 得到候选帧对应的骨架点数据, 组成候选帧序列对应的骨架点数据序列。

其中, 目标对象骨架点识别是指识别目标对象的骨架点。例如, 若目标对象为人, 可以通过各种人体骨架点识别算法进行人体骨架点识别。对一个视频帧进行目标对象骨架点识别, 得到该视频帧对应的骨架点数据。候选帧序列对应的骨架点数据序列包括候选帧序列中各个候选帧分别对应的骨架点数据。骨架点数据包括骨架点的类型和位置。

具体地, 针对候选帧序列中任意一个候选帧, 计算机设备可以对候选帧进行目标对象骨架点识别, 得到候选帧对应的骨架点数据。计算机设备可以得到候选帧序列中各个候选帧分别对应的骨架点数据, 将各个候选帧分别对应的骨架点数据组成候选帧序列对应的骨架点数据序列。例如, 可以将各个骨架点数据直接组合得到骨架点数据序列, 也可以将属于同一类型的各个骨架点数据按照时间顺序排列, 得到各个类型分别对应的骨架点时序数据, 将各个骨架点时序数据组合得到骨架点数据序列, 骨架点数据对应的时间可以是骨架点数据所属的候选帧对应的时间。骨架点数据序列可以反映候选帧序列的动作特征, 有助于识别候选帧序列对应的动作。

步骤 S408, 将骨架点数据序列输入多帧动作识别模型, 得到候选帧序列对应的预测动作。

其中, 多帧动作识别模型是人工智能模型, 用于根据多帧图像识别动作。多帧动作识别模型的输入数据包括图像序列对应的骨架点数据序列, 多帧动作识别模型的输出数据包括预测动作。预测动作是模型识别出的动作。

具体地, 计算机设备可以通过人工智能模型来识别目标动作。计算机设备获取多帧动作识别模型, 将候选帧序列对应的骨架点数据序列输入多帧动作识别模型, 多帧动作识别模型对输入数据进行数据处理, 输出候选帧序列对应的预测动作。

可以理解, 多帧动作识别模型可以是基于训练图像序列对应的骨架点数据序列和动作训练标签进行有监督训练得到的。训练图像序列是已知动作的图像序列。训练图像序列对

应的动作训练标签是训练图像序列反映出的已知动作。

步骤 S410, 在预测动作为目标动作时, 从候选帧序列中确定目标对象的目标动作所对应的目标帧。

具体地, 在多帧动作识别模型输出的预测动作为目标动作时, 计算机设备可以从候选帧序列中确定目标对象的目标动作所对应的目标帧。例如, 从候选帧序列中获取最后一帧作为目标帧。又例如, 从候选帧序列中随机选取一帧作为目标帧。

上述实施例中, 对目标视频的视频帧进行目标对象识别, 将识别出目标对象的视频帧作为候选帧, 获取目标时间段内的各个候选帧组成候选帧序列, 对候选帧序列进行目标动作识别, 从识别出目标动作的候选帧序列中确定目标帧。这样, 综合考虑一定时间内的候选帧来识别是否出现目标动作, 能够有效提高目标动作的识别准确性。对候选帧序列中的候选帧进行目标对象骨架点识别, 得到候选帧对应的骨架点数据, 组成候选帧序列对应的骨架点数据序列, 将骨架点数据序列输入多帧动作识别模型, 得到候选帧序列对应的预测动作, 在预测动作为目标动作时, 从候选帧序列中确定目标对象的目标动作所对应的目标帧。这样, 先提取候选帧序列对应的骨架点数据序列, 骨架点数据序列可以反映动作特征, 通过多帧动作识别模型对骨架点数据序列处理来识别候选帧序列对应的动作, 多帧动作识别模型具有强大的数据处理能力, 能够进一步提高目标动作的识别准确性。

在一个实施例中, 从候选帧序列中确定目标对象的目标动作所对应的目标帧, 包括:

将候选帧序列中的各个候选帧分别输入已训练的单帧动作识别模型, 得到各个候选帧分别对应的动作识别结果; 动作识别结果包括候选帧中目标对象的数量和动作; 获取动作识别结果包括单个目标对象和目标动作的候选帧, 作为目标对象的目标动作所对应的目标帧。

其中, 单帧动作识别模型是人工智能模型, 用于根据单帧图像识别动作。单帧动作识别模型的输入数据包括单帧图像, 单帧动作识别模型的输出数据包括单帧图像中目标对象的数量和动作。

具体地, 在进行目标动作识别时, 计算机设备可以先通过多帧动作识别模型进行预识别, 初筛掉明显与目标动作无关的视频帧, 在预识别到出现目标动作时, 再通过单帧动作识别模型进行进一步识别, 最终得到准确的目标帧。

计算机设备将候选帧序列对应的骨架点数据序列输入多帧动作识别模型, 得到候选帧序列对应的预测动作。可以理解, 受算力资源、光照、角度等多种因素的影响, 通过多帧动作识别模型得到的初筛结果中可能有部分的动作误识别结果, 甚至部分数据中含有多个目标对象。因此, 进一步通过单帧动作识别模型进行目标动作识别。在候选帧序列对应的预测动作为目标动作时, 再将候选帧序列中的各个候选帧分别输入已训练的单帧动作识别模型, 得到各个候选帧分别对应的动作识别结果。候选帧对应的动作识别结果包括候选帧中目标对象的数量和动作。例如, 动作识别结果可以是在候选帧中通过定位框圈出目标对象和标注目标对象的动作。计算机设备从候选帧序列中, 获取动作识别结果包括单个目标对象和目标动作的候选帧, 作为目标对象的目标动作所对应的目标帧。

上述实施例中, 单帧动作识别模型不仅能够识别目标对象的数量, 还能够识别目标动作, 同时实现了过滤多对象场景和避免动作误识别的功能, 能够提高有效目标帧的识别准确性。

在一个具体的应用中, 参考图 5, 获取时间跨度在 1s 内的各个视频帧, 将各个视频帧按照时间顺序排列输入多帧动作识别模型, 若多帧动作识别模型输出的识别结果为目标动作, 则将各个视频帧分别输入单帧动作识别模型。若多帧动作识别模型输出的识别结果不是目标动作, 则可以认为这些视频帧中无目标帧。单帧动作识别模型可以识别视频帧中目标对象的数量和动作, 若单帧动作识别模型针对某个视频帧输出的识别结果为单个目标对象和目标动作, 则将该视频帧作为目标帧。

在一个实施例中，对象状态识别方法还包括：

获取第一训练图像和第一训练图像对应的目标对象位置训练标签、目标对象动作训练标签；将第一训练图像输入待训练的单帧动作识别模型，得到第一训练图像对应的目标对象位置预测标签和目标对象动作预测标签；基于目标对象位置训练标签和目标对象位置预测标签之间的差异、目标对象动作训练标签和目标对象动作预测标签之间的差异，调整待训练的单帧动作识别模型的模型参数，直至满足第一收敛条件，得到已训练的单帧动作识别模型。

其中，第一训练图像是用于训练单帧动作识别模型的图像。第一训练图像对应的目标对象位置训练标签用于指示目标对象在第一训练图像中准确的位置信息。第一训练图像对应的目标对象动作训练标签用于指示目标对象在第一训练图像中准确的动作信息。可以理解，第一训练图像中可以没有目标对象，可以只有单个目标对象，也可以有多个目标对象。例如，若目标对象为人，则第一训练图像可以没有人，可以只有一个人，也可以有多个人，第一训练图像中的多个人可以是同一类型的人，也可以是不同类型的人；若目标对象为狗，则第一训练图像可以没有狗，只有一只狗，也可以有多只狗；等等。第一训练图像中的一个目标对象存在对应的目标对象动作训练标签。目标对象动作训练标签可以是目标动作，也可以是其他动作。例如，目标对象动作训练标签可以是坐姿、站姿、躺姿等。

将第一训练图像输入待训练的单帧动作识别模型，得到第一训练图像对应的目标对象位置预测标签和目标对象动作预测标签。目标对象位置预测标签用于指示预测得到的目标对象在第一训练图像中的位置信息。目标对象动作预测标签用于指示预测得到的目标对象在第一训练图像中的动作信息。

第一收敛条件是判断单帧动作识别模型是否达到收敛的条件。第一收敛条件可以根据实际需要进行设置。第一收敛条件可以是模型损失小于损失阈值、模型迭代次数大于迭代次数阈值、模型损失变化率小于损失变化率阈值等至少一者。

具体地，计算机设备可以获取第一训练图像和第一训练图像对应的目标对象位置训练标签、目标对象动作训练标签，基于第一训练图像和第一训练图像对应的训练标签，对单帧动作识别模型进行有监督训练。

计算机设备可以将第一训练图像输入待训练的单帧动作识别模型，经过单帧动作识别模型的数据处理，单帧动作识别模型输出第一训练图像对应的目标对象位置预测标签和目标对象动作预测标签。进而，计算机设备基于目标对象位置训练标签和目标对象位置预测标签之间的差异、目标对象动作训练标签和目标对象动作预测标签之间的差异计算模型损失，将模型损失进行反向传播调整待训练的单帧动作识别模型的模型参数，直至满足第一收敛条件，得到已训练的单帧动作识别模型。

可以理解，第一训练图像可以有多个。每一轮可以获取多个第一训练图像参与训练。在一轮中，基于同一第一训练图像对应的目标对象位置训练标签和目标对象位置预测标签之间的差异、目标对象动作训练标签和目标对象动作预测标签之间的差异计算模型子损失，得到各个第一训练图像分别对应的模型子损失，基于各个训练图像分别对应的模型子损失得到模型损失，基于模型损失调整待训练的单帧动作识别模型，将调整后的单帧动作识别模型作为下一轮中待训练的单帧动作识别模型。

上述实施例中，基于第一训练图像和第一训练图像对应的目标对象位置训练标签、目标对象动作训练标签，对单帧动作识别模型进行有监督训练，能够使得完成训练的单帧动作识别模型具备目标对象识别能力和动作识别能力。

在一个具体的应用中，参考图6，以目标对象为人体和目标动作为倒下动作为例说明确定目标对象状态的过程。从目标视频中确定背景帧和目标帧，分别对背景帧和目标帧进行图像语义分割，得到背景帧中各个物体分别对应的物体区域和目标帧中人体对应的人体区域。根据物体对应的物体区域和人体区域建立交互信息组。一个交互信息组包括倒下动

作的动作标识（即图 6 中的 P 倒下）、背景帧中物体的物体标识（即图 6 中的地板 O1 或沙发 O2）、物体区域和人体区域之间的重叠面积（即图 6 中的 E1 或 E2）。根据各个交互信息组确定目标对象在目标帧对应的目标对象状态。

可以采用但不限于 Light-Weight RefineNet, SAM (Segment Anything Model) 等深度神经网络模型进行图像语义分割。

在一个实施例中，如图 7 所示，基于各个交互信息组，确定目标对象在目标帧对应的目标对象状态，包括：

步骤 S702，获取目标对象在目标动作对应的各个参考对象状态，获取各个参考对象状态分别对应的参考物体。

其中，目标对象在目标动作可以有至少一个参考对象状态。参考对象状态是目标对象在目标动作可以处于的对象状态。例如，针对倒下动作，目标对象在倒下动作对应的参考对象状态可以是休息状态，也可以是跌倒状态；针对坐下动作，目标对象在坐下动作对应的参考对象状态可以是休息状态，也可以是社交状态；针对跑步动作，目标对象在跑步动作对应的参考对象状态可以是避险状态，也可以是锻炼状态；等等。

参考对象状态存在对应的参考物体。参考物体是指与参考对象状态有关的物体。例如，目标对象处于休息状态，参考物体可以是床、沙发、椅子等物体；目标对象处于跌倒状态，参考物体可以是地板、台阶等物体；目标对象处于锻炼状态，参考物体可以是跑步机、沙袋等物体；等等。

具体地，计算机设备可以获取目标对象在目标动作对应的各个参考对象状态，从各个参考对象状态中确定目标对象在目标帧对应的目标对象状态。

步骤 S704，将背景帧内的参考物体作为目标物体，获取目标物体对应的参考对象状态所对应的参考重合度阈值作为目标重合度阈值。

其中，参考对象状态存在对应的参考重合度阈值。参考重合度阈值是指处于参考对象状态的目标对象与参考物体之间的重合度需要满足的条件。若目标对象在目标动作时与参考物体之间的重合度大于参考重合度阈值，则确定目标对象在目标动作处于该参考物体对应的参考对象状态。例如，针对倒下动作，若目标对象和床、沙发、椅子等物体的重合度大于一定阈值，则确定目标对象处于休息状态，若目标对象和地板、台阶等物体的重合度大于一定阈值，则确定目标对象处于跌倒状态；针对跑步动作，若目标对象和跑步机、跑道等物体的重合度大于一定阈值，则确定目标对象处于锻炼状态。参考对象状态对应的参考重合度阈值可以是对大量已知目标对象处于参考对象状态的已知图像进行统计得到的。例如，计算已知图像中目标对象对应的目标区域和参考物体对应的物体区域之间的区域重合度，计算各个已知图像分别对应的区域重合度的平均值作为参考重合度阈值；获取各个已知图像分别对应的区域重合度的中位值作为参考重合度阈值；等等。

目标物体是指背景帧中包含的参考物体。例如，若目标动作作为倒下动作，则目标物体可以是背景帧中的床、沙发、椅子、地板、台阶等物体。目标重合度阈值是指目标物体对应的参考对象状态所对应的参考重合度阈值。

具体地，计算机设备可以获取各个参考对象状态分别对应的参考物体，将背景帧中存在的参考物体作为目标物体，从目标物体对应的参考对象状态中确定目标对象在目标帧对应的目标对象状态。

步骤 S706，针对目标物体对应的交互信息组，当目标物体对应的物体区域和目标区域之间的区域重合度大于目标重合度阈值时，将目标物体对应的参考对象状态，作为目标对象在目标帧对应的目标对象状态。

其中，物体区域和目标区域之间的区域重合度是指物体区域和目标区域之间的重合程度，具体可以是重叠面积、重叠比例等。

具体地，计算机设备获取目标物体对应的参考对象状态所对应的参考重合度阈值作为

目标重合度阈值,根据目标物体对应的交互信息组确定目标物体对应的物体区域和目标对象对应的目标区域之间的区域重合度,若区域重合度大于目标重合度阈值,则将目标物体对应的参考对象状态作为目标对象在目标帧对应的目标对象状态。

可以理解,若针对一个参考对象状态,参考物体有多个,目标区域和任意一个参考物体对应的物体区域之间的区域重合度大于相应的参考重合度阈值,则将参考物体对应的参考对象状态作为目标对象状态。

可以理解,若区域重合度小于或等于目标重合度阈值,则可以进一步通过其他方式确定目标对象在目标帧对应的目标对象状态。例如,通过已训练的交互模型确定目标对象在目标帧对应的目标对象状态。将目标帧和各个交互信息组输入已训练的交互模型,已训练的交互模型输出目标对象在目标帧对应的目标对象状态。这样,先通过参考对象状态的相关信息确定是否命中参考对象状态,可以提高对象状态识别效率,若未命中参考对象状态,则再通过交互模型识别对象状态,能够提高对象状态识别准确性,弥补参考对象状态的不足。又例如,将目标帧发送至管理终端,管理终端向管理员展示目标帧,以便管理员根据目标帧确定目标对象在目标帧对应的目标对象状态。若区域重合度小于或等于目标重合度阈值,也可以确定目标对象在目标帧对应的目标对象状态为未知状态。

上述实施例中,获取目标对象在目标动作对应的各个参考对象状态,参考对象状态存在对应的参考物体和参考重合度阈值,通过确定背景帧中包含的参考物体对应的物体区域和目标对象在目标帧中对应的目标区域之间的区域重合度,将区域重合度与相应的参考重合度阈值进行比较,可以快速确定目标对象在目标帧对应的目标对象状态。并且,各个参考对象状态存在各自对应的参考重合度阈值,有助于进一步提高对象状态的识别准确性。

在一个实施例中,如图8所示,基于各个交互信息组,确定目标对象在目标帧对应的目标对象状态,包括:

步骤S802,将目标帧和各个交互信息组输入已训练的交互模型。

其中,交互模型是人工智能模型,用于根据图像和交互信息组确定对象状态。交互模型的输入数据包括对象的某个动作对应的图像、根据该图像和对应的背景图像确定的交互信息组,输出数据包括模型识别出来的对象状态。

具体地,计算机设备可以通过人工智能模型来识别目标对象状态。计算机设备获取已训练的交互模型,将目标对象的目标动作对应的目标帧、根据背景帧和目标帧确定的各个交互信息组输入已训练的交互模型,经过交互模型的数据处理,交互模型输出目标对象在目标帧对应的目标对象状态。

步骤S804,基于各个交互信息组在目标帧中分别确定重叠区域。

步骤S806,提取重叠区域的图像特征,得到第一交互特征,基于重叠区域对应的交互信息组中的物体和目标动作,得到第二交互特征,提取目标帧的图像特征,得到第三交互特征。

步骤S808,基于第一交互特征、第二交互特征和第三交互特征,得到综合交互特征。

步骤S810,基于综合交互特征输出目标对象在目标帧对应的目标对象状态。

其中,重叠区域是指背景帧中物体对应的物体区域和目标帧中目标对象对应的目标区域之间的重叠部分。第一交互特征是重叠区域的图像特征,包含重叠区域的信息。第二交互特征是根据重叠区域对应的交互信息组中的物体和目标动作确定的特征,包含重叠区域对应的物体和目标动作的信息。第三交互特征是目标帧的图像特征,包含目标帧的信息。综合交互特征是基于第一交互特征、第二交互特征和第三交互特征得到的特征,包含三种交互特征中比较重要的信息。

具体地,计算机设备将目标帧和各个交互信息组输入已训练的交互模型。在交互模型中,基于各个交互信息组在目标帧中分别确定重叠区域,得到至少一个重叠区域,提取重叠区域的图像特征,得到第一交互特征,提取目标帧的图像特征,得到第三交互特征。例

如，可以通过卷积层提取图像特征。在交互模型中，基于重叠区域对应的交互信息组中的物体和目标动作生成第二交互特征。例如，提取物体的特征，提取目标动作的特征，将提取到的特征进行拼接，得到第二交互特征。在交互模型中，基于第一交互特征、第二交互特征和第三交互特征，得到综合交互特征。例如，融合第一交互特征、第二交互特征和第三交互特征，得到综合交互特征，可以通过全连接层进行特征融合；将第一交互特征、第二交互特征和第三交互特征两两分别进行注意力处理，得到多个注意力特征，融合多个注意力特征得到综合交互特征，可以通过注意力机制进行注意力处理，通过注意力处理可以提取特征之间彼此相互影响的部分；等等。在交互模型中，基于综合交互特征确定目标对象在目标帧对应的目标对象状态。交互模型最终输出目标对象在目标帧对应的目标对象状态。

上述实施例中，将目标帧和各个交互信息组输入已训练的交互模型，交互模型具有强大的数据处理能力，能够快速输出准确的目标对象状态，并且交互信息组可以提供目标对象和物体之间的交互信息，帮助交互模型理解目标帧，从而交互模型输出的目标对象状态进一步具有更高的准确性。

在一个实施例中，对象状态识别方法还包括：

获取训练对象的训练动作对应的第二训练图像和第二训练图像对应的训练对象状态、训练信息组；训练对象状态是训练对象在第二训练图像对应的对象状态，训练信息组是基于第二训练图像中训练对象所处区域的位置信息、第二训练图像对应的背景图像中物体所处区域的位置信息和物体信息、训练动作确定的；将第二训练图像和训练信息组输入待训练的交互模型，得到第二训练图像对应的预测对象状态；基于训练对象状态和预测对象状态之间的差异，调整待训练的交互模型的模型参数，直至满足第二收敛条件，得到已训练的交互模型。

其中，第二训练图像是用于训练交互模型的图像。第二训练图像是训练对象的训练动作对应的图像。第二训练图像对应的训练对象状态用于指示训练对象在训练图像对应的准确的对象状态。训练信息组是根据第二训练图像的图像分割结果和第二训练图像对应的背景图像的图像分割结果确定的交互信息组。训练信息组的确定方式可以参考前述交互信息组的确定方式。

将第二训练图像和第二训练图像对应的训练信息组输入已训练的交互模型，得到第二训练图像对应的预测对象状态。预测对象状态是模型预测出来的训练对象在第二训练图像对应的对象状态。

第二收敛条件是判断交互模型是否达到收敛的条件。第二收敛条件可以根据实际需要进行设置。第二收敛条件可以是模型损失小于损失阈值、模型迭代次数大于迭代次数阈值、模型损失变化率小于损失变化率阈值等至少一者。

可以理解，训练对象和目标对象可以是相同的对象，也可以是不同的对象。训练动作和目标动作可以是相同的动作，也可以是不同的动作。第一训练图像和第二训练图像可以是相同的训练图像，也可以是不同的训练图像。

具体地，计算机设备可以获取训练对象的训练动作对应的第二训练图像和第二训练图像对应的训练对象状态、训练信息组，基于这些数据对交互模型进行有监督训练。计算机设备将第二训练图像和训练信息组输入待训练的交互模型，经过交互模型的数据处理，交互模型输出第二训练图像对应的预测对象状态，进而基于训练对象状态和预测对象状态之间的差异计算模型损失，将模型损失反向传播调整待训练的交互模型的模型参数，直至满足第二收敛条件，得到已训练的交互模型。

可以理解，第二训练图像可以有多个。每一轮可以获取多个第二训练图像参与训练。

上述实施例中，基于第二训练图像和第二训练图像对应的训练对象状态、训练信息组，对交互模型进行有监督训练，能够使得完成训练的交互模型具备对象状态识别能力。

在一个实施例中，目标视频是对目标对象的生活环境实时采集图像得到的视频流。对象状态识别方法还包括：

在目标对象状态为异常对象状态时，向生活环境对应的管理终端发送针对异常对象状态的告警消息。

其中，目标视频是对目标对象的生活环境实时采集图像得到的视频流。例如，目标视频可以是在目标对象的家中，通过安装的监测摄像头实时采集到的监测视频；目标视频可以是在目标对象所处的养老机构或医院内，通过安装的监测摄像头实时采集到的监测视频；目标视频可以是在目标对象经常去的场所内，通过安装的监测摄像头实时采集到的监测视频。

异常对象状态是指异常的对象状态。异常对象状态可以根据需要进行设置。例如，可以将有危险的对象状态作为异常对象状态；可以将与日常对象状态不同的对象状态作为异常对象状态；等等。例如，目标对象的生活比较规律，目标对象在一天中各个时间段的对象状态基本是稳定的、比较有规律的，若目标对象在某个时间段的当前对象状态与往常在该时间段的历史对象状态不同，则可以确定目标对象在该时间段的当前对象状态为异常对象状态。

生活环境对应的管理终端是针对目标对象的管理终端，用于管理目标对象在生活环境的对象状态。例如，若目标对象的生活环境为目标对象的居家环境，则管理终端可以是目标对象的家人、监护人等；若目标对象的生活环境为目标对象在养老机构或医院内所处的生活环境，则管理终端可以是养老机构或医院的工作人员，也可以是目标对象的家人、监护人等。

针对异常对象状态的告警消息用于告警目标对象处于异常对象状态。告警消息可以是文本、图像、视频、音频等至少一种媒体形式的消息。告警消息可以通过短信、应用程序、电话等各种方式展示。

具体地，目标视频是对目标对象的生活环境实时采集图像得到的视频流，也就是，目标视频是在不断生成的视频流，计算机设备可以边获取目标视频边从目标视频中解码出视频帧进行对象状态识别。若出现目标对象的目标动作对应的目标帧，则根据目标帧和背景帧确定目标对象在目标帧对应的目标对象状态，进一步识别目标对象状态是否为异常对象状态。若目标对象状态为异常对象状态，则计算机设备可以向生活环境对应的管理终端发送针对异常对象状态的告警消息，从而管理终端可以及时获知目标对象处于异常对象状态，及时采取相应的保护措施。

上述实施例中，在目标对象状态为异常对象状态时，向生活环境对应的管理终端发送针对异常对象状态的告警消息，有助于及时帮助目标对象恢复正常对象状态，保障目标对象的安全。

在一个实施例中，目标视频是通过设置在目标对象的生活环境中的目标设备对生活环境实时采集图像得到的视频流。对象状态识别方法还包括：

在目标对象状态为异常对象状态时，向目标设备发送针对异常对象状态的告警消息；目标设备用于展示针对异常对象状态的协助需求确认消息，并响应于针对协助需求确认消息的同意接受协助事件，确定针对目标对象的目标协助方式，向目标协助方式对应的协作设备发送针对目标对象的协作消息。

其中，协助需求确认消息是用于确定目标对象是否需要协助的消息。协助需求确认消息也可以称为协助需求询问消息。若目标对象需要帮助，目标对象和目标设备进行交互，以使目标设备向其他设备发送针对目标对象的协助消息。例如，协助需求确认消息可以是询问目标对象是否需要帮助的语音消息，目标对象可以与目标设备对话来反馈是否需要帮助。又例如，协助需求确认消息可以是询问目标对象是否需要帮助的通知消息，目标对象可以在目标设备上进行操作来反馈是否需要帮助。

同意接受协助事件是表示同意接受协助、确认需要协助的事件。同意接受协助事件可以是基于用户的触发操作生成的。例如，响应于针对协助需求确认消息的触发操作，触发同意接受协助事件；显示同意控件，响应于针对同意控件的触发操作，触发同意接受协助事件。同意接受协助事件也可以是基于用户的隔空操作生成的。隔空操作是以非接触式的方式向目标设备传达信息的操作。例如，协助需求确认消息可以是询问目标对象是否需要帮助的语音消息，目标对象可以发出表示同意接收协助的语音消息，若目标设备接收到目标对象发出的语音消息，则生成同意接受协助事件。当然，同意接受协助事件也可以是自动触发的。例如，若用户在一定时间内没有拒绝接受协助，则自动触发同意接受协助事件。

针对目标对象的目标协助方式是在目标对象处于异常对象状态时，对目标对象进行帮助的具体方式。目标协助方式可以是预设的协作方式，例如，目标协助方式包括紧急联系人协助方式、专业机构协助方式、机器人协助方式中任意一者。又例如，目标协助方式可以是候选协助方式中选择的协助方式。协助需求确认消息可以进一步包括多种候选协助方式以供目标对象选择，若目标对象需要协助，目标对象可以对候选协助方式进行选择，将目标对象选择的候选协助方式作为目标协助方式。目标对象可以通过接触式或非接触式的方式从候选协助方式中选择目标协助方式。当然，目标协助方式也可以是自定义的协助方式。目标对象可以向目标设备提供自定义的协助方式。例如，目标对象可以向目标设备提供好友的联系方式，以使目标设备基于联系方式向好友的好友终端发送针对目标对象的协助消息。

目标设备向目标协助方式对应的协助设备发送针对目标对象的协助消息。针对目标对象的协助消息包括目标对象的位置信息和目标对象状态。例如，候选协助方式包括紧急联系人协助方式、专业机构协助方式、机器人协助方式。若目标对象选择紧急联系人协助方式，则协助设备可以是目标对象的紧急联系人对应的通讯设备。若目标对象选择专业机构协助方式，则协助设备可以是专业结构对应的通讯设备。若目标对象选择机器人协助方式，则协助设备是与生活环境的距离在一定范围内的机器人，向机器人发送针对目标对象的协助消息，从而机器人可以根据目标对象的位置信息查找到目标对象，根据目标对象的目标对象状态采取相应的措施帮助目标对象。

这样，在监测到目标对象出现异常的时候，及时通过协助需求确认消息询问目标对象是否需要帮助，能够有效帮助目标对象尽快恢复正常。

在一个实施例中，目标视频是对目标对象的生活环境实时采集图像得到的视频流。对象状态识别方法还包括：

在目标对象状态为异常对象状态时，从目标视频中获取目标帧的相邻帧集合；对相邻帧集合中的相邻帧进行目标对象识别，将识别出目标对象的相邻帧作为辅助帧；从辅助帧中确定辅助对象；辅助对象和目标帧中的目标对象是属于同一类型的不同对象；向辅助对象对应的对象终端发送针对异常对象状态的告警消息。

其中，目标帧的相邻帧集合是指在目标视频中与目标帧之间的采集时间距离小于时间距离阈值的视频帧。也就是，目标帧的相邻帧集合包括目标视频中在目标帧前后的若干个视频帧。

辅助帧是从目标帧的相邻帧集合中选取的视频帧。辅助对象是指与出现异常对象状态的目标对象不同的、且辅助帧中存在的其他目标对象。例如，若目标对象为人，目标帧中出现异常对象状态为张三，辅助对象可以是辅助帧中出现的李四。可以理解，在检测到目标帧中的目标对象出现异常对象状态时，辅助对象未检测到出现异常对象状态。

辅助对象对应的对象终端是指辅助对象对应的通讯设备。例如，辅助对象对应的对象终端可以是辅助对象的手机、智能手表等。

具体地，在目标对象在目标帧对应的目标对象状态为异常对象状态时，计算机设备可以进一步从目标视频中确定辅助对象，向辅助对象对应的对象终端发送针对异常对象状态

的告警消息，以便辅助对象及时采取措施帮助目标对象。

在目标对象状态为异常对象状态时，从目标视频中获取目标帧的相邻帧集合，例如，获取距离目标帧 5 分钟内的前后视频帧组成相邻帧集合。进而，对相邻帧集合中的相邻帧进行目标对象识别，将识别出目标对象的相邻帧作为辅助帧，从辅助帧中确定辅助对象。例如，将识别出目标对象且距离目标帧最近的相邻帧作为辅助帧，从辅助帧中选取与目标帧中出现异常对象状态的目标对象不同的至少一个其他目标对象作为辅助对象。进而，向辅助对象对应的对象终端发送针对异常对象状态的告警消息。

举例说明，若目标对象为人，在目标帧中的用户 A 出现异常对象状态时，若目标帧的相邻帧集合中出现了用户 B，则将用户 B 作为辅助对象，向用户 B 对应的通讯设备发送针对用户 A 的告警消息，以便用户 B 通过通讯设备及时获知用户 A 出现异常。

上述实施例，在目标对象状态为异常对象状态时，从目标帧的相邻帧集合中确定辅助对象，向辅助对象对应的对象终端发送针对异常对象状态的告警消息，辅助对象通常距离目标对象比较近，有助于快速帮助目标对象恢复正常对象状态，保障目标对象的安全。

在一个具体的实施例中，本申请方法可以应用于居家场景中的跌倒检测，有效提高跌倒识别准确性。参考图 9 和图 10，服务器获取居家场景中的摄像头采集的视频流并解码成图像，先对图像进行人形检测（即人体检测），检测到出现目标人物即通过动作识别模块识别图像帧中目标人物执行的动作，同时根据各个图像的人体检测结果也可以识别视频流中的背景帧（即环境上下文检测）。可以理解，将识别到目标人物出现倒下动作的图像作为前景帧（即目标帧）。对前景帧和背景帧分别进行图像语义分割，前景帧中主要关注目标人物的像素级语义分割区域，背景帧中则关注和目标人物区域重叠或相连的物体的分割结果。根据图像语义分割结果计算目标人物和周边环境各物体的重合率，将重合率和倒下动作、物体共同构成交互三元组（即交互信息组）。根据交互三元组确定目标人物的跌倒识别结果。可以将交互三元组输入交互模型，交互模型输出最终的跌倒识别结果。

若跌倒识别结果为目标人物出现跌倒，则服务器可以向前端返回跌倒识别结果，前端可以根据跌倒识别结果进行进一步的图像显示、异常行为告警等操作。例如，前端可以显示前景帧，在前景帧中圈出跌倒的目标人物。

可以理解，目标人物可以根据需要进行设置。例如，目标人物可以是老人、小孩。

可以理解，随着老年人口的增加，需要更先进的家用设备来监测老人的活动，以保证老人的安全。通过本申请方法可以在卧室、客厅等区域中，当类似床、沙发等和倒下动作强关联的家具存在时，更有效区分老人在不同情境下的倒下动作，区分老人是处于休息状态还是跌倒状态。

当然，本申请方法不仅应用于居家安全守护，还可以应用于医疗/养老机构中的病患辅助看护等场景。本申请方法也可以用于检测除跌倒以外的其他状态。

在一个具体的实施例中，本申请方法可以应用于云边协同系统，通过云边协同系统进行跌倒检测。云边协同系统是云计算和边缘计算的融合，旨在实现数据的高效处理和传输。云边协同系统包括云端设备、边缘设备。边缘设备，即智能摄像设备，用于捕捉实时视频后解码为图像，并在本地设备中完成初步分析，包括人体检测、人体骨骼点检测、基于骨骼点时序数据的动作识别等，得到交互三元组，根据交互三元组确定目标人物是否跌倒的初步判断结果。例如，若交互三元组中的重合率大于相应的阈值，则确定初步判断结果为目标人物出现跌倒。当初步判断结果为目标人物出现跌倒时，将本地处理得到的结果、当前帧（即前景帧）、交互三元组上传至云端设备。云端设备根据边缘设备发送过来的数据进行二次分析，得到最终判断结果。例如，将前景帧和交互三元组输入交互模型，交互模型输出最终判断结果。在最终判断结果为目标人物出现跌倒时，云端设备可根据业务逻辑将最终判断结果（比如跌倒告警及相关画面）推送至用户终端或监测中心，云端设备也可以自动拨打目标人物的紧急联系人电话。

这样，通过边缘设备和云端设备的协同工作，实现高效、准确的跌倒检测。边缘设备进行初步判断，实现数据过滤，在初步判断结果为目标人物出现跌倒时，才将相关数据发送至云端设备，云端设备进行二次判断，得到最终判断结果。

应该理解的是，虽然如上所述的各实施例所涉及的流程图中的各个步骤按照箭头的指示依次显示，但是这些步骤并不是必然按照箭头指示的顺序依次执行。除非本文中有明确的说明，这些步骤的执行并没有严格的顺序限制，这些步骤可以以其它的顺序执行。而且，如上所述的各实施例所涉及的流程图中的至少一部分步骤可以包括多个步骤或者多个阶段，这些步骤或者阶段并不必然是在同一时刻执行完成，而是可以在不同的时刻执行，这些步骤或者阶段的执行顺序也不必然是依次进行，而是可以与其它步骤或者其它步骤中的步骤或者阶段的至少一部分轮流或者交替地执行。

基于同样的发明构思，本申请实施例还提供了一种用于实现上述所涉及的对象状态识别方法的对象状态识别装置。该装置所提供的解决问题的实现方案与上述方法中所记载的实现方案相似，故下面所提供的一个或多个对象状态识别装置实施例中的具体限定可以参见上文中对于对象状态识别方法的限定，在此不再赘述。

在一个实施例中，如图 11 所示，提供了一种对象状态识别装置，包括：视频获取模块 1102、图像确定模块 1104、图像分割模块 1106、信息组生成模块 1108 和状态识别模块 1110，其中：

视频获取模块 1102，用于获取目标视频。

图像确定模块 1104，用于获取目标视频对应的背景帧，从目标视频中确定目标对象的目标动作所对应的目标帧。

图像分割模块 1106，用于对背景帧和目标帧进行图像分割，得到背景帧中各个物体分别对应的物体区域和目标帧中目标对象对应的目标区域。

信息组生成模块 1108，用于基于目标区域和物体区域的位置信息、物体区域对应的物体、目标动作，生成交互信息组，得到目标对象分别和各个物体之间的交互信息组。

状态识别模块 1110，用于基于各个交互信息组，确定目标对象在目标帧对应的目标对象状态。

在一个实施例中，图像确定模块 1104 还用于：

对目标视频中的视频帧进行目标对象识别，每发现连续目标数量个视频帧均未识别出目标对象，从连续目标数量个视频帧中获取最后一个视频帧作为背景帧。

在一个实施例中，图像确定模块 1104 还用于：

基于通用对象识别算法，对目标视频中的视频帧进行对象识别，将首次识别到的活体对象作为目标对象；

基于目标对象的保护优先级确定目标对象识别算法；

基于目标对象识别算法，对目标视频中的视频帧进行目标对象识别。

在一个实施例中，目标数量和采用的目标对象识别算法对应的识别准确度呈负相关。

在一个实施例中，图像确定模块 1104 还用于：

对目标视频的视频帧进行目标对象识别，将识别出目标对象的视频帧作为候选帧；

获取目标时间段内的各个候选帧组成候选帧序列；

对候选帧序列中的候选帧进行目标对象骨架点识别，得到候选帧对应的骨架点数据，组成候选帧序列对应的骨架点数据序列；

将骨架点数据序列输入多帧动作识别模型，得到候选帧序列对应的预测动作；

在预测动作为目标动作时，从候选帧序列中确定目标对象的目标动作所对应的目标帧。

在一个实施例中，图像确定模块 1104 还用于：

将候选帧序列中的各个候选帧分别输入已训练的单帧动作识别模型，得到各个候选帧分别对应的动作识别结果；动作识别结果包括候选帧中目标对象的数量和动作；

获取动作识别结果包括单个目标对象和目标动作的候选帧，作为目标对象的目标动作所对应的目标帧。

在一个实施例中，对象状态识别装置还用于：

获取第一训练图像和第一训练图像对应的目标对象位置训练标签、目标对象动作训练标签；

将第一训练图像输入待训练的单帧动作识别模型，得到第一训练图像对应的目标对象位置预测标签和目标对象动作预测标签；

基于目标对象位置训练标签和目标对象位置预测标签之间的差异、目标对象动作训练标签和目标对象动作预测标签之间的差异，调整待训练的单帧动作识别模型的模型参数，直至满足第一收敛条件，得到已训练的单帧动作识别模型。

在一个实施例中，状态识别模块 1110 还用于：

获取目标对象在目标动作对应的各个参考对象状态，获取各个参考对象状态分别对应的参考物体；

将背景帧内的参考物体作为目标物体，获取目标物体对应的参考对象状态所对应的参考重合度阈值作为目标重合度阈值；

针对目标物体对应的交互信息组，当目标物体对应的物体区域和目标区域之间的区域重合度大于目标重合度阈值时，将目标物体对应的参考对象状态，作为目标对象在目标帧对应的目标对象状态。

在一个实施例中，状态识别模块 1110 还用于：

当目标物体对应的物体区域和目标区域之间的区域重合度小于或等于目标重合度阈值时，将目标帧和各个交互信息组输入已训练的交互模型，得到目标对象在目标帧对应的目标对象状态。

在一个实施例中，状态识别模块 1110 还用于：

将目标帧和各个交互信息组输入已训练的交互模型；

基于各个交互信息组在目标帧中分别确定重叠区域；

提取重叠区域的图像特征，得到第一交互特征，基于重叠区域对应的交互信息组中的物体和目标动作，得到第二交互特征，提取目标帧的图像特征，得到第三交互特征；

基于第一交互特征、第二交互特征和第三交互特征，得到综合交互特征；

基于综合交互特征输出目标对象在目标帧对应的目标对象状态。

在一个实施例中，对象状态识别装置还用于：

获取训练对象的训练动作对应的第二训练图像和第二训练图像对应的训练对象状态、训练信息组；训练对象状态是训练对象在第二训练图像对应的对象状态，训练信息组是基于第二训练图像中训练对象所处区域的位置信息、第二训练图像对应的背景图像中物体所处区域的位置信息和物体信息、训练动作确定的；

将第二训练图像和训练信息组输入待训练的交互模型，得到第二训练图像对应的预测对象状态；

基于训练对象状态和预测对象状态之间的差异，调整待训练的交互模型的模型参数，直至满足第二收敛条件，得到已训练的交互模型。

在一个实施例中，目标视频是对目标对象的生活环境实时采集图像得到的视频流。对象状态识别装置还用于：

在目标对象状态为异常对象状态时，向生活环境对应的管理终端发送针对异常对象状态的告警消息。

在一个实施例中，目标视频是对目标对象的生活环境实时采集图像得到的视频流。对象状态识别装置还用于：

在目标对象状态为异常对象状态时，从目标视频中获取目标帧的相邻帧集合；

对相邻帧集合中的相邻帧进行目标对象识别，将识别出目标对象的相邻帧作为辅助帧；从辅助帧中确定辅助对象；辅助对象和目标帧中的目标对象是属于同一类型的不同对象；

向辅助对象对应的对象终端发送针对异常对象状态的告警消息。

在一个实施例中，目标视频是通过设置在目标对象的生活环境中的目标设备对生活环境实时采集图像得到的视频流。对象状态识别装置还用于：

在目标对象状态为异常对象状态时，向目标设备发送针对异常对象状态的告警消息；目标设备用于展示针对异常对象状态的协助需求确认消息，并响应于针对协助需求确认消息的同意接受协助事件，确定针对目标对象的目标协助方式，向目标协助方式对应的协作设备发送针对目标对象的协作消息。

在一个实施例中，协助需求确认消息包括候选协助方式，目标设备还用于响应于针对候选协助方式的选择操作，确定目标协助方式。

在一个实施例中，当目标协助方式为机器人协助方式时，协助设备为与生活环境距离在预设阈值内的机器人。

上述对象状态识别装置，根据背景帧和目标帧对目标对象在目标帧中的目标动作进行分析，以识别目标对象在目标帧对应的目标对象状态。背景帧可以提供物体的信息，目标帧可以提供目标对象的信息，根据物体的信息和目标对象的信息建立交互信息组，通过交互信息组分析目标对象和物体之间的交互关系，能够有效提高对象状态的识别准确性。在识别对象状态的时候，考虑目标对象和环境的交互关系，能够有效区分因同一动作容易混淆的对象状态，提高对象状态的识别准确性。

上述对象状态识别装置中的各个模块可全部或部分通过软件、硬件及其组合来实现。上述各模块可以硬件形式内嵌于或独立于计算机设备中的处理器中，也可以以软件形式存储于计算机设备中的存储器中，以便于处理器调用执行以上各个模块对应的操作。

在一个实施例中，提供了一种计算机设备，该计算机设备可以是服务器，其内部结构图可以如图 12 所示。该计算机设备包括处理器、存储器、输入/输出接口(Input/Output，简称 I/O) 和通信接口。其中，处理器、存储器和输入/输出接口通过系统总线连接，通信接口通过输入/输出接口连接到系统总线。其中，该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质和内存。该非易失性存储介质存储有操作系统、计算机可读指令和数据库。该内存为易失性存储介质中的操作系统和计算机可读指令的运行提供环境。该计算机设备的数据库用于存储各种模型、各种参考信息等数据。该计算机设备的输入/输出接口用于处理器与外部设备之间交换信息。该计算机设备的通信接口用于与外部的终端通过网络连接通信。该计算机可读指令被处理器执行时以实现一种对象状态识别方法。

在一个实施例中，提供了一种计算机设备，该计算机设备可以是终端，其内部结构图可以如图 13 所示。该计算机设备包括处理器、存储器、输入/输出接口、通信接口、显示单元和输入装置。其中，处理器、存储器和输入/输出接口通过系统总线连接，通信接口、显示单元和输入装置通过输入/输出接口连接到系统总线。其中，该计算机设备的处理器用于提供计算和控制能力。该计算机设备的存储器包括非易失性存储介质和内存。该非易失性存储介质存储有操作系统和计算机可读指令。该内存为易失性存储介质中的操作系统和计算机可读指令的运行提供环境。该计算机设备的输入/输出接口用于处理器与外部设备之间交换信息。该计算机设备的通信接口用于与外部的终端进行有线或无线方式的通信，无线方式可通过 WIFI、移动蜂窝网络、NFC（近场通信）或其他技术实现。该计算机可读指令被处理器执行时以实现一种对象状态识别方法。该计算机设备的显示单元用于形成视觉可见的画面，可以是显示屏、投影装置或虚拟现实成像装置。显示屏可以是液晶显示屏或者电子墨水显示屏，该计算机设备的输入装置可以是显示屏上覆盖的触摸

层，也可以是计算机设备外壳上设置的按键、轨迹球或触控板，还可以是外接的键盘、触控板或鼠标等。

本领域技术人员可以理解，图 12、图 13 中示出的结构，仅仅是与本申请方案相关的部分结构的框图，并不构成对本申请方案所应用于其上的计算机设备的限定，具体的计算机设备可以包括比图中所示更多或更少的部件，或者组合某些部件，或者具有不同的部件布置。

在一个实施例中，还提供了一种计算机设备，包括存储器和处理器，存储器中存储有计算机可读指令，该处理器执行计算机可读指令时实现上述各方法实施例中的步骤。

在一个实施例中，提供了一种计算机可读存储介质，存储有计算机可读指令，该计算机可读指令被处理器执行时实现上述各方法实施例中的步骤。

在一个实施例中，提供了一种计算机程序产品，该计算机程序产品包括计算机可读指令，该计算机可读指令被处理器执行时实现上述各方法实施例中的步骤。

需要说明的是，本申请所涉及的用户信息（包括但不限于用户设备信息、用户个人信息等）和数据（包括但不限于用于分析的数据、存储的数据、展示的数据等），均为经用户授权或者经过各方充分授权的信息和数据，且相关数据的收集、使用和处理需要符合相关规定。

本领域普通技术人员可以理解实现上述实施例方法中的全部或部分流程，是可以通过计算机可读指令来指令相关的硬件来完成，所述的计算机可读指令可存储于一非易失性计算机可读存储介质中，该计算机可读指令在执行时，可包括如上述各方法的实施例的流程。其中，本申请所提供的各实施例中所使用的对存储器、数据库或其它介质的任何引用，均可包括非易失性和易失性存储器中的至少一种。非易失性存储器可包括只读存储器（Read-Only Memory, ROM）、磁带、软盘、闪存、光存储器、高密度嵌入式非易失性存储器、阻变存储器（ReRAM）、磁变存储器（Magnetoresistive Random Access Memory, MRAM）、铁电存储器（Ferroelectric Random Access Memory, FRAM）、相变存储器（Phase Change Memory, PCM）、石墨烯存储器等。易失性存储器可包括随机存取存储器（Random Access Memory, RAM）或外部高速缓冲存储器等。作为说明而非局限，RAM 可以是多种形式，比如静态随机存取存储器（Static Random Access Memory, SRAM）或动态随机存取存储器（Dynamic Random Access Memory, DRAM）等。本申请所提供的各实施例中所涉及的数据库可包括关系型数据库和非关系型数据库中至少一种。非关系型数据库可包括基于区块链的分布式数据库等，不限于此。本申请所提供的各实施例中所涉及的处理器可为通用处理器、中央处理器、图形处理器、数字信号处理器、可编程逻辑器、基于量子计算的数据处理逻辑器等，不限于此。

以上实施例的各技术特征可以进行任意的组合，为使描述简洁，未对上述实施例中的各个技术特征所有可能的组合都进行描述，然而，只要这些技术特征的组合不存在矛盾，都应当认为是本说明书记载的范围。

以上所述实施例仅表达了本申请的几种实施方式，其描述较为具体和详细，但并不能因此而理解为对本申请专利范围的限制。应当指出的是，对于本领域的普通技术人员来说，在不脱离本申请构思的前提下，还可以做出若干变形和改进，这些都属于本申请的保护范围。因此，本申请的保护范围应以所附权利要求为准。

权利要求书

- 1、一种对象状态识别方法，其特征在于，由计算机设备执行，所述方法包括：
获取目标视频；
获取所述目标视频对应的背景帧，从所述目标视频中确定目标对象的目标动作所对应的目标帧；
对所述背景帧和所述目标帧进行图像分割，得到所述背景帧中各个物体分别对应的物体区域和所述目标帧中所述目标对象对应的目标区域；
基于所述目标区域和所述物体区域的位置信息、所述物体区域对应的物体、所述目标动作，生成交互信息组，得到所述目标对象分别和所述各个物体之间的交互信息组；及
基于各个交互信息组，确定所述目标对象在所述目标帧对应的目标对象状态。
- 2、根据权利要求1所述的方法，其特征在于，所述获取所述目标视频对应的背景帧，包括：
对所述目标视频中的视频帧进行目标对象识别，每发现连续目标数量个视频帧均未识别出目标对象，从所述连续目标数量个视频帧中获取最后一个视频帧作为背景帧。
- 3、根据权利要求2所述的方法，其特征在于，所述对所述目标视频中的视频帧进行目标对象识别，包括：
基于通用对象识别算法，对所述目标视频中的视频帧进行对象识别，将首次识别到的活体对象作为目标对象；
基于所述目标对象的保护优先级确定目标对象识别算法；及
基于所述目标对象识别算法，对所述目标视频中的视频帧进行目标对象识别。
- 4、根据权利要求2或3所述的方法，其特征在于，所述目标数量和采用的目标对象识别算法对应的识别准确度呈负相关。
- 5、根据权利要求1至3中任一项所述的方法，其特征在于，所述从所述目标视频中确定目标对象的目标动作所对应的目标帧，包括：
对所述目标视频的视频帧进行目标对象识别，将识别出目标对象的视频帧作为候选帧；
获取目标时间段内的各个候选帧组成候选帧序列；
对所述候选帧序列中的候选帧进行目标对象骨架点识别，得到候选帧对应的骨架点数据，组成所述候选帧序列对应的骨架点数据序列；
将所述骨架点数据序列输入多帧动作识别模型，得到所述候选帧序列对应的预测动作；
及
在所述预测动作为目标动作时，从所述候选帧序列中确定所述目标对象的所述目标动作所对应的目标帧。
- 6、根据权利要求5所述的方法，其特征在于，所述从所述候选帧序列中确定所述目标对象的所述目标动作所对应的目标帧，包括：
将所述候选帧序列中的各个候选帧分别输入已训练的单帧动作识别模型，得到所述各个候选帧分别对应的动作识别结果；所述动作识别结果包括候选帧中目标对象的数量和动作；及
获取动作识别结果包括单个目标对象和所述目标动作的候选帧，作为所述目标对象的所述目标动作所对应的目标帧。
- 7、根据权利要求6所述的方法，其特征在于，所述方法还包括：
获取第一训练图像和所述第一训练图像对应的目标对象位置训练标签、目标对象动作训练标签；
将所述第一训练图像输入待训练的单帧动作识别模型，得到所述第一训练图像对应的目标对象位置预测标签和目标对象动作预测标签；及
基于所述目标对象位置训练标签和所述目标对象位置预测标签之间的差异、所述目标

对象动作训练标签和所述目标对象动作预测标签之间的差异,调整所述待训练的单帧动作识别模型的模型参数,直至满足第一收敛条件,得到所述已训练的单帧动作识别模型。

8、根据权利要求1至7中任一项所述的方法,其特征在于,所述基于各个交互信息组,确定所述目标对象在所述目标帧对应的目标对象状态,包括:

获取所述目标对象在所述目标动作对应的各个参考对象状态,获取所述各个参考对象状态分别对应的参考物体;

将所述背景帧内的参考物体作为目标物体,获取所述目标物体对应的参考对象状态所对应的参考重合度阈值作为目标重合度阈值;及

针对所述目标物体对应的交互信息组,当所述目标物体对应的物体区域和所述目标区域之间的区域重合度大于所述目标重合度阈值时,将所述目标物体对应的参考对象状态,作为所述目标对象在所述目标帧对应的目标对象状态。

9、根据权利要求8所述的方法,其特征在于,所述方法还包括:

当所述目标物体对应的物体区域和所述目标区域之间的区域重合度小于或等于所述目标重合度阈值时,将所述目标帧和所述各个交互信息组输入已训练的交互模型,得到所述目标对象在所述目标帧对应的目标对象状态。

10、根据权利要求1至7中任一项所述的方法,其特征在于,所述基于各个交互信息组,确定所述目标对象在所述目标帧对应的目标对象状态,包括:

将所述目标帧和各个交互信息组输入已训练的交互模型;

基于所述各个交互信息组在所述目标帧中分别确定重叠区域;

提取所述重叠区域的图像特征,得到第一交互特征,基于所述重叠区域对应的交互信息组中的物体和目标动作,得到第二交互特征,提取所述目标帧的图像特征,得到第三交互特征;

基于所述第一交互特征、所述第二交互特征和所述第三交互特征,得到综合交互特征;及

基于所述综合交互特征输出所述目标对象在所述目标帧对应的目标对象状态。

11、根据权利要求10所述的方法,其特征在于,所述方法还包括:

获取训练对象的训练动作对应的第二训练图像和所述第二训练图像对应的训练对象状态、训练信息组;所述训练对象状态是所述训练对象在所述第二训练图像对应的对象状态,所述训练信息组是基于所述第二训练图像中所述训练对象所处区域的位置信息、所述第二训练图像对应的背景图像中物体所处区域的位置信息和物体信息、所述训练动作确定的;

将所述第二训练图像和所述训练信息组输入待训练的交互模型,得到所述第二训练图像对应的预测对象状态;及

基于所述训练对象状态和所述预测对象状态之间的差异,调整所述待训练的交互模型的模型参数,直至满足第二收敛条件,得到所述已训练的交互模型。

12、根据权利要求1至11中任一项所述的方法,其特征在于,所述目标视频是对所述目标对象的生活环境实时采集图像得到的视频流,所述方法还包括:

在所述目标对象状态为异常对象状态时,向所述生活环境对应的管理终端发送针对所述异常对象状态的告警消息。

13、根据权利要求1至11中任一项所述的方法,其特征在于,所述目标视频是对所述目标对象的生活环境实时采集图像得到的视频流,所述方法还包括:

在所述目标对象状态为异常对象状态时,从所述目标视频中获取所述目标帧的相邻帧集合;

对所述相邻帧集合中的相邻帧进行目标对象识别,将识别出目标对象的相邻帧作为辅助帧;

从所述辅助帧中确定辅助对象；所述辅助对象和所述目标帧中的目标对象是属于同一类型的不同对象；及

向所述辅助对象对应的对象终端发送针对所述异常对象状态的告警消息。

14、根据权利要求 1 至 11 中任一项所述的方法，其特征在于，所述目标视频是通过设置在所述目标对象的生活环境中的目标设备对所述生活环境实时采集图像得到的视频流，所述方法还包括：

在所述目标对象状态为异常对象状态时，向所述目标设备发送针对异常对象状态的告警消息；所述目标设备用于展示针对所述异常对象状态的协助需求确认消息，并响应于针对所述协助需求确认消息的同意接受协助事件，确定针对所述目标对象的目标协助方式，向所述目标协助方式对应的协作设备发送针对所述目标对象的协作消息。

15、根据权利要求 14 所述的方法，其特征在于，所述协助需求确认消息包括候选协助方式，所述目标设备还用于响应于针对所述候选协助方式的选择操作，确定所述目标协助方式。

16、根据权利要求 14 所述的方法，其特征在于，当所述目标协助方式为机器人协助方式时，所述协助设备为与所述生活环境的距离在预设阈值内的机器人。

17、一种对象状态识别装置，其特征在于，所述装置包括：

视频获取模块，用于获取目标视频；

图像确定模块，用于获取所述目标视频对应的背景帧，从所述目标视频中确定目标对象的目标动作所对应的目标帧；

图像分割模块，用于对所述背景帧和所述目标帧进行图像分割，得到所述背景帧中各个物体分别对应的物体区域和所述目标帧中所述目标对象对应的目标区域；

信息组生成模块，用于基于所述目标区域和所述物体区域的位置信息、所述物体区域对应的物体、所述目标动作，生成交互信息组，得到所述目标对象分别和所述各个物体之间的交互信息组；及

状态识别模块，用于基于各个交互信息组，确定所述目标对象在所述目标帧对应的目标对象状态。

18、一种计算机设备，包括存储器和处理器，所述存储器存储有计算机可读指令，其特征在于，所述处理器执行所述计算机可读指令时实现权利要求 1 至 16 中任一项所述的方法的步骤。

19、一种计算机可读存储介质，其上存储有计算机可读指令，其特征在于，所述计算机可读指令被处理器执行时实现权利要求 1 至 16 中任一项所述的方法的步骤。

20、一种计算机程序产品，包括计算机可读指令，其特征在于，该计算机可读指令被处理器执行时实现权利要求 1 至 16 中任一项所述的方法的步骤。