

手持物体识别方法及装置

本申请要求于 2020 年 7 月 29 日提交中国国家知识产权局、申请号为 202010756462.1，发明名称为“手持物体识别方法及装置”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

本申请涉及物体识别领域，尤其涉及一种手持物体识别方法及装置。

背景技术

近几年，随着深度学习、认知计算等人工智能技术的发展及在诸多领域的应用，人工智能逐渐使我们的工作生活变得更加智能化。智能机器人领域在人工智能大发展的浪潮中也实现了质的飞跃，其逐渐被应用于教育、医疗和服务业等各个行业，并且越来越智能化地影响工业和服务业的发展，也潜移默化地提升了大众的生活水平。

对于服务型智能机器人而言，其在操作运行过程中，离不开视觉功能的支持。智能机器人在人机交互过程中，各种各样的物体扮演着非常重要的角色，智能机器人往往需要通过检测场景中相关的物体来增强对场景的理解，进而实现辅助机器人即时定位和地图构建（simultaneous localization and mapping, SLAM）、导航、运动控制，智能教育等功能。而一个完整的学习到应用的过程需要在交互的环境下进行。随着自然语言处理与理解技术的发展，机器人可以在与人的语音交互中解析语言，从而获取物体的概念以及属性等相关信息。在智能机器人学习了不同种类的物体之后，智能机器人能够被应用于更为广泛的室内场景中，并检测出更多认识的物体，实现从手持学习到复杂多变的室内场景的应用，这需要机器人自适应手持物体图像与实际室内场景中获取的物体图像的分布不匹配的问题，对智能机器人在场景中的实际应用而言是一项具有挑战性的任务。

该任务的核心部分，手持物体的定位。在手持物体定位相关方法上，可以大致分为两类，第一类是基于 RGB-D 的方法，其根据深度信息对场景进行重建，再对点云中的手和物体进行分割，得到手持物体，其明显不足在于深度信息的获取是代价高昂的，且在点云上获取分割结果精度也较低；第二类是基于 RGB 的方法，其进行 2D 物体的检测，显示的得到物体和手的位置，再通过分类的方式，确定并显示手持关系；但是这些方法在物体和手出现明显遮挡时是不可用的。

发明内容

本申请实施例提供一种手持物体识别方法及装置，采用本申请实施例可精准地确定出手持物体，实现了对于 RGB 图像，在物体或手出现明显遮挡时也可准确实现手持物体的识别，并且识别准确度也比较高。

第一方面，本申请实施例提供一种手持物体识别方法，包括：

获取待识别图像中一个或多个检测对象中每个检测对象的位置信息，并获取每个检测对象的第一标签，每个检测对象的位置信息为该检测对象在待识别图像中的位置信息，检

测对象的第一标签用于指示检测对象的类型，检测对象的类型用于表征检测对象的手持关系；根据每个检测对象的第一标签从一个或多个检测对象中获取手持物体，并从一个或多个检测对象的位置信息获取手持物体的位置信息；根据手持物体的位置信息对待识别图像中手持物体进行识别，以得到手持物体的识别结果。

在获取检测对象的位置信息后，通过用于指示检测对象的手持关系的第一标签确定出手持物体，在物体或手出现明显遮挡时可准确确定出手持物体，从而识别出手持物体。

在一个可选的实施例中，第一标签包括子类标签，子类标签用于指示检测对象的第一类型，获取每个检测对象的第一标签，包括：

在获取待识别图像的第一特征向量后，对待识别图像的第一特征向量进行第一卷积操作，以得到待识别图像的第二特征向量；对待识别图像的第二特征向量进行第一全连接操作，以得到每个检测对象的第一标签，其中，第一类型包括手持状态下的物体和非手持状态下的物体。

可选地，第一类型还包括持有物体的手和未持有物体的手。

在一个可选的实施例中，第一标签还包括超类标签，超类标签用于指示检测对象的第二类型，第一类型为第二类型的子类，获取每个检测对象的第一标签，还包括：

在获取待识别图像的第一特征向量后，对第一特征向量进行第二卷积操作，以得到待识别图像的第三特征向量；将第二特征向量和第三特征向量进行融合操作，以得到待识别图像的第四特征向量；对第四特征向量进行第三卷积操作，以得到待识别图像的第五特征向量；对第五特征向量进行第二全连接操作，以得到每个检测对象的超类标签；其中，第二类型包括手和物体。

通过引入超类标签，使得在进行区分检测对象的手持关系的同时，可保持其语义的一致性，进而保证检测对象的手持关系的准确性。

在一个可选的实施例中，获取待识别图像中一个或多个检测对象中每个检测对象的位置信息，包括：

在获取待识别图像的第一特征向量后，根据待识别图像的第一特征向量对待识别图像中一个或多个检测对象进行边框回归操作，以得到一个或多个检测对象中每个检测对象的位置信息。

在一个可选的实施例中，根据手持物体的位置信息对待识别图像中手持物体进行识别，以得到手持物体的识别结果，包括：

根据手持物体的位置信息从待识别图像中裁剪出包含手持物体的图像块，图像块为待识别图像的一部分；对图像块中的手持物体进行特征提取，以得到手持物体的第一特征向量；

根据手持物体的第一特征向量确定手持物体的参考标签及第一置信度，第一置信度为手持物体的标签为参考标签的概率；根据手持物体的参考标签和第一置信度获取手持物体的识别结果。

在一个可选的实施例中，识别结果包括手持物体的标签或者第二标签，第二标签用于指示未识别手持物体，增量特征库中包括 S 个特征向量及与 S 个特征向量中每个特征向量对应的标签， S 为大于或者等于 0 的整数，根据手持物体的参考标签和第一置信度获取手

持物体的识别结果，包括：

当第一置信度高于预设置信度时，将手持物体的参考标签确定为手持物体的标签；当第一置信度不高于预设置信度时，根据手持物体的第一特征向量和 S 个特征向量获取 S 个第二置信度， S 个第二置信度与 S 个特征向量一一对应， S 个特征向量中第 j 个特征向量所对应的第二置信度用于表征手持物体的标签为第 j 个特征向量对应的标签的概率；

若 S 个第二置信度均低于预设置信度，则将第二标签确定为手持物体的标签；若 S 个第二置信度中存在高于预设置信度的置信度，则将目标特征向量对应的标签确定为手持物体的标签，目标特征向量为高于预设置信度中的最高置信度对应的特征向量。

在一个可选的实施例中， S 个特征向量中每个特征向量包括 k 个子特征向量，根据手持物体的第一特征向量和 S 个特征向量获取 S 个第二置信度，包括：

对于 S 个特征向量中的任一特征向量 A ，根据手持物体的第一特征向量和特征向量 A 的 k 个子特征向量计算得到 k 个第三置信度；第三置信度用于表征手持物体的标签为特征向量 A 对应的标签的概率；根据 k 个第三置信度计算得到特征向量 A 对应的第二置信度。

在一个可选的实施例中，根据手持物体的第一特征向量确定手持物体的识别结果后，本申请的方法还包括：

向用户展示手持物体的识别结果。

比如通过显示屏显示或者通过语音播放的方式向用户展示手持物体的识别结果。

在一个可选的实施例中，本申请的方法还包括：

接收用户的第一信息，该第一信息包括手持物体的第三标签；将手持物体的第三标签和手持物体的第一特征向量保存至增量识别库中。

可选地，用户的第一信息是通过语音方式输入的，或者通过手持物体识别装置的虚拟键盘输入的。

在向用户展示手持物体的识别结果后，若用户对识别结果不满意，获取用户输入的标签，将用户输入的标签和在识别时获取的手持物体的特征向量保持至增量特征库中，使得在进行下次识别时，能够从增量识别库中获取手持物体的标签。

在一个可选的实施例中，本申请的方法还包括：

接收用户的第二信息，该第二信息用于指示用户翻转手持物体，且第二信息包括手持物体的第三标签；获取手持物体的多帧第一图像，多帧第一图像为手持物体不同角度的图像；根据多帧第一图像获取手持物体的多个第三特征向量；多个第三特征向量与多帧第一图像一一对应；根据聚类算法对多个第三特征向量进行聚类，以得到 k 个子特征向量；将手持物体的第三标签和第二特征向量保存至增量识别库中，其中，第二特征向量是根据 k 个子特征向量得到的，或者第二特征向量包括 k 个子特征向量。

在向用户展示手持物体的识别结果后，若用户对识别结果不满意，获取用户输入的标签，并引导用户翻转手持物体，获取不同角度的手持物体的图像，并基于不同角度的手持物体的图像得到手持物体的特征向量，将该特征向量和用户输入的标签保存至增量特征库中，从而在后续识别时，能从增强特征库中获取手持物体的标签；由于新获取的手持物体的特征向量是基于不同角度的手持物体的图像获取的，因此使得不管图像块中手持物体属于何种角度，都能够准确得到手持物体的标签。

第二方面，本申请实施例提供另一种手持物体识别方法，包括：

获取待识别图像中一个或多个检测对象中每个检测对象的位置信息，并获取每个检测对象的第一标签，每个检测对象的位置信息为该检测对象在待识别图像中的位置信息，检测对象的第一标签用于指示检测对象的类型，检测对象的类型用于表征检测对象的手持关系；根据每个检测对象的第一标签从一个或多个检测对象中获取手持物体，并从一个或多个检测对象的位置信息获取手持物体的位置信息；根据手持物体的位置信息从待识别图像中裁剪出包含手持物体的图像块，图像块为待识别图像的一部分；向物体识别服务器发送物体识别请求，物体识别请求携带有图像块；接收物体识别服务器的响应消息，响应消息携带有手持物体的识别结果。

在一个可选的实施例中，第一标签包括子类标签，子类标签用于指示检测对象的第一类型，获取每个检测对象的第一标签，包括：

在获取待识别图像的第一特征向量后，对待识别图像的第一特征向量进行第一卷积操作，以得到待识别图像的第二特征向量；

对待识别图像的第二特征向量进行第一全连接操作，以得到每个检测对象的第一标签，其中，第一类型包括手持状态下的物体和非手持状态下的物体。

进一步地，第一类型还包括持有物体的手和未持有物体的手。

在一个可选的实施例中，第一标签还包括超类标签，超类标签用于指示检测对象的第二类型，第一类型为第二类型的子类，获取每个检测对象的第一标签，还包括：

在获取待识别图像的第一特征向量后，对第一特征向量进行第二卷积操作，以得到待识别图像的第三特征向量；

将第二特征向量和第三特征向量进行融合操作，以得到待识别图像的第四特征向量；对第四特征向量进行第三卷积操作，以得到待识别图像的第五特征向量；对第五特征向量进行第二全连接操作，以得到每个检测对象的超类标签；其中，第二类型包括手和物体。

在一个可选的实施例中，获取待识别图像中一个或多个检测对象中每个检测对象的位置信息，包括：

在获取待识别图像的第一特征向量后，根据待识别图像的第一特征向量对待识别图像中一个或多个检测对象进行边框回归操作，以得到一个或多个检测对象中每个检测对象的位置信息。

在一个可选的实施例中，接收物体识别服务器的响应消息之后，本申请的方法还包括：向用户展示手持物体的识别结果。

在一个可选的实施例中，本申请的方法还包括：

接收用户的第一信息，第一信息包括手持物体的第三标签；向物体识别服务器发送第一消息，第一消息携带有手持物体的第三标签，第一消息用于指示物体识别服务器将手持物体的第三标签和手持物体的第一特征向量保存至增量识别库中。

在一个可选的实施例中，本申请的方法还包括：

接收用户的第二信息，第二信息用于指示用户翻转手持物体，且第二信息包括手持物体的第三标签；获取手持物体的多帧第一图像，多帧第一图像为手持物体不同角度的图像；向物体识别服务器发送第二消息，第二消息携带有手持物体的第三标签和多帧第一图像，

第二消息用于指示物体识别服务器根据多帧第一图像获取手持物体的第二特征向量，并将手持物体的第三标签和手持物体的第二特征向量保存至增量识别库中。

第三方面，本申请实施例提供另一种手持物体识别方法，包括：

接收手持物体识别装置发送的物体识别请求，物体识别请求携带有包括手持物体的图像块；对图像块中的手持物体进行特征提取，以得到手持物体的第一特征向量；根据手持物体的第一特征向量确定手持物体的参考标签及第一置信度，第一置信度为手持物体的标签为参考标签的概率；根据手持物体的参考标签和第一置信度获取手持物体的识别结果；向手持物体识别装置发送用于响应物体识别请求的响应消息，该响应消息携带有手持物体的识别结果。

在一个可选的实施例中，识别结果包括手持物体的标签或者第二标签，第二标签用于指示未识别手持物体，增量特征库中包括 S 个特征向量及与 S 个特征向量中每个特征向量对应的标签， S 为大于或者等于 0 的整数，根据手持物体的参考标签和第一置信度获取手持物体的识别结果，包括：

当第一置信度高于预设置信度时，将手持物体的参考标签确定为手持物体的标签；当第一置信度不高于预设置信度时，根据手持物体的第一特征向量和 S 个特征向量获取 S 个第二置信度， S 个第二置信度与 S 个特征向量一一对应， S 个特征向量中第 j 个特征向量所对应的第二置信度用于表征手持物体的标签为第 j 个特征向量对应的标签的概率；若 S 个第二置信度均低于预设置信度，则将第二标签确定为手持物体的标签；若 S 个第二置信度中存在高于预设置信度的置信度，则将目标特征向量对应的标签确定为手持物体的标签，目标特征向量为高于预设置信度中的最高置信度对应的特征向量。

在一个可选的实施例中， S 个特征向量中每个特征向量包括 k 个子特征向量，根据手持物体的第一特征向量和 S 个特征向量获取 S 个第二置信度，包括：

对于 S 个特征向量中的任一特征向量 A ，根据手持物体的第一特征向量和特征向量 A 的 k 个子特征向量计算得到 k 个第三置信度；第三置信度用于表征手持物体的标签为特征向量 A 对应的标签的概率；根据 k 个第三置信度计算得到特征向量 A 对应的第二置信度。

在一个可选的实施例中，本申请的方法还包括：

接收手持物体识别装置发送的第一消息，第一消息携带有手持物体的第三标签，将手持物体的第三标签和手持物体的第一特征向量保存至增量识别库中。

在一个可选的实施例中，本申请的方法还包括：

接收手持物体识别装置发送的第二消息，第二消息携带有手持物体的第三标签和多帧第一图像，多帧第一图像为手持物体不同角度的图像，根据多帧第一图像获取手持物体的多个第三特征向量；多个第三特征向量与多帧第一图像一一对应；根据聚类算法对多个第三特征向量进行聚类，以得到 k 个子特征向量；将手持物体的第三标签和第二特征向量保存至增量识别库中，其中，第二特征向量是根据 k 个子特征向量得到的，或者第二特征向量包括 k 个子特征向量。

第四方面，本申请实施例还提供一种手持物体识别装置，包括：

存储器和与该存储器耦合的处理器，其中，存储器中存储有指令，当处理器执行该指令时，执行如第一方面或第二方面方法中的部分或者全部。

第五方面，本申请实施例还提供一种物体识别服务器，包括：

存储器和与该存储器耦合的处理器，其中，存储器中存储有指令，当处理器执行该指令时，执行如第三方面方法中的部分或者全部。

第六方面，本申请实施例提供一种芯片系统，该芯片系统应用于电子设备；芯片系统包括一个或多个接口电路，以及一个或多个处理器；所述接口电路和所述处理器通过线路互联；所述接口电路用于从所述电子设备的存储器接收信号，并向所述处理器发送所述信号，所述信号包括所述存储器中存储的计算机指令；当所述处理器执行所述计算机指令时，所述电子设备执行如第一方面或第二方面或第三方面所述方法的部分或全部。

第七方面，本申请实施例提供一种计算机可读存储介质，该计算机存储介质存储有计算机程序，所述计算机程序包括程序指令，所述程序指令当被处理器执行时使所述处理器执行如第一方面或第二方面或第三方面所述方法的部分或全部。

第八方面，本申请实施例提供一种计算机程序产品，其特征在于，包括计算机指令，当所述计算机指令在电子设备上运行时，使得所述电子设备执行如第一方面或第二方面或第三方面所述方法的部分或全部。

本申请的这些方面或其他方面在以下实施例的描述中会更加简明易懂。

附图说明

为了更清楚地说明本申请实施例或现有技术中的技术方案，下面将对实施例或现有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示出一种人工智能主体框架示意图；

图1a为本申请实施例提供的一种手持物体识别的应用场景示意图；

图1b为本申请实施例提供的另一种手持物体识别的应用场景示意图；

图2为本申请实施例提供了一种系统架构示意图；

图3为本申请实施例提供了另一种系统架构示意图；

图4为本申请实施例提供的一种手持物体识别方法的流程示意图；

图5为本申请实施例提供的一种手持物体定位模型的架构示意图；

图5a为本申请实施例提供的一种手持物体定位结果示意图；

图5b为本申请实施例提供的另一种手持物体定位结果示意图；

图5c为本申请实施例提供的一种增量特征库中的特征向量获取示意图；

图6为本申请实施例提供的一种具体的应用流程示意图；

图7为本申请实施例提供的另一种手持物体识别方法的流程示意图；

图8为本申请实施例提供的另一种手持物体识别方法的流程示意图；

图9为本申请实施例提供的一种手持物体识别装置的结构示意图；

图10为本申请实施例提供的另一种手持物体识别装置的结构示意图；

图11为本申请实施例提供的一种物体识别服务器的结构示意图；

图12为本申请实施例提供的另一种手持物体识别装置的结构示意图；

图 13 为本申请实施例提供的另一种物体识别服务器的结构示意图。

具体实施方式

下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行描述，显然，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，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，都属于本发明保护的范围。

图 1 示出一种人工智能主体框架示意图，该主体框架描述了人工智能系统总体工作流程，适用于通用的人工智能领域需求。

下面从“智能信息链”（水平轴）和“IT 价值链”（垂直轴）两个维度对上述人工智能主题框架进行阐述。

“智能信息链”反映从数据的获取到处理的一系列过程。举例来说，可以是智能信息感知、智能信息表示与形成、智能推理、智能决策、智能执行与输出的一般过程。在这个过程中，数据经历了“数据—信息—知识—智慧”的凝练过程。

“IT 价值链”从人工智能的底层基础设施、信息（提供和处理技术实现）到系统的产业生态过程，反映人工智能为信息技术产业带来的价值。

（1）基础设施

基础设施为人工智能系统提供计算能力支持，实现与外部世界的沟通，并通过基础平台实现支撑。通过传感器与外部沟通；计算能力由智能芯片（CPU、NPU、GPU、ASIC、FPGA 等硬件加速芯片）提供；基础平台包括分布式计算框架及网络等相关的平台保障和支持，可以包括云存储和计算、互联互通网络等。举例来说，传感器和外部沟通获取数据，这些数据提供给基础平台提供的分布式计算系统中的智能芯片进行计算。

（2）数据

基础设施的上一层的数据用于表示人工智能领域的数据来源。数据涉及到图形、图像、语音、文本，还涉及到传统设备的物联网数据，包括已有系统的业务数据以及力、位移、液位、温度、湿度等感知数据。

（3）数据处理

数据处理通常包括数据训练，机器学习，深度学习，搜索，推理，决策等方式。

其中，机器学习和深度学习可以对数据进行符号化和形式化的智能信息建模、抽取、预处理、训练等。

推理是指在计算机或智能系统中，模拟人类的智能推理方式，依据推理控制策略，利用形式化的信息进行机器思维和求解问题的过程，典型的功能是搜索与匹配。

决策是指智能信息经过推理后进行决策的过程，通常提供分类、排序、预测等功能。

（4）通用能力

对数据经过上面提到的数据处理后，进一步基于数据处理的结果可以形成一些通用的能力，比如可以是算法或者一个通用系统，例如，翻译，文本的分析，计算机视觉的处理，语音识别，图像的识别等等。

（5）智能产品及行业应用

智能产品及行业应用指人工智能系统在各领域的产品和应用，是对人工智能整体解决方案的封装，将智能信息决策产品化、实现落地应用，其应用领域主要包括：智能制造、智能交通、智能家居、智能医疗、智能安防、自动驾驶，智慧城市，智能终端等。

参见图 1a，图 1a 为本申请实施例提供的一种手持物体识别的应用场景示意图。如图 1a 所示的智能超市导航场景，在逛商城过程中，用户拿起商品，手动或者语音开启智能手机的手持物体识别功能；智能手机获取包含手持商品的图像，然后从包含手持商品的图像中确定出商品在包含手持商品的图像中的位置信息，根据该位置信息从包含手持商品的图像中获取包含商品的图像块，该图像块为包含手持商品的一部分；对该图像块中的手持商品进行特征提取，以得到手持商品的特征向量，根据手持商品的特征向量对该手持商品进行识别，以得到商品的名称。

可选地，智能手机在获取上述包含手持商品的图像块后，向物体识别服务器发送携带该图像块的物体识别请求，对该图像块中的手持商品进行特征提取，以得到手持商品的特征向量，根据手持商品的特征向量对该手持商品进行识别，以得到商品的名称；向智能手机发送用于响应物体识别请求的响应消息，该响应消息携带上述商品的名称。

可选地，智能手机基于商品的名称获取商品的相关信息，比如商品产地，功能等，并通过语音解说，比如“尊敬的顾客，您好！这是 XX 平台的台灯，具有 XX 功能……”；还可以通过智能手机的屏幕显示商品的相关信息；当然还可以通过其他方式告知商品的相关信息。

参见图 1b，图 1b 为本申请实施例提供的另一种手持物体识别的应用场景示意图。如图 1b 所示的智能家庭教育机器人的场景：小朋友拿着一个物体，跑到智能机器人前，问智能机器人“你看我拿的是什么”，机器人开启手持物体识别功能，智能手机获取包含手持物体的图像，然后从包含手持物体的图像中确定出手持物体在包含手持物体的图像中的位置信息，根据该位置信息从包含手持物体的图像中获取包含手持物体的图像块，该图像块为包含手持物体的一部分；对该图像块中的手持物体进行特征提取，以得到手持物体的特征向量，根据手持物体的特征向量对该手持物体进行识别，以得到物体的名称。

可选地，智能手机在获取上述包含手持物体的图像块后，向物体识别服务器发送携带该图像块的物体识别请求，对该图像块中的手持物体进行特征提取，以得到手持物体的特征向量，根据手持物体的特征向量对该手持物体进行识别，以得到物体的名称；向智能手机发送用于响应物体识别请求的响应消息，该响应消息携带上述物体的名称。

机器人根据物体的名称获取该物体的相关信息，比如物体的名称和用途等，并通过语音方式告知小朋友该物体的名称和用途，比如“小朋友，你拿的是 XX，它是 XX”；若是未识别，则提示小朋友或者家长进行反馈该物体的相关信息，比如“小朋友，我不认识它，你能告诉我吗？”或者识别错误接收到用户反馈的物体的相关信息，机器人保存用户反馈的该物体的相关信息，以使下次检测到该物体时能准确识别。

参见附图 2，本申请实施例提供了一种系统架构 200。数据采集设备 260 用于采集训练

数据并存入数据库 230，训练设备 220 基于数据库 230 中维护的训练数据生成手持物体定位模型 201。其中，对于手持物体定位模型，训练数据包括包含检测对象的图像样本、检测对象的位置信息样本、检测对象的子类标签样本、可选地，训练数据还包括检测对象的超类标签样本；下面将更详细地描述训练设备 220 如何基于训练数据得到手持物体定位模型，手持物体定位模型 201 能够从输入的图像中确定出手持物体在图像中的位置信息。

神经网络中的每一层的工作可以用数学表达式 $\vec{y} = a(W \cdot \vec{x} + b)$ 来描述：从物理层面神经网络中的每一层的工作可以理解为通过五种对输入空间（输入向量的集合）的操作，完成输入空间到输出空间的变换（即矩阵的行空间到列空间），这五种操作包括：1、升维/降维；2、放大/缩小；3、旋转；4、平移；5、“弯曲”。其中 1、2、3 的操作由 $W \cdot \vec{x}$ 完成，4 的操作由 $+b$ 完成，5 的操作则由 $a()$ 来实现。这里之所以用“空间”二字来表述是因为被分类的对象并不是单个事物，而是一类事物，空间是指这类事物所有个体的集合。其中， W 是权重向量，该向量中的每一个值表示该层神经网络中的一个神经元的权重值。该向量 W 决定着上文所述的输入空间到输出空间的空间变换，即每一层的权重 W 控制着如何变换空间。训练深度神经网络的目的，也就是最终得到训练好的神经网络的所有层的权重矩阵（由很多层的向量 W 形成的权重矩阵）。因此，神经网络的训练过程本质上就是学习控制空间变换的方式，更具体的就是学习权重矩阵。

因为希望深度神经网络的输出尽可能的接近真正想要预测的值，所以可以通过比较当前网络的预测值和真正想要的目标值，再根据两者之间的差异情况来更新每一层神经网络的权重向量（当然，在第一次更新之前通常会有初始化的过程，即为神经网络中的各层预先配置参数），比如，如果网络的预测值高了，就调整权重向量让它预测低一些，不断的调整，直到神经网络能够预测出真正想要的目标值。因此，就需要预先定义“如何比较预测值和目标值之间的差异”，这便是损失函数（loss function）或目标函数（objective function），它们是用于衡量预测值和目标值的差异的重要方程。其中，以损失函数举例，损失函数的输出值（loss）越高表示差异越大，那么深度神经网络的训练就变成了尽可能缩小这个 loss 的过程。

训练设备 220 得到的手持物体定位模型 201 可以应用不同的系统或设备中。在附图 2 中，执行设备 210 配置有 I/O 接口 212，与外部设备进行数据交互，“用户”可以通过客户设备 240 向 I/O 接口 212 输入数据，比如待识别图像。

执行设备 210 可以调用数据存储系统 250 中的数据、代码等，也可以将数据、指令等存入数据存储系统 250 中。

计算模块 211 使用手持物体定位模型 201 对待识别图像进行处理，以得到手持物体在待识别图像中的位置信息；计算模块 211 根据手持物体在待识别图像中位置信息获取包含手持物体的图像块，该图像块为待识别图像中的一部分；计算模块 211 对图像块中的手持物体进行特征提取，以得到手持物体的特征向量；根据手持物体的特征向量获取手持物体的识别结果。

最后，I/O 接口 212 将手持物体的识别结果返回给客户设备 240，提供给用户。

更深层地，训练设备 220 可以针对不同的目标，基于不同的数据生成相应的手持物体定位模型 201，以给用户提供更佳的结果。

在附图 2 中所示情况下,用户可以手动指定输入执行设备 210 中的数据,例如,在 I/O 接口 212 提供的界面中操作。另一种情况下,客户设备 240 可以自动地向 I/O 接口 212 输入数据并获得结果,如果客户设备 240 自动输入数据需要获得用户的授权,用户可以在客户设备 240 中设置相应权限。用户可以在客户设备 240 查看执行设备 210 输出的结果,比如手持物体的识别结果,具体的呈现形式可以是显示、声音、动作等具体方式。客户设备 240 也可以作为数据采集端将采集到训练数据存入数据库 230。

值得注意的,附图 2 仅是本发明实施例提供的一种系统架构的示意图,图中所示设备、器件、模块等之间的位置关系不构成任何限制,例如,在附图 2 中,数据存储系统 250 相对执行设备 210 是外部存储器,在其它情况下,也可以将数据存储系统 250 置于执行设备 210 中。

参见附图 3,本发明实施例提供了另一种系统架构 300。执行设备 210 由一个或多个服务器实现,可选的,与其它计算设备配合,例如:数据存储、路由器、负载均衡器等设备;执行设备 210 可以布置在一个物理站点上,或者分布在多个物理站点上。执行设备 210 可以使用数据存储系统 250 中的数据,或者调用数据存储系统 250 中的程序代码实现训练得到手持物体定位模型,并将待识别图像输入到手持物体定位模型中进行处理,以得到手持物体在待识别图像中的位置信息。

用户可以操作各自的本地设备(例如本地设备 301 和本地设备 302)与执行设备 210 进行交互。每个本地设备可以表示任何计算设备,例如个人计算机、计算机工作站、智能手机、平板电脑、智能摄像头、智能汽车或其他类型蜂窝电话、媒体消费设备、可穿戴设备、机顶盒、游戏机等。

每个用户的本地设备可以通过任何通信机制/通信标准的通信网络与执行设备 210 进行交互,通信网络可以是广域网、局域网、点对点连接等方式,或它们的任意组合。

在另一种实现中,执行设备 210 的一个方面或多个方面可以由每个本地设备实现,例如,本地设备 301 可以为执行设备 210 提供本地数据或反馈计算结果。

需要注意的,执行设备 210 的所有功能也可以由本地设备实现。例如,本地设备 301 实现执行设备 210 的功能并为用户提供服务,或者为本地设备 302 的用户提供服务。

比如,本地设备 301 在获取包含手持物体的待识别图像后,向执行设备 210 发送携带待识别图像的物体识别请求;执行设备 210 根据手持物体定位模型获取手持物体在待识别图像中的位置信息;可选地,执行设备 210 根据手持物体在待识别图像中的位置信息获取包含手持物体的图像块,该图像块为待识别图像中的一部分;执行设备 210 该图像块中的手持物体进行特征提取,以得到手持物体的特征向量,根据手持物体的特征向量对该手持物体进行识别,以得到物体的名称;向本地设备 301 发送用于响应物体识别请求的响应消息,该响应消息携带上述物体的名称。

可选地,执行设备 210 在获取手持物体的位置信息后,本地设备 301 发送用于响应物体识别请求的响应消息,该响应消息携带上述手持物体的位置信息;本地设备 301 根据手持物体在待识别图像中的位置信息获取包含手持物体的图像块,该图像块为待识别图像中的一部分;执行设备 210 该图像块中的手持物体进行特征提取,以得到手持物体的特征向

量，根据手持物体的特征向量对该手持物体进行识别，以得到手持物体的名称。

本地设备 301 在获取手持物体的名称后，向用户展示物体的名称。

下面具体介绍上述设备如何实现手持物体定位和识别的。

参见图 4，图 4 为本申请实施例提供的一种手持物体识别方法的流程示意图。如图 4 所示，该方法包括：

S401、获取待识别图像中一个或多个检测对象中每个检测对象的位置信息，并获取检测每个检测对象的第一标签，每个检测对象的位置信息为该检测对象在待识别图像中的位置信息，检测对象的第一标签用于指示检测对象的类型，检测对象的类型用于表征检测对象的手持关系。

其中，上述检测对象包括手和非人物体。

可选地，在手持物体识别装置的手持物体识别功能开启后，通过该手持物体识别装置的摄像头或者从相册中获取待识别图像。比如说用户在逛超市时看到不认识的物品，用户拿起该物品；在唤起手持物体识别装置（比如智能手机、智能手表等）的手持物体识别功能后，识别装置获取上述待识别图像，该待识别图像包含上述不认识的图像。当然，也可以是先获取待识别图像，再开启手持物体识别功能；也可以是两者同时执行。

可选地，手持物体识别功能可以是用户通过语音唤起的或者是用户通过手动操作开启的。

在一个示例中，获取待识别图像中一个或多个检测对象中每个检测对象的位置信息，包括：

在获取待识别图像的第一特征向量后，根据待识别图像的第一特征向量对待识别图像中的一个或多个检测对象进行边框回归操作，以得到一个或多个检测对象中每个检测对象的位置信息。

边框回归操作的具体流程：获取待识别图像中一个或多个检测对象中的每个检测对象的初始边框，该初始边框所覆盖的区域包括该检测对象；然后对初始边框不断进行平移和缩放，直至获取该检测对象的目标边框，该目标边框与该检测对象的最小边框的交并比大于预设交并比；其中，检测对象的最小边框为覆盖区域的面积最小，且覆盖区域包括该检测对象的边框。

其中，检测对象的位置信息以该检测对象的目标边框的信息表示；可选地，边框的信息可以包括边框对角线两端的坐标；还可以是边框中心点坐标和边框的尺寸；还可以是边框的左上角定点坐标、右上角定点坐标、左下角定点坐标或右上角定点坐标及边框的尺寸。

在一个可行的实施例中，第一标签包括子类标签，该子类标签用于指示检测对象的第一类型，获取检测每个检测对象的第一标签，包括：

在获取待识别图像的第一特征向量后，对该待识别图像的第一特征向量进行第一卷积操作，以得到待识别图像的第二特征向量；然后对该待识别图像的第二特征向量进行第一全连接操作，以得到待识别图像中一个或多个检测对象中每个检测对象的第一标签，其中，第一类型包括手持状态下的物体和非手持状态下的物体。

可选地，第一类型还包括持有物体的手和未持有物体的手。

进一步地, 在一个可行的实施例中, 第一标签还包括超类标签, 该超类标签用于指示检测对象的第二类型, 第一类型为第二类型的子类, 获取每个检测对象的第一标签包括:

在获取待识别图像的第一特征向量后, 对该第一特征向量进行第二卷积操作, 以得到待识别图像的第三特征向量; 将待识别图像的第二特征向量和第三特征向量进行融合, 以得到待识别图像的第四特征向量; 对该第四特征向量进行第三卷积操作, 以得到待识别图像的第五特征向量; 对该第五特征向量进行第二全连接操作, 以得到待识别图像中一个或多个检测对象中每个检测对象的超类标签, 其中, 超类标签包括手和物体。

如图 5 所示, 将待识别图像输入到骨干网络中进行处理, 以得到不同尺度的特征向量; 然后对不同尺度的特征向量进行堆叠操作, 以得到待识别图像的第一特征向量。对待识别图像的第一特征向量进行边框回归操作, 以得到待识别图像中一个或多个检测对象中每个检测对象的位置信息。具体地, 如图 5 所示, 将待识别图像的第一特征向量通过第一普通卷积神经网络、第一分离式卷积神经网络和第二普通卷积神经网络进行处理, 以得到待识别图像中的一个或多个检测对象中每个检测对象的位置信息。

在此需要说明的是, 边框回归的主要思想是: 寻找一种函数关系使得输入初始边框 P 经过映射得到一个跟检测对象的最小边框 G 更接近的目标边框 $\hat{G}$; 其中, 检测对象的最小边框为覆盖区域的面积最小, 且该覆盖区域包括该检测对象的边框。上述函数关系可通过上述第一普通卷积神经网络、第一分离式卷积神经网络和第二普通卷积神经网络来实现。

其中, 检测对象的位置信息以该检测对象的目标边框的信息表示; 可选地, 边框的信息可以包括边框对角线两端的坐标; 还可以是边框中心点坐标和边框的尺寸; 还可以是边框的左上角定点坐标、右上角定点坐标、左下角定点坐标或右上角定点坐标及边框的尺寸。

如图 5 所示, 将待识别图像的第一特征向量输入到第二分离式卷积神经网络和第三普通卷积神经网络中进行处理, 实现对待识别图像的第一特征向量进行第一卷积操作, 得到待识别图像的第二特征向量, 再将待识别图像的第二特征向量输入第一全连接网络中进行处理, 实现对待识别图像的第二特征向量的第一全连接操作, 得到待识别图像的一个或多个检测对象中每个检测对象的子类标签, 该子类标签用于指示检测对象的第一类型。

可选地, 该第一类型包括被手持的物体或未被手持的物体, 或者第一类型包括被手持的物体、未被手持的物体、手持物体的手或者未手持物体的手, 如图 5a 所示。

假设第一类型包括四种类型, 比如第一类型包括被手持的物体、未被手持的物体、手持物体的手和未手持物体的手时, 将待识别图像的第二特征向量输入到第一全连接网络中进行处理, 得到待识别图像中每个检测对象为上述四种类型中每种类型的概率, 将最大概率所对应的类型确定为该检测对象的第一类型。

在此需要指出的是, 获取待识别图像中一个或多个检测对象中每个检测对象的位置信息, 与获取每个检测对象的第一标签可以串行处理的, 也可是并行处理的。

在一个可选的示例中, 在确定手持物体的位置信息后, 只显示被手持物体的边框及子类标签, 如图 5b 所示。

可选地, 如图 5 所示, 将待识别图像的第一特征向量输入到第三分离式卷积神经网络和第四普通卷积神经网络中进行处理, 实现对待识别图像的第一特征向量进行第二卷积操作, 得到待识别图像的第三特征向量; 对待识别图像的第二特征向量和第三特征向量进行

融合操作，以得到待识别图像的第四特征向量；将待识别图像的第四特征向量输入到第四分离式卷积神经网络和第五普通卷积神经网络中进行处理，实现对待识别图像的第四特征向量进行第三卷积操作，得到待识别图像的第五特征向量；再将待识别图像的第五特征向量输入第二全连接网络中进行处理，实现对待识别图像的第五特征向量的第二全连接操作，得到待识别图像的一个或多个检测对象中每个检测对象的超类标签，该超类标签用于指示检测对象的第二类型，其中，第一类型为第二类型的子类。

可选地，第二类型包括手或物体。手持物体的手和未手持物体的手均属于手，被手持的物体和未被手持的物体均属于物体。

其中，图 5 所示的第二分离式卷积神经网络、第三普通卷积神经网络、第三分离式卷积神经网络、第四普通卷积神经网络、第四分离式卷积神经网络、第五普通卷积神经网络、第一全连接网络和第二全连接网络构成分类检测网络。通过分类检测网络，可以对 RGB 图像进行手持物体定位，进而可以隐式判断出手持关系并确定出手持物体。

在一个可选的实施例，获取待识别图像中一个或多个检测对象中每个检测对象的位置信息，并获取每个检测对象的第一标签具体可通过手持物体定位模型来实现，基于手持物体定位模型获取每个检测对象的位置信息和第一标签具体参加上述相关描述在此不再叙述。手持物体定位模型的具体架构如图 5 所示。

在使用手持物体定位模型之前，本申请的方法还包括：

获取手持物体定位模型。

其中，手持物体定位模型包括上述骨干网络、执行上述堆叠操作的堆叠模块、上述普通卷积神经网络、分离式卷积神经网络、执行上述融合操作的融合模块和全连接神经网络。

其中，手持物体定位模型是使用 mobilenetV3 作为基础网络架构，采用 SSD 的 Head 结构实现的。下表 1 示意了 mobilenetV3 的网络结构。

Input	Operator	exp size	#out	SE	NL	s
$224^2 \times 3$	conv2d	-	16	-	HS	2
$112^2 \times 16$	bneck, 3×3	16	16	-	RE	1
$112^2 \times 16$	bneck, 3×3	64	24	-	RE	2
$56^2 \times 24$	bneck, 3×3	72	24	-	RE	1
$56^2 \times 24$	bneck, 5×5	72	40	√	RE	1
$28^2 \times 40$	bneck, 5×5	120	40	√	RE	2
$28^2 \times 40$	bneck, 5×5	120	40	√	RE	1
$28^2 \times 40$	bneck, 3×3	240	80	-	HS	1
$14^2 \times 80$	bneck, 3×3	200	80	-	HS	2
$14^2 \times 80$	bneck, 3×3	184	80	-	HS	1
$14^2 \times 80$	bneck, 3×3	184	80	-	HS	1
$14^2 \times 80$	bneck, 3×3	480	112	√	HS	1
$14^2 \times 112$	bneck, 3×3	672	112	√	HS	1
$14^2 \times 112$	bneck, 5×5	672	160	√	HS	1

$14^2 \times 112$	bneck, 5×5	672	160	√	HS	1
$7^2 \times 160$	bneck, 5×5	960	160	√	HS	2
$7^2 \times 160$	conv2d, 1×1	-	960	-	HS	1
$7^2 \times 960$	Pool, 7×7	-	-	-	HS	-
$1^2 \times 960$	conv2d, 1×1 , NBN	-	1280	-	HS	1
$1^2 \times 1280$	conv2d, 1×1 , NBN	-	k	-		-

表 1

其中, SE 表示在该块中是否有挤压和激发; NL 表示使用的非线性类型; HS 表示激活函数 h-swish, RE 表示激活函数 ReLU; NBN 表示无批次归一化, s 表示步长。

可选地, 手持物体定位模型可以是手持物体识别装置训练得到的, 还可以是其他设备训练好后, 手持物体识别装置从其他设备中获取的。

具体地, 手持物体识别装置训练得到手持物体定位模型, 具体包括:

获取初始模型和训练数据, 训练数据包括样本图像和样本图像中的一个或多个检测对象的位置信息样本和第一标签样本; 将样本图像输入到初始模型中进行处理, 以得到样本图像中一个或多个检测对象的位置信息预测结果和第一标签预测结果; 基于第一损失函数及样本图像的一个或多个检测对象的位置信息样本和位置信息预测结果计算得到第一损失值; 基于聚焦损失 (focal loss) 函数及样本图像中一个或多个检测对象的第一标签样本和第一标签预测结果计算得到第二损失值; 然后基于第一损失值和第二损失值调整初始模型中的参数, 从而得到调整后的模型; 然后判断调整后的模型是否满足预设条件, 比如调整后的模型的精度是否达到预设精度等; 若调整后的模型满足预设条件, 则将该调整后的模型确定为手持物体定位模型; 若调整后的模型不满足预设条件, 则重新按照上述方法对上述调整后的模型进行训练, 直至得到上述手持物体定位模型。

可选地, 第一损失函数可以为回归损失函数, 比如 L1 loss, 第二损失函数为聚焦损失 (focal loss) 函数。

通过上述方法, 可以显示地检测手和物体, 并隐式地识别出手持物体, 并且在手不可见或较难被检测出时, 依然可以检测物体并确认手持关系。

S402、根据每个检测对象的第一标签从一个或多个检测对象获取手持物体, 并从一个或多个检测对象的位置信息中获取手持物体的位置信息。

其中, 手持物体的位置信息为手持物体在待识别图像中的位置信息。

S403、根据手持物体的位置信息对待识别图像中的手持物体进行识别, 以得到手持物体的识别结果。

具体地, 根据上述手持物体的位置信息从待识别图像中裁剪出包含该手持物体的图像块, 该图像块为待识别图像中的一部分; 对该图像块中的手持图像进行特征提取, 以得到手持物体的第一特征向量; 根据该手持物体的第一特征向量确定出手持物体的参考标签和第一置信度, 该第一置信度为手持物体的标签为参考标签的概率; 根据该手持物体的参考标签和第一置信度获取手持物体的识别结果。

在一个可选的示例中, 根据该手持物体的第一特征向量与确定出手持物体的参考标签和第一置信度, 包括:

计算手持物体的第一特征向量与第一特征库中每个第一特征向量之间的距离，其中，第一特征库中包括一个或多个特征向量及其对应的标签，并根据手持物体的第一特征向量与第一特征库中每个特征向量之间的距离确定手持物体的第一特征向量与第一特征库中每个特征向量之间的置信度，将第一特征库中，与手持物体的第一特征向量之间置信度最大的特征向量对应的标签确定为手持物体的参考标签，该手持物体的特征向量与第一特征向量之间的置信度越大，该手持物体的特征向量与第一特征向量之间的距离越小，第一置信度为手持物体的第一特征向量与第一特征库中每个特征向量之间的置信度中的最大置信度。

在一个可选的示例中，手持物体的识别结果包括手持物体的目标标签或者第二标签，该第二标签用于指示未识别出手持物体，增量特征库中包括 S 个特征向量及与 S 个特征向量中每个特征向量对应的标签， S 为大于或者等于 0 的整数，根据手持物体的参考标签和第一置信度获取手持物体的识别结果，包括：

当第一置信度高于预设置信度时，将手持物体的参考标签确定为手持物体的标签；当第一置信度不高于预设置信度时，根据手持物体的第一特征向量和 S 个特征向量获取 S 个第二置信度， S 个第二置信度与 S 个特征向量一一对应， S 个特征向量中第 j 个特征向量所对应的第二置信度用于表征手持物体的标签为第 j 个特征向量对应的标签的概率；

若 S 个第二置信度均低于预设置信度，则将第二标签确定为手持物体的标签；若 S 个第二置信度中存在高于预设置信度的置信度，则将目标特征向量对应的标签确定为手持物体的标签，目标特征向量为高于预设置信度中的最高置信度对应的特征向量。

在一个可选的示例中，增量特征库中的每个特征向量包括 k 个子特征向量，计算手持物体的特征向量与增量特征库中每个特征向量之间的第二置信度，得到 S 个第二置信度，包括：

对于增量特征库中任一特征向量 A ，分别计算该特征向量 A 的 k 个子特征向量与手持物体的特征向量之间的置信度，得到 k 个第三置信度；根据该 k 个第三置信度获取该特征向量 A 的第二置信度。

可选地，特征向量 A 的第二置信度为上述 k 个第三置信度中的最大置信度；或者对该 k 个第三置信度进行加权求平均得到的。

在一个可选的示例中，在获取手持物体的目标标签，比如手持物体的名称后，根据手持物体的目标标签获取手持物体的相关信息，比如手持物体的功能、工作原理等。

在一个可选的实施例，在对待识别视频进行手持物体识别时，按照上述方法对待识别视频中的图像逐帧进行手持物体识别；对于待识别视频的非最后一帧图像中的任一帧图像，若按照上述方法获取该图像中的手持物体的标签，则停止后续图像的识别；若按照上述方法未获取该图像中的手持物体的标签，则按照上述方法对该图像的后一帧继续进行识别，直到获取手持物体的标签；对于待识别视频的最后一帧图像，若按照上述方法未获取手持物体的标签，则第二标签确定为手持物体识别结果。

在一个示例中，在获取手持物体的识别结果后，向用户展示手持物体的识别结果。

具体地，在手持物体识别装置的显示屏上显示手持物体和手持物体的识别结果，或者通过语音的方式，以告知用户手持物体的识别结果，比如手持物体识别装置播放“您手上拿的是 XX”，或者“无法识别出您手上拿的物体”；进一步地，在手持物体识别装置的显

示屏幕上显示手持物体的相关信息，或者通过语音的方式告知用户手持物体的相关信息。

在一个可行的实施例，在向用户展示手持物体的识别结果后，接收用户的第一信息，该第一信息包括手持物体的第三标签；将第三标签和手持物体的特征向量保存至第二特征库中。

在一个示例中，用户对于手持物体识别装置识别出的结果不满意时，比如手持物体为“茶杯”，手持物体识别装置识别出来的结果为“碗”，此时用户通过语音的方式或者通过手动输入方式告知手持物体识别装置用户手持的物体为“茶杯”，手持物体识别装置将手持物体的第三标签（比如茶杯）和手持物体的第一特征向量保存至第二特征库，以使手持物体识别装置后续能够正确识别出标签为第三标签的手持物体。

在一个可行的实施例，接收用户的第二信息，该第二信息用于指示用户翻转手持物体，该第二信息包括手持物体的第三标签；在用户翻转手持物体时，获取手持物体的多帧第一图像，该多帧第一图像为手持物体不同角度的图像；根据多帧第一图像获取手持物体的第二特征向量。

可选地，对多帧第一图像中的手持物体进行特征提取，以得到手持物体的多个第三特征向量；对该多个第三特征向量进行聚类，以得到 k 个子特征向量；将 k 个子特征向量作为手持物体的第二特征向量和第三标签保存至增量特征库中。

可选地，对多帧第一图像中的手持物体进行特征提取，以得到手持物体的多个第三特征向量；对该多个第三特征向量进行聚类，以得到 k 个子特征向量；对该 k 个子特征向量进行求平均或者加权求平均，以得到手持物体的第二特征向量，并将该手持物体的第二特征向量和第三标签保存至增量特征库中。

在一个具体的示例中，手持物体识别装置通过语音方式引导用户对手持物体进行多视觉旋转，手持物体识别装置的摄像头获取多帧第一图像；该多帧第一图像为手持物体不同角度的图像；比如手持物体的正视图、左视图、右视图、后视图和下视图等；如图 5c 所示，对手持物体的多帧第一图像进行 CNN 语义特征提取，以得到多帧第一图像分别对应的多个参考特征向量；可选地，多个参考特征向量为将多帧第一图像输入到物体识别模型中后，从该物体识别模型的最后池化层的输出的特征向量；对多个参考特征向量进行多视角聚类，聚类中心个数为 k ，得到 k 个聚类中心的特征，该 k 个聚类中心的特征为上述手持物体的第二特征向量的 k 个子特征向量，并将 k 个子特征向量和手持式物体的标签保存至第二特征库中，或者对 k 个聚类中心的特征进行求平均或者加权求平均得到手持物体的第二特征向量，将手持物体的第二特征向量和手持物体的标签保存至第二特征库中。在用户下次需要对该手持物体进行识别时，手持物体识别装置在第二特征库中成功匹配到手持物体，并获取该手持物体的标签。

可选地，对多个参考特征向量进行多视角聚类，具体可采用 k -means 聚类算法；当然还可以采用其他聚类算法，在此不做具体限定。

在此需要指出的是，聚类中心个数与多帧第一图像中手持物体的角度相关联；比如多帧第一图像包括手持物体的 5 个角度的图像，则聚类中心的个数为 5。

在将手持物体的第二特征向量保存至增量特征库中后，后续再进行手持物体的识别时，就可以从增量识别库中的特征向量匹配成功，从而得到手持物体的标签。

举例说明，如图 6 所示，用户通过语音方式触发机器人开启手持物体识别功能，比如用户对机器人说：“小艺，看看我手里拿的是什么”；机器人通过其摄像头获取包含手持物体的待识别图像；然后机器人的手持物体定位模块基于待识别图像获取手持物体在该待识别图像中的位置信息；机器人根据手持物体的位置信息从待识别图像中裁剪出包含手持物体的图像块，从该图像块中提取出手持物体的第一特征向量；基于手持物体的第一特征向量进行物体识别，得到手持物体的参考标签和第一置信度，第一置信度用于表征手持物体的标签为参考标签的概率；若手持物体的第一置信度超过预设置信度，则将手持物体的参考标签确定为手持物体的标签；若手持物体的第一置信度低于预设置信度，根据手持物体的第一特征向量检索增量特征库，获取手持物体的第一特征向量与增量特征库中每个特征向量的第二置信度，该第二置信度用于表征手持物体的标签为该特征向量所对应标签的概率；将手持物体的第一特征向量与增量特征库中每个特征向量的第二置信度中，高于预设置信度且最大置信度的特征向量所对应的标签确定为手持物体的标签；若手持物体的第一特征向量与增量特征库中每个特征向量的第二置信度均低于预设置信度，则将手持物体的标签确定第二标签，第二用于指示未识别出手持物体，具体过程参见上述实施例的相关描述，在此不再叙述；机器人在获取手持物体的标签后，若机器人识别出手持物体，在通过语音的方式告知用户：小艺说：“我猜这是 B，对不对啊”；用户说：“答对了”，此时机器人结束此次物体识别；用户说：“答错了，这是 A”；若机器人未识别出手持物体，也通过语音的方式告知用户：小艺说：“哎呀，我不认识啊，告知我这是什么，下次我就认识了”，用户说：“这是 A”；在机器人识别错了或者未识别出手持物体，机器人的增量模块将手持物体的第一特征向量和从用户处获取的手持物体的标签保存至是增量特征库中，或者机器人引导用户翻转手持物体，同时通过摄像头获取手持物体的多帧第一图像，该多帧第一图像为手持物体不同角度的图像；增量模块根据多帧第一图像获取手持物体的第二特征向量，具体过程参见上述实施例的相关描述，在此不再叙述；将手持物体的第二特征向量和从用户处获取的手持物体的标签保存至是增量特征库中。

可以看出，在获取检测对象的位置信息后，通过用于指示检测对象的手持关系的第一标签确定出手持物体，在物体或手出现明显遮挡时可准确的确定出手持物体，从而识别出手持物体。通过引入超类标签，使得在进行区分检测对象的手持关系的同时，可保持其语义的一致性，进而保证检测对象的手持关系的准确性。在向用户展示手持物体的识别结果后，若用户对识别结果不满意，获取用户输入的标签，将用户输入的标签和在识别时获取的手持物体的特征向量保持至增量特征库中，使得在进行下次识别时，能够从增量识别库中获取手持物体的标签。其中，手持物体的特征向量是在向用户展示手持物体的识别结果后，若用户对识别结果不满意，获取用户输入的标签，并引导用户翻转手持物体，获取不同角度的手持物体的图像，并基于不同角度的手持物体的图像得到的，由于新获取的手持物体的特征向量是基于不同角度的手持物体的图像获取的，因此使得不管图像块中手持物体属于何种角度，都能够准确得到手持物体的标签。

参见图 7，图 7 为本申请实施例提供的另一种手持物体识别方法流程示意图。如图 7 所示，该方法包括：

S701、获取待识别图像中一个或多个检测对象中每个检测对象的位置信息，并获取每个检测对象的第一标签，每个检测对象的位置信息为该检测对象在待识别图像中的位置信息，检测对象的第一标签用于指示检测对象的类型，检测对象的类型用于表征检测对象的手持关系。

在一个可行的实施例中，第一标签包括子类标签，子类标签用于指示检测对象的第一类型，获取每个检测对象的第一标签，包括：

在获取待识别图像的第一特征向量后，对待识别图像的第一特征向量进行第一卷积操作，以得到待识别图像的第二特征向量；对待识别图像的第二特征向量进行第一全连接操作，以得到每个检测对象的第一标签，其中，第一类型包括手持状态下的物体和非手持状态下的物体。

在一个可行的实施例中，第一类型还包括持有物体的手和未持有物体的手。

在一个可行的实施例中，第一标签还包括超类标签，超类标签用于指示检测对象的第二类型，第一类型为第二类型的子类，获取每个检测对象的第一标签，还包括：

在获取待识别图像的第一特征向量后，对第一特征向量进行第二卷积操作，以得到待识别图像的第三特征向量；将第二特征向量和第三特征向量进行融合操作，以得到待识别图像的第四特征向量；对第四特征向量进行第三卷积操作，以得到待识别图像的第五特征向量；对第五特征向量进行第二全连接操作，以得到每个检测对象的超类标签；其中，第二类型包括手和物体。

在一个可行的实施例中，获取待识别图像中一个或多个检测对象中每个检测对象的位置信息，包括：

在获取待识别图像的第一特征向量后，根据待识别图像的第一特征向量对待识别图像中一个或多个检测对象进行边框回归操作，以得到一个或多个检测对象中每个检测对象的位置信息。

S702、根据每个检测对象的第一标签从一个或多个检测对象中获取手持物体，并从一个或多个检测对象的位置信息获取手持物体的位置信息。

S703、根据手持物体的位置信息从待识别图像中裁剪出包含手持物体的图像块，图像块为待识别图像的一部分。

S704、向物体识别服务器发送物体识别请求，物体识别请求携带有图像块；接收物体识别服务器的响应消息，响应消息携带有手持物体的识别结果。

在一个可行的实施例中，接收物体识别服务器的响应消息之后，本申请的方法还包括：向用户展示手持物体的识别结果。

在一个可行的实施例中，本申请的方法还包括：

接收用户的第一信息，第一信息包括手持物体的第三标签；向物体识别服务器发送第一消息，第一消息携带有手持物体的第三标签，第一消息用于指示物体识别服务器将手持物体的第三标签和手持物体的第一特征向量保存至增量识别库中。

在一个可行的实施例中，本申请的方法还包括：

接收用户的第二信息，第二信息用于指示用户翻转手持物体，且第二信息包括手持物体的第三标签；获取手持物体的多帧第一图像，多帧第一图像为手持物体不同角度的图像；

向物体识别服务器发送第二消息，第二消息携带有手持物体的第三标签和多帧第一图像，第二消息用于指示物体识别服务器根据多帧第一图像获取手持物体的第二特征向量，并将手持物体的第三标签和手持物体的第二特征向量保存至增量识别库中。

在此需要说明的是，S701-S704 的具体实现过程可参见上述步骤 S401 和 S403 的相关描述，在此不再叙述。

参见图 8，图 8 为本申请实施例提供的另一种手持物体识别方法流程示意图。如图 8 所示，该方法包括：

S801、接收手持物体识别装置发送的物体识别请求，物体识别请求携带有包括手持物体的图像块。

在一个可选的实施例中，物体识别请求中包括手持物体的位置信息和待识别图像，物体识别服务器根据手持物体的位置信息从待识别图像中获取包括手持物体的图像块。

在一个可选的实施例中，图 4 所示实施例里中的手持物体定位模型运行在物体识别服务器中，在手持物体识别装置需要进行手持物体识别时，向物体识别服务器发送包含待识别图像的物体识别请求，物体识别服务器基于手持物体定位模型从待识别图像中获取手持物体信息，并基于手持物体的位置信息从待识别图像中获取包含手持物体的图像块，并按照图 8 所示实施例的方法获取手持物体的识别结果，并向手持物体识别装置发送包含手持物体识别结果，且用于响应上述物体识别请求的响应消息。

S802、对图像块中的手持物体进行特征提取，以得到手持物体的第一特征向量。

S803、根据手持物体的第一特征向量确定手持物体的参考标签及第一置信度，第一置信度为手持物体的标签为参考标签的概率。

S804、根据手持物体的参考标签和第一置信度获取手持物体的识别结果。

在一个可行的实施例中，识别结果包括手持物体的标签或者第二标签，第二标签用于指示未识别手持物体，增量特征库中包括 S 个特征向量及与 S 个特征向量中每个特征向量对应的标签，S 为大于或者等于 0 的整数，根据手持物体的参考标签和第一置信度获取手持物体的识别结果，包括：

当第一置信度高于预设置信度时，将手持物体的参考标签确定为手持物体的标签；当第一置信度不高于预设置信度时，根据手持物体的第一特征向量和 S 个特征向量获取 S 个第二置信度，S 个第二置信度与 S 个特征向量一一对应，S 个特征向量中第 j 个特征向量所对应的第二置信度用于表征手持物体的标签为第 j 个特征向量对应的标签的概率；若 S 个第二置信度均低于预设置信度，则将第二标签确定为手持物体的标签；若 S 个第二置信度中存在高于预设置信度的置信度，则将目标特征向量对应的标签确定为手持物体的标签，目标特征向量为高于预设置信度中的最高置信度对应的特征向量。

可选地，S 个特征向量中每个特征向量包括 k 个子特征向量，根据手持物体的第一特征向量和 S 个特征向量获取 S 个第二置信度，包括：

对于 S 个特征向量中的任一特征向量 A，根据手持物体的第一特征向量和特征向量 A 的 k 个子特征向量计算得到 k 个第三置信度；第三置信度用于表征手持物体的标签为特征向量 A 对应的标签的概率；根据 k 个第三置信度计算得到特征向量 A 对应的第二置信度。

S805、向手持物体识别装置发送用于响应物体识别请求的响应消息，响应消息携带有

手持物体的识别结果。

在一个可行的实施例中，本实施例的方法还包括：

接收手持物体识别装置发送的第一消息，第一消息携带有手持物体的第三标签，将手持物体的第三标签和手持物体的第一特征向量保存至增量识别库中。

在一个可行的实施例中，本实施例的方法还包括：

接收手持物体识别装置发送的第二消息，第二消息携带有手持物体的第三标签和多帧第一图像，多帧第一图像为手持物体不同角度的图像，根据多帧第一图像获取手持物体的多个第三特征向量；多个第三特征向量与多帧第一图像一一对应；根据聚类算法对多个第三特征向量进行聚类，以得到 k 个子特征向量；将手持物体的第三标签和第二特征向量保存至增量识别库中，其中，第二特征向量是根据 k 个子特征向量得到的，或者第二特征向量包括 k 个子特征向量。

在此需要说明的是，S801-S805 的具体实现过程可参见上述步骤 S402 和 S403 的相关描述，在此不再叙述。

参见图 9，图 9 为本申请实施例提供的一种手持物体识别装置的结构示意图。如图 9 所示，该手持物体识别装置 900 包括：

获取单元 901，用于获取待识别图像中一个或多个检测对象中每个检测对象的位置信息，并获取每个检测对象的第一标签，每个检测对象的位置信息为该检测对象在待识别图像中的位置信息，检测对象的第一标签用于指示检测对象的类型，检测对象的类型用于表征检测对象的手持关系；根据每个检测对象的第一标签从一个或多个检测对象中获取手持物体，并从一个或多个检测对象的位置信息获取手持物体的位置信息；

识别单元 902，用于根据手持物体的位置信息对待识别图像中手持物体进行识别，以得到手持物体的识别结果。

在一个可行的实施例中，第一标签包括子类标签，子类标签用于指示检测对象的第一类型，在获取每个检测对象的第一标签的方面，获取单元 901 具体用于：

在获取待识别图像的第一特征向量后，对待识别图像的第一特征向量进行第一卷积操作，以得到待识别图像的第二特征向量；对待识别图像的第二特征向量进行第一全连接操作，以得到每个检测对象的第一标签，其中，第一类型包括手持状态下的物体和非手持状态下的物体。

在一个可行的实施例中，第一类型还包括持有物体的手和未持有物体的手。

在一个可行的实施例中，第一标签还包括超类标签，超类标签用于指示检测对象的第二类型，第一类型为第二类型的子类，在获取每个检测对象的第一标签的方面，获取单元 901 还用于：

在获取待识别图像的第一特征向量后，对第一特征向量进行第二卷积操作，以得到待识别图像的第三特征向量；将第二特征向量和第三特征向量进行融合操作，以得到待识别图像的第四特征向量；对第四特征向量进行第三卷积操作，以得到待识别图像的第五特征向量；对第五特征向量进行第二全连接操作，以得到每个检测对象的超类标签；其中，第二类型包括手和物体。

在一个可行的实施例中，在获取待识别图像中一个或多个检测对象中每个检测对象的

位置信息的方面，获取单元 901 具体用于：

在获取待识别图像的第一特征向量后，根据待识别图像的第一特征向量对待识别图像中一个或多个检测对象进行边框回归操作，以得到一个或多个检测对象中每个检测对象的位置信息。

在一个可行的实施例中，识别单元 902 具体用于：

根据手持物体的位置信息从待识别图像中裁剪出包含手持物体的图像块，图像块为待识别图像的一部分；对图像块中的手持物体进行特征提取，以得到手持物体的第一特征向量；根据手持物体的第一特征向量确定手持物体的参考标签及第一置信度，第一置信度为手持物体的标签为参考标签的概率；根据手持物体的参考标签和第一置信度获取手持物体的识别结果。

在一个可行的实施例中，识别结果包括手持物体的标签或者第二标签，第二标签用于指示未识别手持物体，增量特征库中包括 S 个特征向量及与 S 个特征向量中每个特征向量对应的标签， S 为大于或者等于 0 的整数，在根据手持物体的参考标签和第一置信度获取手持物体的识别结果的方面，识别单元 902 具体用于：

当第一置信度高于预设置信度时，将手持物体的参考标签确定为手持物体的标签；当第一置信度不高于预设置信度时，根据手持物体的第一特征向量和 S 个特征向量获取 S 个第二置信度， S 个第二置信度与 S 个特征向量一一对应， S 个特征向量中第 j 个特征向量所对应的第二置信度用于表征手持物体的标签为第 j 个特征向量对应的标签的概率；若 S 个第二置信度均低于预设置信度，则将第二标签确定为手持物体的标签；若 S 个第二置信度中存在高于预设置信度的置信度，则将目标特征向量对应的标签确定为手持物体的标签，目标特征向量为高于预设置信度中的最高置信度对应的特征向量。

在一个可行的实施例中， S 个特征向量中每个特征向量包括 k 个子特征向量，在根据手持物体的第一特征向量和 S 个特征向量获取 S 个第二置信度的方面，识别单元 902 具体用于：

对于 S 个特征向量中的任一特征向量 A ，根据手持物体的第一特征向量和特征向量 A 的 k 个子特征向量计算得到 k 个第三置信度；第三置信度用于表征手持物体的标签为特征向量 A 对应的标签的概率；根据 k 个第三置信度计算得到特征向量 A 对应的第二置信度。

在一个可行的实施例中，手持物体识别装置 900 还包括：

展示单元 903，用于在识别单元根据手持物体的第一特征向量确定手持物体的识别结果后，向用户展示手持物体的识别结果。

在一个可行的实施例中，手持物体识别装置 900 还包括：

接收单元 904，用于接收用户的第一信息，第一信息包括手持物体的第三标签；

保存单元 905，用于将手持物体的第三标签和手持物体的第一特征向量保存至增量识别库中。

在一个可行的实施例中，手持物体识别装置 900 还包括：

接收单元 904，用于接收用户的第二信息，第二信息用于指示用户翻转手持物体，且第二信息包括手持物体的第三标签；

获取单元 901，还用于获取手持物体的多帧第一图像，多帧第一图像为手持物体不同

角度的图像；根据多帧第一图像获取手持物体的多个第三特征向量；多个第三特征向量与多帧第一图像一一对应；

聚类单元 906，用于根据聚类算法对多个第三特征向量进行聚类，以得到 k 个子特征向量；

保存单元 905，用于将手持物体的第三标签和第二特征向量保存至增量识别库中，其中，第二特征向量是根据 k 个子特征向量得到的，或者第二特征向量包括 k 个子特征向量。

需要说明的是，上述各单元（获取单元 901、识别单元 902、展示单元 903、接收单元 904、保存单元 905 和聚类单元 906）用于执行上述步骤 S401-S403 的相关内容，在此不再叙述。

在本实施例中，手持物体识别装置 900 是以单元的形式来呈现。这里的“单元”可以指特定应用集成电路（application-specific integrated circuit, ASIC），执行一个或多个软件或固件程序的处理器和存储器，集成逻辑电路，和/或其他可以提供上述功能的器件。此外，以上获取单元 901、识别单元 902、保存单元 905 和聚类单元 906 可通过图 12 所示的手持物体识别装置的处理器 1201 来实现。

参见图 10，图 10 为本申请实施例提供的另一种手持物体识别装置的结构示意图。如图 10 所示，包括：

获取单元 1001，用于获取待识别图像中一个或多个检测对象中每个检测对象的位置信息，并获取每个检测对象的第一标签，每个检测对象的位置信息为该检测对象在待识别图像中的位置信息，检测对象的第一标签用于指示检测对象的类型，检测对象的类型用于表征检测对象的手持关系；

获取单元 1001，还用于根据每个检测对象的第一标签从一个或多个检测对象中获取手持物体，并从一个或多个检测对象的位置信息获取手持物体的位置信息；

裁剪单元 1002，用于根据手持物体的位置信息从待识别图像中裁剪出包含手持物体的图像块，图像块为待识别图像的一部分；

发送单元 1003，用于向物体识别服务器发送物体识别请求，物体识别请求携带有图像块；

接收单元 1004，用于接收物体识别服务器的响应消息，响应消息携带有手持物体的识别结果。

在一个可行的实施例中，第一标签包括子类标签，子类标签用于指示检测对象的第一类型，在获取每个检测对象的第一标签的方面，获取单元 1001 具体用于：

在获取待识别图像的第一特征向量后，对待识别图像的第一特征向量进行第一卷积操作，以得到待识别图像的第二特征向量；

对待识别图像的第二特征向量进行第一全连接操作，以得到每个检测对象的第一标签，其中，第一类型包括手持状态下的物体和非手持状态下的物体。

在一个可行的实施例中，第一类型还包括持有物体的手和未持有物体的手。

在一个可行的实施例中，第一标签还包括超类标签，超类标签用于指示检测对象的第二类型，第一类型为第二类型的子类，在获取每个检测对象的第一标签的方面，获取单元

1001 还具体用于：

在获取待识别图像的第一特征向量后，对第一特征向量进行第二卷积操作，以得到待识别图像的第三特征向量；将第二特征向量和第三特征向量进行融合操作，以得到待识别图像的第四特征向量；对第四特征向量进行第三卷积操作，以得到待识别图像的第五特征向量；对第五特征向量进行第二全连接操作，以得到每个检测对象的超类标签；其中，第二类型包括手和物体。

在一个可行的实施例中，在获取待识别图像中一个或多个检测对象中每个检测对象的位置信息的方面，获取单元 1001 还用于：

在获取待识别图像的第一特征向量后，根据待识别图像的第一特征向量对待识别图像中一个或多个检测对象进行边框回归操作，以得到一个或多个检测对象中每个检测对象的位置信息。

在一个可行的实施例中，手持物体识别装置 1000 还包括：

展示单元 1005，用于接收单元接收物体识别服务器的响应消息之后，向用户展示手持物体的识别结果。

在一个可行的实施例中，接收单元 1004，还用于接收用户的第一信息，第一信息包括手持物体的第三标签；

发送单元 1003，还用于向物体识别服务器发送第一消息，第一消息携带有手持物体的第三标签，第一消息用于指示物体识别服务器将手持物体的第三标签和手持物体的第一特征向量保存至增量识别库中。

在一个可行的实施例中，手持物体识别装置 1000 还包括：

接收单元 1004，还用于接收用户的第二信息，第二信息用于指示用户翻转手持物体，且第二信息包括手持物体的第三标签；

获取单元 1001，还用于获取手持物体的多帧第一图像，多帧第一图像为手持物体不同角度的图像；

发送单元 1003，还用于向物体识别服务器发送第二消息，第二消息携带有手持物体的第三标签和多帧第一图像，第二消息用于指示物体识别服务器根据多帧第一图像获取手持物体的第二特征向量，并将手持物体的第三标签和手持物体的第二特征向量保存至增量识别库中。

需要说明的是，上述各单元（获取单元 1001、裁剪单元 1002、发送单元 1003、接收单元 1004 和展示单元 1005）用于执行上述步骤 S701-S704 的相关内容，在此不再叙述。

在本实施例中，手持物体识别装置 1000 是以单元的形式来呈现。这里的“单元”可以指特定应用集成电路（application-specific integrated circuit, ASIC），执行一个或多个软件或固件程序的处理器和存储器，集成逻辑电路，和/或其他可以提供上述功能的器件。此外，以上获取单元 1001、裁剪单元 1002 和展示单元 1005 可通过图 12 所示的手持物体识别装置的处理器 1201 来实现。

参见图 11，图 11 为本申请实施例提供的一种物体识别服务器的结构示意图。如图 11 所示，物体识别服务器 1100 包括：

接收单元 1101, 用于接收手持物体识别装置发送的物体识别请求, 物体识别请求携带有包括手持物体的图像块;

特征提取单元 1102, 用于对图像块中的手持物体进行特征提取, 以得到手持物体的第一特征向量;

确定单元 1103, 用于根据手持物体的第一特征向量确定手持物体的参考标签及第一置信度, 第一置信度为手持物体的标签为参考标签的概率;

获取单元 1104, 用于根据手持物体的参考标签和第一置信度获取手持物体的识别结果;

发送单元 1105, 用于向手持物体识别装置发送用于响应物体识别请求的响应消息, 该响应消息携带有手持物体的识别结果。

在一个可行的实施例, 识别结果包括手持物体的标签或者第二标签, 第二标签用于指示未识别手持物体, 增量特征库中包括 S 个特征向量及与 S 个特征向量中每个特征向量对应的标签, S 为大于或者等于 0 的整数, 获取单元 1104 具体用于:

当第一置信度高于预设置信度时, 将手持物体的参考标签确定为手持物体的标签;

当第一置信度不高于预设置信度时, 根据手持物体的第一特征向量和 S 个特征向量获取 S 个第二置信度, S 个第二置信度与 S 个特征向量一一对应, S 个特征向量中第 j 个特征向量所对应的第二置信度用于表征手持物体的标签为第 j 个特征向量对应的标签的概率;

若 S 个第二置信度均低于预设置信度, 则将第二标签确定为手持物体的标签; 若 S 个第二置信度中存在高于预设置信度的置信度, 则将目标特征向量对应的标签确定为手持物体的标签, 目标特征向量为高于预设置信度中的最高置信度对应的特征向量。

在一个可行的实施例, S 个特征向量中每个特征向量包括 k 个子特征向量, 在根据手持物体的第一特征向量和 S 个特征向量获取 S 个第二置信度的方面, 获取单元 1104 具体用于:

对于 S 个特征向量中的任一特征向量 A , 根据手持物体的第一特征向量和特征向量 A 的 k 个子特征向量计算得到 k 个第三置信度; 第三置信度用于表征手持物体的标签为特征向量 A 对应的标签的概率; 根据 k 个第三置信度计算得到特征向量 A 对应的第二置信度。

在一个可行的实施例, 物体识别服务器 1100 还包括:

接收单元 1101, 还用于接收手持物体识别装置发送的第一消息, 第一消息携带有手持物体的第三标签,

保存单元 1106, 用于将手持物体的第三标签和手持物体的第一特征向量保存至增量识别库中。

在一个可行的实施例, 物体识别服务器 1100 还包括:

接收单元 1101, 还用于接收手持物体识别装置发送的第二消息, 第二消息携带有手持物体的第三标签和多帧第一图像, 多帧第一图像为手持物体不同角度的图像,

获取单元 1104, 还用于根据多帧第一图像获取手持物体的多个第三特征向量; 多个第三特征向量与多帧第一图像一一对应;

聚类单元 1107, 用于根据聚类算法对多个第三特征向量进行聚类, 以得到 k 个子特征向量;

保存单元 1106, 用于将手持物体的第三标签和第二特征向量保存至增量识别库中, 其

中,第二特征向量是根据 k 个子特征向量得到的,或者第二特征向量包括 k 个子特征向量。

需要说明的是,上述各单元(接收单元 1101、特征提取单元 1102、确定单元 1103、获取单元 1104、发送单元 1105、保存单元 1106 和聚类单元 1107)用于执行上述步骤 S801-S805 的相关内容,在此不再叙述。

在本实施例中,物体识别服务器 1100 是以单元的形式来呈现。这里的“单元”可以指特定应用集成电路(application-specific integrated circuit, ASIC),执行一个或多个软件或固件程序的处理器和存储器,集成逻辑电路,和/或其他可以提供上述功能的器件。此外,以上特征提取单元 1102、确定单元 1103、获取单元 1104、保存单元 1106 和聚类单元 1107 可通过图 13 所示的物体识别服务器的处理器 1301 来实现。

如图 12 所示手持物体识别装置 1200 可以以图 12 中的结构来实现,该手持物体识别装置 1200 包括至少一个处理器 1201、至少一个存储器 1202、至少一个通信接口 1203 以及至少一个显示屏 1204。所述处理器 1201、所述存储器 1202、所述通信接口 1203 和显示器 1204 通过所述通信总线连接并完成相互间的通信。

处理器 1201 可以是通用中央处理器(CPU)、微处理器、特定应用集成电路(application-specific integrated circuit, ASIC),或一个或多个用于控制以上方案程序执行的集成电路。

通信接口 1203,用于与其他设备或通信网络通信,如以太网,无线接入网(RAN),无线局域网(Wireless Local Area Networks, WLAN)等。

显示器 1024,可以用于显示处理器 1201 处理的结果,比如手持物体的识别结果等。

存储器 1202 可以是只读存储器(read-only memory, ROM)或可存储静态信息和指令的其他类型的静态存储设备,随机存取存储器(random access memory, RAM)或者可存储信息和指令的其他类型的动态存储设备,也可以是电可擦可编程只读存储器(Electrically Erasable Programmable Read-Only Memory, EEPROM)、只读光盘(Compact Disc Read-Only Memory, CD-ROM)或其他光盘存储、光碟存储(包括压缩光碟、激光碟、光碟、数字通用光碟、蓝光光碟等)、磁盘存储介质或者其他磁存储设备、或者能够用于携带或存储具有指令或数据结构形式的期望的程序代码并能够由计算机存取的任何其他介质,但不限于此。存储器可以是独立存在,通过总线与处理器相连接。存储器也可以和处理器集成在一起。

其中,所述存储器 1202 用于存储执行以上方案的应用程序代码,并由处理器 1201 来控制执行。所述处理器 1201 用于执行所述存储器 1202 中存储的应用程序代码,并在显示器 1204 上显示处理器 1201 的处理结果。

存储器 1202 存储的代码可执行以上提供的任一种手持物体识别方法,比如:

获取待识别图像中一个或多个检测对象中每个检测对象的位置信息,并获取每个检测对象的第一标签,每个检测对象的位置信息为该检测对象在待识别图像中的位置信息,检测对象的第一标签用于指示检测对象的类型,检测对象的类型用于表征检测对象的手持关系;根据每个检测对象的第一标签从一个或多个检测对象中获取手持物体,并从一个或多个检测对象的位置信息获取手持物体的位置信息;根据手持物体的位置信息对待识别图像中手持物体进行识别,以得到手持物体的识别结果;

或者,

获取待识别图像中一个或多个检测对象中每个检测对象的位置信息，并获取每个检测对象的第一标签，每个检测对象的位置信息为该检测对象在待识别图像中的位置信息，检测对象的第一标签用于指示检测对象的类型，检测对象的类型用于表征检测对象的手持关系；根据每个检测对象的第一标签从一个或多个检测对象中获取手持物体，并从一个或多个检测对象的位置信息获取手持物体的位置信息；根据手持物体的位置信息从待识别图像中裁剪出包含手持物体的图像块，图像块为待识别图像的一部分；向物体识别服务器发送物体识别请求，物体识别请求携带有图像块；接收物体识别服务器的响应消息，响应消息携带有手持物体的识别结果。

在此需要说明的是，对于上述手持物体识别方法的过程可参见 S301-S304 及 S801-S806 的相关描述，在此不再叙述。

如图 13 所示物体识别服务器 1300 可以以图 13 中的结构来实现，该物体识别服务器 1300 包括至少一个处理器 1301，至少一个存储器 1302 以及至少一个通信接口 1303。所述处理器 1301、所述存储器 1302 和所述通信接口 1303 通过所述通信总线连接并完成相互间的通信。

处理器 1301 可以是通用中央处理器（CPU），微处理器，特定应用集成电路（application-specific integrated circuit，ASIC），或一个或多个用于控制以上方案程序执行的集成电路。

通信接口 1303，用于与其他设备或通信网络通信，如以太网，无线接入网（RAN），无线局域网（Wireless Local Area Networks，WLAN）等。

存储器 1302 可以是只读存储器（read-only memory，ROM）或可存储静态信息和指令的其他类型的静态存储设备，随机存取存储器（random access memory，RAM）或者可存储信息和指令的其他类型的动态存储设备，也可以是电可擦可编程只读存储器（Electrically Erasable Programmable Read-Only Memory，EEPROM）、只读光盘（Compact Disc Read-Only Memory，CD-ROM）或其他光盘存储、光碟存储（包括压缩光碟、激光碟、光碟、数字通用光碟、蓝光光碟等）、磁盘存储介质或者其他磁存储设备、或者能够用于携带或存储具有指令或数据结构形式的期望的程序代码并能够由计算机存取的任何其他介质，但不限于此。存储器可以是独立存在，通过总线与处理器相连接。存储器也可以和处理器集成在一起。

其中，所述存储器 1302 用于存储执行以上方案的应用程序代码，并由处理器 1301 来控制执行。所述处理器 1301 用于执行所述存储器 1302 中存储的应用程序代码。

存储器 1302 存储的代码可执行以上提供的任一种手持物体识别方法，比如：

接收手持物体识别装置发送的物体识别请求，物体识别请求携带有包括手持物体的图像块；对图像块中的手持物体进行特征提取，以得到手持物体的第一特征向量；根据手持物体的第一特征向量确定手持物体的参考标签及第一置信度，第一置信度为手持物体的标签为参考标签的概率；根据手持物体的参考标签和第一置信度获取手持物体的识别结果；向手持物体识别装置发送用于响应物体识别请求的响应消息，该响应消息携带有手持物体的识别结果。

在此需要说明的是，对于上述手持物体识别方法的过程可参见 S401-S403、S701-S704 及 S801-S805 的相关描述，在此不再叙述。

本申请实施例还提供一种计算机存储介质，其中，该计算机存储介质可存储有程序，该程序执行时包括上述方法实施例中记载的任何一种手持物体识别方法的部分或全部步骤。

需要说明的是，对于前述的各方法实施例，为了简单描述，故将其都表述为一系列的动作组合，但是本领域技术人员应该知悉，本申请并不受所描述的动作顺序的限制，因为依据本申请，某些步骤可以采用其他顺序或者同时进行。其次，本领域技术人员也应该知悉，说明书中所描述的实施例均属于优选实施例，所涉及的动作和模块并不一定是本申请所必须的。

在上述实施例中，对各个实施例的描述都各有侧重，某个实施例中未详述的部分，可以参见其他实施例的相关描述。

在本申请所提供的几个实施例中，应该理解到，所揭露的装置，可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，例如所述单元的划分，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，可以是电性或其它的形式。

所述作为分离部件说明的单元可以是或者也可以不是物理上分开的，作为单元显示的部件可以是或者也可以不是物理单元，即可以位于一个地方，或者也可以分布到多个网络单元上。可以根据实际的需要选择其中的部分或者全部单元来实现本实施例方案的目的。

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。

所述集成的单元如果以软件功能单元的形式实现并作为独立的产品销售或使用，可以存储在一个计算机可读存储器中。基于这样的理解，本申请的技术方案本质上或者说对现有技术做出贡献的部分或者该技术方案的全部或部分可以以软件产品的形式体现出来，该计算机软件产品存储在一个存储器中，包括若干指令用以使得一台计算机设备（可为个人计算机、服务器或者网络设备等）执行本申请各个实施例所述方法的全部或部分步骤。而前述的存储器包括：U 盘、只读存储器（ROM，Read-Only Memory）、随机存取存储器（RAM，Random Access Memory）、移动硬盘、磁碟或者光盘等各种可以存储程序代码的介质。

本领域普通技术人员可以理解上述实施例的各种方法中的全部或部分步骤是可以通程序来指令相关的硬件来完成，该程序可以存储于一计算机可读存储器中，存储器可以包括：闪存盘、只读存储器（英文：Read-Only Memory，简称：ROM）、随机存取器（英文：Random Access Memory，简称：RAM）、磁盘或光盘等。

以上对本申请实施例进行了详细介绍，本文中应用了具体个例对本申请的原理及实施方式进行了阐述，以上实施例的说明只是用于帮助理解本申请的方法及其核心思想；同时，对于本领域的一般技术人员，依据本申请的思想，在具体实施方式及应用范围上均会有改变之处，综上所述，本说明书内容不应理解为对本申请的限制。

权利要求

1. 一种手持物体识别方法，其特征在于，包括：

获取待识别图像中一个或多个检测对象中每个检测对象的位置信息，并获取所述每个检测对象的第一标签，所述每个检测对象的位置信息为该检测对象在所述待识别图像中的位置信息，所述检测对象的第一标签用于指示所述检测对象的类型，所述检测对象的类型用于表征所述检测对象的手持关系；

根据所述每个检测对象的第一标签从所述一个或多个检测对象中获取所述手持物体，并从所述一个或多个检测对象的位置信息获取所述手持物体的位置信息；

根据所述手持物体的位置信息对所述待识别图像中所述手持物体进行识别，以得到所述手持物体的识别结果。

2. 根据权利要求1所述的方法，其特征在于，所述第一标签包括子类标签，所述子类标签用于指示所述检测对象的第一类型，所述获取所述每个检测对象的第一标签，包括：

在获取所述待识别图像的第一特征向量后，对所述待识别图像的第一特征向量进行第一卷积操作，以得到所述待识别图像的第二特征向量；

对所述待识别图像的第二特征向量进行第一全连接操作，以得到所述每个检测对象的第一标签，其中，所述第一类型包括手持状态下的物体和非手持状态下的物体。

3. 根据权利要求2所述的方法，其特征在于，所述第一类型还包括持有物体的手和未持有物体的手。

4. 根据权利要求3所述的方法，其特征在于，所述第一标签还包括超类标签，所述超类标签用于指示所述检测对象的第二类型，所述第一类型为所述第二类型的子类，所述获取所述每个检测对象的第一标签还包括：

在获取所述待识别图像的第一特征向量后，对所述第一特征向量进行第二卷积操作，以得到所述待识别图像的第三特征向量；

将所述第二特征向量和所述第三特征向量进行融合操作，以得到所述待识别图像的第四特征向量；

对所述第四特征向量进行第三卷积操作，以得到所述待识别图像的第五特征向量；

对所述第五特征向量进行第二全连接操作，以得到所述每个检测对象的超类标签；其中，第二类型包括手和物体。

5. 根据权利要求1-4任一项所述的方法，其特征在于，所述获取所述待识别图像中一个或多个检测对象中每个检测对象的位置信息，包括：

在获取待识别图像的第一特征向量后，根据所述待识别图像的第一特征向量对所述待识别图像中一个或多个检测对象进行边框回归操作，以得到所述一个或多个检测对象中每个检测对象的位置信息。

6. 根据权利要求 1-5 任一项所述的方法, 其特征在于, 所述根据所述手持物体的位置信息对所述待识别图像中所述手持物体进行识别, 以得到所述手持物体的识别结果, 包括:

根据所述手持物体的位置信息从所述待识别图像中裁剪出包含所述手持物体的图像块, 所述图像块为所述待识别图像的一部分;

对所述图像块中的所述手持物体进行特征提取, 以得到所述手持物体的第一特征向量;

根据所述手持物体的第一特征向量确定所述手持物体的参考标签及第一置信度, 所述第一置信度为所述手持物体的标签为所述参考标签的概率;

根据所述手持物体的参考标签和第一置信度获取所述手持物体的识别结果。

7. 根据权利要求 6 所述的方法, 其特征在于, 所述识别结果包括所述手持物体的标签或者第二标签, 所述第二标签用于指示未识别所述手持物体, 增量特征库中包括 S 个特征向量及与 S 个特征向量中每个特征向量对应的标签, 所述 S 为大于或者等于 0 的整数, 所述根据所述手持物体的参考标签和第一置信度获取所述手持物体的识别结果, 包括:

当所述第一置信度高于预设置信度时, 将所述手持物体的参考标签确定为所述手持物体的标签;

当所述第一置信度不高于所述预设置信度时, 根据所述手持物体的第一特征向量和所述 S 个特征向量获取 S 个第二置信度, 所述 S 个第二置信度与所述 S 个特征向量一一对应, 所述 S 个特征向量中第 j 个特征向量所对应的第二置信度用于表征所述手持物体的标签为所述第 j 个特征向量对应的标签的概率;

若所述 S 个第二置信度均低于所述预设置信度, 则将所述第二标签确定为所述手持物体的标签; 若 S 个第二置信度中存在高于所述预设置信度的置信度, 则将目标特征向量对应的标签确定为所述手持物体的标签, 所述目标特征向量为所述高于所述预设置信度中的最高置信度对应的特征向量。

8. 根据权利要求 7 所述的方法, 其特征在于, 所述 S 个特征向量中每个特征向量包括 k 个子特征向量, 所述根据所述手持物体的第一特征向量和所述 S 个特征向量获取 S 个第二置信度, 包括:

对于所述 S 个特征向量中的任一特征向量 A , 根据所述手持物体的第一特征向量和所述特征向量 A 的 k 个子特征向量计算得到 k 个第三置信度; 所述第三置信度用于表征所述手持物体的标签为所述特征向量 A 对应的标签的概率;

根据所述 k 个第三置信度计算得到所述特征向量 A 对应的第二置信度。

9. 根据权利要求 7 或 8 所述的方法, 其特征在于, 所述根据所述手持物体的第一特征向量确定所述手持物体的识别结果后, 所述方法还包括:

向用户展示所述手持物体的识别结果。

10. 根据权利要求 9 所述的方法, 其特征在于, 所述方法还包括:

接收所述用户的第一信息，所述第一信息包括所述手持物体的第三标签；
将所述手持物体的第三标签和所述手持物体的第一特征向量保存至所述增量识别库中。

11. 根据权利要求 9 所述的方法，其特征在于，所述方法还包括：

接收所述用户的第二信息，所述第二信息用于指示所述用户翻转所述手持物体，且所述第二信息包括所述手持物体的第三标签；

获取所述手持物体的多帧第一图像，所述多帧第一图像为所述手持物体不同角度的图像；

根据所述多帧第一图像获取所述手持物体的多个第三特征向量；所述多个第三特征向量与所述多帧第一图像一一对应；

根据聚类算法对所述多个第三特征向量进行聚类，以得到 k 个子特征向量；

将所述手持物体的第三标签和第二特征向量保存至所述增量识别库中，其中，所述第二特征向量是根据 k 个子特征向量得到的，或者所述第二特征向量包括所述 k 个子特征向量。

12. 一种手持物体识别方法，其特征在于，包括：

获取待识别图像中一个或多个检测对象中每个检测对象的位置信息，并获取所述每个检测对象的第一标签，所述每个检测对象的位置信息为该检测对象在所述待识别图像中的位置信息，所述检测对象的第一标签用于指示所述检测对象的类型，所述检测对象的类型用于表征所述检测对象的手持关系；

根据所述每个检测对象的第一标签从所述一个或多个检测对象中获取所述手持物体，并从所述一个或多个检测对象的位置信息获取所述手持物体的位置信息；

根据所述手持物体的位置信息从所述待识别图像中裁剪出包含所述手持物体的图像块，所述图像块为所述待识别图像的一部分；

向物体识别服务器发送物体识别请求，所述物体识别请求携带有所述图像块；

接收所述物体识别服务器的响应消息，所述响应消息携带有所述手持物体的识别结果。

13. 根据权利要求 12 所述的方法，其特征在于，所述第一标签包括子类标签，所述子类标签用于指示所述检测对象的第一类型，所述获取所述每个检测对象的第一标签，包括：

在获取所述待识别图像的第一特征向量后，对所述待识别图像的第一特征向量进行第一卷积操作，以得到所述待识别图像的第二特征向量；

对所述待识别图像的第二特征向量进行第一全连接操作，以得到所述每个检测对象的第一标签，其中，所述第一类型包括手持状态下的物体和非手持状态下的物体。

14. 根据权利要求 13 所述的方法，其特征在于，所述第一类型还包括持有物体的手和未持有物体的手。

15. 根据权利要求 14 所述的方法，其特征在于，所述第一标签还包括超类标签，所述

超类标签用于指示所述检测对象的第二类型，所述第一类型为所述第二类型的子类，所述获取所述每个检测对象的第一标签，还包括：

在获取所述待识别图像的第一特征向量后，对所述第一特征向量进行第二卷积操作，以得到所述待识别图像的第三特征向量；

将所述第二特征向量和所述第三特征向量进行融合操作，以得到所述待识别图像的第四特征向量；

对所述第四特征向量进行第三卷积操作，以得到所述待识别图像的第五特征向量；

对所述第五特征向量进行第二全连接操作，以得到所述每个检测对象的超类标签；其中，第二类型包括手和物体。

16. 根据权利要求 12-15 任一项所述的方法，其特征在于，所述获取所述待识别图像中一个或多个检测对象中每个检测对象的位置信息，包括：

在获取所述待识别图像的第一特征向量后，根据所述待识别图像的第一特征向量对所述待识别图像中一个或多个检测对象进行边框回归操作，以得到所述一个或多个检测对象中每个检测对象的位置信息。

17. 根据权利要求 12-16 任一项所述的方法，其特征在于，所述接收所述物体识别服务器的响应消息之后，所述方法还包括：

向用户展示所述手持物体的识别结果。

18. 根据权利要求 17 所述的方法，其特征在于，所述方法还包括：

接收所述用户的第一信息，所述第一信息包括所述手持物体的第三标签；

向所述物体识别服务器发送第一消息，所述第一消息携带有所述手持物体的第三标签，所述第一消息用于指示所述物体识别服务器将所述手持物体的第三标签和所述手持物体的第一特征向量保存至所述增量识别库中。

19. 根据权利要求 17 所述的方法，其特征在于，所述方法还包括：

接收所述用户的第二信息，所述第二信息用于指示所述用户翻转所述手持物体，且所述第二信息包括所述手持物体的第三标签；

获取所述手持物体的多帧第一图像，所述多帧第一图像为所述手持物体不同角度的图像；

向所述物体识别服务器发送第二消息，所述第二消息携带有所述手持物体的第三标签和所述多帧第一图像，所述第二消息用于指示所述物体识别服务器根据所述多帧第一图像获取所述手持物体的第二特征向量，并将所述手持物体的第三标签和所述手持物体的第二特征向量保存至所述增量识别库中。

20. 一种手持物体识别方法，其特征在于，包括：

接收手持物体识别装置发送的物体识别请求，所述物体识别请求携带有包括手持物体

的图像块；

对所述图像块中的所述手持物体进行特征提取，以得到所述手持物体的第一特征向量；

根据所述手持物体的第一特征向量确定所述手持物体的参考标签及第一置信度，所述第一置信度为所述手持物体的标签为所述参考标签的概率；

根据所述手持物体的参考标签和第一置信度获取所述手持物体的识别结果；

向所述手持物体识别装置发送用于响应所述物体识别请求的响应消息，所述响应消息携带有所述手持物体的识别结果。

21. 根据权利要求 20 所述的方法，其特征在于，所述识别结果包括所述手持物体的标签或者第二标签，所述第二标签用于指示未识别所述手持物体，增量特征库中包括 S 个特征向量及与 S 个特征向量中每个特征向量对应的标签，所述 S 为大于或者等于 0 的整数，所述根据所述手持物体的参考标签和第一置信度获取所述手持物体的识别结果，包括：

当所述第一置信度高于预设置信度时，将所述手持物体的参考标签确定为所述手持物体的标签；

当所述第一置信度不高于所述预设置信度时，根据所述手持物体的第一特征向量和所述 S 个特征向量获取 S 个第二置信度，所述 S 个第二置信度与所述 S 个特征向量一一对应，所述 S 个特征向量中第 j 个特征向量所对应的第二置信度用于表征所述手持物体的标签为所述第 j 个特征向量对应的标签的概率；

若所述 S 个第二置信度均低于所述预设置信度，则将所述第二标签确定为所述手持物体的标签；若 S 个第二置信度中存在高于所述预设置信度的置信度，则将目标特征向量对应的标签确定为所述手持物体的标签，所述目标特征向量为所述高于所述预设置信度中的最高置信度对应的特征向量。

22. 根据权利要求 21 所述的方法，其特征在于，所述 S 个特征向量中每个特征向量包括 k 个子特征向量，所述根据所述手持物体的第一特征向量和所述 S 个特征向量获取 S 个第二置信度，包括：

对于所述 S 个特征向量中的任一特征向量 A，根据所述手持物体的第一特征向量和所述特征向量 A 的 k 个子特征向量计算得到 k 个第三置信度；所述第三置信度用于表征所述手持物体的标签为所述特征向量 A 对应的标签的概率；

根据所述 k 个第三置信度计算得到所述特征向量 A 对应的第二置信度。

23. 根据权利要求 20-22 任一项所述的方法，其特征在于，所述方法还包括：

接收所述手持物体识别装置发送的第一消息，所述第一消息携带有所述手持物体的第三标签，

将所述手持物体的第三标签和所述手持物体的第一特征向量保存至所述增量识别库中。

24. 根据权利要求 20-22 任一项所述的方法，其特征在于，所述方法还包括：

接收所述手持物体识别装置发送的第二消息，所述第二消息携带有所述手持物体的第

三标签和多帧第一图像，所述多帧第一图像为所述手持物体不同角度的图像，

根据所述多帧第一图像获取所述手持物体的多个第三特征向量；所述多个第三特征向量与所述多帧第一图像一一对应；

根据聚类算法对所述多个第三特征向量进行聚类，以得到 k 个子特征向量；

将所述手持物体的第三标签和第二特征向量保存至所述增量识别库中，其中，所述第二特征向量是根据 k 个子特征向量得到的，或者所述第二特征向量包括所述 k 个子特征向量。

25. 一种手持物体识别装置，其特征在于，包括：

获取单元，用于获取待识别图像中一个或多个检测对象中每个检测对象的位置信息，并获取所述每个检测对象的第一标签，所述每个检测对象的位置信息为该检测对象在所述待识别图像中的位置信息，所述检测对象的第一标签用于指示所述检测对象的类型，所述检测对象的类型用于表征所述检测对象的手持关系；

所述获取单元，还用于根据所述每个检测对象的第一标签从所述一个或多个检测对象中获取所述手持物体，并从所述一个或多个检测对象的位置信息获取所述手持物体的位置信息；

识别单元，用于根据所述手持物体的位置信息对所述待识别图像中所述手持物体进行识别，以得到所述手持物体的识别结果。

26. 根据权利要求 25 所述的装置，其特征在于，所述第一标签包括子类标签，所述子类标签用于指示所述检测对象的第一类型，在所述获取所述每个检测对象的第一标签的方面，所述获取单元具体用于：

在获取所述待识别图像的第一特征向量后，对所述待识别图像的第一特征向量进行第一卷积操作，以得到所述待识别图像的第二特征向量；

对所述待识别图像的第二特征向量进行第一全连接操作，以得到所述每个检测对象的第一标签，其中，所述第一类型包括手持状态下的物体和非手持状态下的物体。

27. 根据权利要求 26 所述的装置，其特征在于，所述第一类型还包括持有物体的手和未持有物体的手。

28. 根据权利要求 27 所述的装置，其特征在于，所述第一标签还包括超类标签，所述超类标签用于指示所述检测对象的第二类型，所述第一类型为所述第二类型的子类，在所述获取所述每个检测对象的第一标签的方面，所述获取单元还用于：

在获取所述待识别图像的第一特征向量后，对所述第一特征向量进行第二卷积操作，以得到所述待识别图像的第三特征向量；

将所述第二特征向量和所述第三特征向量进行融合操作，以得到所述待识别图像的第四特征向量；

对所述第四特征向量进行第三卷积操作，以得到所述待识别图像的第五特征向量；

对所述第五特征向量进行第二全连接操作，以得到所述每个检测对象的超类标签；其中，第二类型包括手和物体。

29. 根据权利要求 25-28 任一项所述的装置，其特征在于，在所述获取所述待识别图像中一个或多个检测对象中每个检测对象的位置信息的方面，所述获取单元具体用于：

在获取待识别图像的第一特征向量后，根据所述待识别图像的第一特征向量对所述待识别图像中一个或多个检测对象进行边框回归操作，以得到所述一个或多个检测对象中每个检测对象的位置信息。

30. 根据权利要求 25-29 任一项所述的装置，其特征在于，所述识别单元具体用于：

根据所述手持物体的位置信息从所述待识别图像中裁剪出包含所述手持物体的图像块，所述图像块为所述待识别图像的一部分；

对所述图像块中的所述手持物体进行特征提取，以得到所述手持物体的第一特征向量；

根据所述手持物体的第一特征向量确定所述手持物体的参考标签及第一置信度，所述第一置信度为所述手持物体的标签为所述参考标签的概率；

根据所述手持物体的参考标签和第一置信度获取所述手持物体的识别结果。

31. 根据权利要求 30 所述的装置，其特征在于，所述识别结果包括所述手持物体的标签或者第二标签，所述第二标签用于指示未识别所述手持物体，增量特征库中包括 S 个特征向量及与 S 个特征向量中每个特征向量对应的标签，所述 S 为大于或者等于 0 的整数，在所述根据所述手持物体的参考标签和第一置信度获取所述手持物体的识别结果的方面，所述识别单元具体用于：

当所述第一置信度高于预设置信度时，将所述手持物体的参考标签确定为所述手持物体的标签；

当所述第一置信度不高于所述预设置信度时，根据所述手持物体的第一特征向量和所述 S 个特征向量获取 S 个第二置信度，所述 S 个第二置信度与所述 S 个特征向量一一对应，所述 S 个特征向量中第 j 个特征向量所对应的第二置信度用于表征所述手持物体的标签为所述第 j 个特征向量对应的标签的概率；

若所述 S 个第二置信度均低于所述预设置信度，则将所述第二标签确定为所述手持物体的标签；若 S 个第二置信度中存在高于所述预设置信度的置信度，则将目标特征向量对应的标签确定为所述手持物体的标签，所述目标特征向量为所述高于所述预设置信度中的最高置信度对应的特征向量。

32. 根据权利要求 31 所述的装置，其特征在于，所述 S 个特征向量中每个特征向量包括 k 个子特征向量，在所述根据所述手持物体的第一特征向量和所述 S 个特征向量获取 S 个第二置信度的方面，所述识别单元具体用于：

对于所述 S 个特征向量中的任一特征向量 A ，根据所述手持物体的第一特征向量和所述特征向量 A 的 k 个子特征向量计算得到 k 个第三置信度；所述第三置信度用于表征所述

手持物体的标签为所述特征向量 A 对应的标签的概率；

根据所述 k 个第三置信度计算得到所述特征向量 A 对应的第二置信度。

33. 根据权利要求 31 或 32 所述的装置，其特征在于，所述装置还包括：

展示单元，用于在所述识别单元根据所述手持物体的第一特征向量确定所述手持物体的识别结果后，向用户展示所述手持物体的识别结果。

34. 根据权利要求 33 所述的装置，其特征在于，所述装置还包括：

接收单元，用于接收所述用户的第一信息，所述第一信息包括所述手持物体的第三标签；

保存单元，用于将所述手持物体的第三标签和所述手持物体的第一特征向量保存至所述增量识别库中。

35. 根据权利要求 33 所述的装置，其特征在于，所述装置还包括：

接收单元，用于接收所述用户的第二信息，所述第二信息用于指示所述用户翻转所述手持物体，且所述第二信息包括所述手持物体的第三标签；

所述获取单元，还用于获取所述手持物体的多帧第一图像，所述多帧第一图像为所述手持物体不同角度的图像；根据所述多帧第一图像获取所述手持物体的多个第三特征向量；所述多个第三特征向量与所述多帧第一图像一一对应；

聚类单元，用于根据聚类算法对所述多个第三特征向量进行聚类，以得到 k 个子特征向量；

保存单元，用于将所述手持物体的第三标签和第二特征向量保存至所述增量识别库中，其中，所述第二特征向量是根据 k 个子特征向量得到的，或者所述第二特征向量包括所述 k 个子特征向量。

36. 一种手持物体识别装置，其特征在于，包括：

获取单元，用于获取待识别图像中一个或多个检测对象中每个检测对象的位置信息，并获取所述每个检测对象的第一标签，所述每个检测对象的位置信息为该检测对象在所述待识别图像中的位置信息，所述检测对象的第一标签用于指示所述检测对象的类型，所述检测对象的类型用于表征所述检测对象的手持关系；

所述获取单元，还用于根据所述每个检测对象的第一标签从所述一个或多个检测对象中获取所述手持物体，并从所述一个或多个检测对象的位置信息获取所述手持物体的位置信息；

裁剪单元，用于根据所述手持物体的位置信息从所述待识别图像中裁剪出包含所述手持物体的图像块，所述图像块为所述待识别图像的一部分；

发送单元，用于向物体识别服务器发送物体识别请求，所述物体识别请求携带有所述图像块；

接收单元，用于接收所述物体识别服务器的响应消息，所述响应消息携带有所述手持

物体的识别结果。

37. 根据权利要求 36 所述的装置，其特征在于，所述第一标签包括子类标签，所述子类标签用于指示所述检测对象的第一类型，在所述获取所述每个检测对象的第一标签的方面，所述获取单元具体用于：

在获取所述待识别图像的第一特征向量后，对所述待识别图像的第一特征向量进行第一卷积操作，以得到所述待识别图像的第二特征向量；

对所述待识别图像的第二特征向量进行第一全连接操作，以得到所述每个检测对象的第一标签，其中，所述第一类型包括手持状态下的物体和非手持状态下的物体。

38. 根据权利要求 37 所述的装置，其特征在于，所述第一类型还包括持有物体的手和未持有物体的手。

39. 根据权利要求 38 所述的装置，其特征在于，所述第一标签还包括超类标签，所述超类标签用于指示所述检测对象的第二类型，所述第一类型为所述第二类型的子类，在所述获取所述每个检测对象的第一标签的方面，所述获取单元还具体用于：

在获取所述待识别图像的第一特征向量后，对所述第一特征向量进行第二卷积操作，以得到所述待识别图像的第三特征向量；

将所述第二特征向量和所述第三特征向量进行融合操作，以得到所述待识别图像的第四特征向量；

对所述第四特征向量进行第三卷积操作，以得到所述待识别图像的第五特征向量；

对所述第五特征向量进行第二全连接操作，以得到所述每个检测对象的超类标签；其中，所述第二类型包括手和物体。

40. 根据权利要求 36-39 任一项所述的装置，其特征在于，在所述获取所述待识别图像中一个或多个检测对象中每个检测对象的位置信息的方面，所述获取单元还用于：

在获取待识别图像的第一特征向量后，根据所述待识别图像的第一特征向量对所述待识别图像中一个或多个检测对象进行边框回归操作，以得到所述一个或多个检测对象中每个检测对象的位置信息。

41. 根据权利要求 36-40 任一项所述的装置，其特征在于，所述装置还包括：

展示单元，用于所述接收单元接收所述物体识别服务器的响应消息之后，向用户展示所述手持物体的识别结果。

42. 根据权利要求 41 所述的装置，其特征在于，

所述接收单元，还用于接收所述用户的第一信息，所述第一信息包括所述手持物体的第三标签；

所述发送单元，还用于向所述物体识别服务器发送第一消息，所述第一消息携带有所

述手持物体的第三标签，所述第一消息用于指示所述物体识别服务器将所述手持物体的第三标签和所述手持物体的第一特征向量保存至所述增量识别库中。

43. 根据权利要求 41 所述的装置，其特征在于，所述装置还包括：

所述接收单元，还用于接收所述用户的第二信息，所述第二信息用于指示所述用户翻转所述手持物体，且所述第二信息包括所述手持物体的第三标签；

所述获取单元，还用于获取所述手持物体的多帧第一图像，所述多帧第一图像为所述手持物体不同角度的图像；

所述发送单元，还用于向所述物体识别服务器发送第二消息，所述第二消息携带有所述手持物体的第三标签和所述多帧第一图像，所述第二消息用于指示所述物体识别服务器根据所述多帧第一图像获取所述手持物体的第二特征向量，并将所述手持物体的第三标签和所述手持物体的第二特征向量保存至所述增量识别库中。

44. 一种物体识别服务器，其特征在于，包括：

接收单元，用于接收手持物体识别装置发送的物体识别请求，所述物体识别请求携带有包括手持物体的图像块；

特征提取单元，用于对所述图像块中的所述手持物体进行特征提取，以得到所述手持物体的第一特征向量；

确定单元，用于根据所述手持物体的第一特征向量确定所述手持物体的参考标签及第一置信度，所述第一置信度为所述手持物体的标签为所述参考标签的概率；

获取单元，用于根据所述手持物体的参考标签和第一置信度获取所述手持物体的识别结果；

发送单元，用于向所述手持物体识别装置发送用于响应所述物体识别请求的响应消息，所述响应消息携带有所述手持物体的识别结果。

45. 根据权利要求 44 所述的服务器，其特征在于，所述识别结果包括所述手持物体的标签或者第二标签，所述第二标签用于指示未识别所述手持物体，增量特征库中包括 S 个特征向量及与 S 个特征向量中每个特征向量对应的标签，所述 S 为大于或者等于 0 的整数，所述获取单元具体用于：

当所述第一置信度高于预设置信度时，将所述手持物体的参考标签确定为所述手持物体的标签；

当所述第一置信度不高于所述预设置信度时，根据所述手持物体的第一特征向量和所述 S 个特征向量获取 S 个第二置信度，所述 S 个第二置信度与所述 S 个特征向量一一对应，所述 S 个特征向量中第 j 个特征向量所对应的第二置信度用于表征所述手持物体的标签为所述第 j 个特征向量对应的标签的概率；

若所述 S 个第二置信度均低于所述预设置信度，则将所述第二标签确定为所述手持物体的标签；若 S 个第二置信度中存在高于所述预设置信度的置信度，则将目标特征向量对应的标签确定为所述手持物体的标签，所述目标特征向量为所述高于所述预设置信度中的

最高置信度对应的特征向量。

46. 根据权利要求 45 所述的服务器，其特征在于，所述 S 个特征向量中每个特征向量包括 k 个子特征向量，在所述根据所述手持物体的第一特征向量和所述 S 个特征向量获取 S 个第二置信度的方面，所述获取单元具体用于：

对于所述 S 个特征向量中的任一特征向量 A，根据所述手持物体的第一特征向量和所述特征向量 A 的 k 个子特征向量计算得到 k 个第三置信度；所述第三置信度用于表征所述手持物体的标签为所述特征向量 A 对应的标签的概率；

根据所述 k 个第三置信度计算得到所述特征向量 A 对应的第二置信度。

47. 根据权利要求 44-46 任一项所述的服务器，其特征在于，所述服务器还包括：

所述接收单元，还用于接收所述手持物体识别装置发送的第一消息，所述第一消息携带有所述手持物体的第三标签，

保存单元，用于将所述手持物体的第三标签和所述手持物体的第一特征向量保存至所述增量识别库中。

48. 根据权利要求 44-46 任一项所述的服务器，其特征在于，所述服务器还包括：

所述接收单元，还用于接收所述手持物体识别装置发送的第二消息，所述第二消息携带有所述手持物体的第三标签和多帧第一图像，所述多帧第一图像为所述手持物体不同角度的图像，

所述获取单元，还用于根据所述多帧第一图像获取所述手持物体的多个第三特征向量；所述多个第三特征向量与所述多帧第一图像一一对应；

聚类单元，用于根据聚类算法对所述多个第三特征向量进行聚类，以得到 k 个子特征向量；

保存单元，用于将所述手持物体的第三标签和第二特征向量保存至所述增量识别库中，其中，所述第二特征向量是根据所述 k 个子特征向量得到的，或者所述第二特征向量包括所述 k 个子特征向量。

49. 一种手持物体识别装置，其特征在于，包括：

存储器，用于存储指令；以及

处理器，与所述存储器耦合；

其中，当所述处理器执行所述指令时，执行如权利要求 1-19 任一项所述的方法。

50. 一种物体识别服务器，其特征在于，包括：

存储器，用于存储指令；以及

处理器，与所述存储器耦合；

其中，当所述处理器执行所述指令时，执行如权利要求 20-24 任一项所述的方法。

51. 一种芯片系统，其特征在于，所述芯片系统应用于电子设备；所述芯片系统包括一个或多个接口电路，以及一个或多个处理器；所述接口电路和所述处理器通过线路互联；所述接口电路用于从所述电子设备的存储器接收信号，并向所述处理器发送所述信号，所述信号包括所述存储器中存储的计算机指令；当所述处理器执行所述计算机指令时，所述电子设备执行如权利要求 1-24 中任意一项所述方法。

52. 一种计算机可读存储介质，其特征在于，所述计算机可读存储介质存储有计算机程序，所述计算机程序包括程序指令，所述程序指令当被处理器执行时使所述处理器执行如权利要求 1-24 任一项所述的方法。

53. 一种计算机程序产品，其特征在于，包括计算机指令，当所述计算机指令在电子设备上运行时，使得所述电子设备执行如权利要求 1-24 任一项所述的方法。