

(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
2020 年 1 月 16 日 (16.01.2020)



(10) 国际公布号
WO 2020/011014 A1

(51) 国际专利分类号:
G06K 9/00 (2006.01) *G06N 3/04* (2006.01)

(21) 国际申请号: PCT/CN2019/092791

(22) 国际申请日: 2019 年 6 月 25 日 (25.06.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201810771226.X 2018 年 7 月 13 日 (13.07.2018) CN

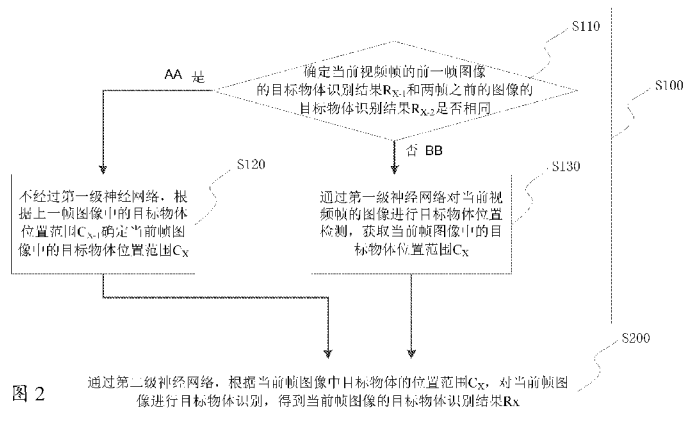
(71) 申请人: 腾讯科技(深圳)有限公司 (TENCENT TECHNOLOGY (SHENZHEN) COMPANY LIMITED) [CN/CN]; 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。

(72) 发明人: 程君(CHENG, Jun); 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。 尚海豹(SHANG, Haibao); 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。 李峰(LI, Feng); 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。 李昊沅(LI, Haoyuan); 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。 左小祥(ZUO, Xiaoxiang); 中国广东省深圳市南山区高新区科技中一路腾讯大厦 35 层, Guangdong 518057 (CN)。

(74) 代理人: 北京三高永信知识产权代理有限公司(BEIJING SAN GAO YONG XIN INTELLECTUAL PROPERTY AGENCY CO., LTD.); 中国北京市

(54) Title: METHOD AND SYSTEM FOR DETECTING AND RECOGNIZING OBJECT IN REAL-TIME VIDEO, STORAGE MEDIUM AND DEVICE

(54) 发明名称: 实时视频的目标检测与识别方法、系统、存储介质及设备



S110 DETERMINE WHETHER A TARGET OBJECT RECOGNITION RESULT R_{X-1} OF THE PREVIOUS IMAGE OF THE CURRENT VIDEO FRAME IS IDENTICAL TO A TARGET OBJECT RECOGNITION RESULT R_{X-2} OF AN IMAGE PRIOR TO THE BOTH IMAGES

S120 DETERMINE, NOT BY MEANS OF A FIRST-LEVEL NEURAL NETWORK, THE POSITION RANGE C_X OF THE TARGET OBJECT IN THE CURRENT IMAGE, ACCORDING TO THE POSITION RANGE C_{X-1} OF THE TARGET OBJECT IN THE PREVIOUS IMAGE

S130 PERFORM, BY MEANS OF THE FIRST-LEVEL NEURAL NETWORK, A TARGET OBJECT DETECTION ON THE IMAGE OF THE CURRENT VIDEO FRAME, TO ACQUIRE THE POSITION RANGE C_X OF THE TARGET OBJECT IN THE CURRENT IMAGE

S200 PERFORM, BY MEANS OF A SECOND-LEVEL NEURAL NETWORK, TARGET OBJECT RECOGNITION ON THE CURRENT IMAGE ACCORDING TO THE POSITION RANGE C_X OF THE TARGET OBJECT IN THE CURRENT IMAGE, SO AS TO OBTAIN A TARGET OBJECT RECOGNITION RESULT R_X OF THE CURRENT IMAGE

AA YES
BB NO

(57) Abstract: The present application proposes a method and a system for detecting and recognizing an object in a real-time video. The method comprises: determining whether a target object recognition result R_{X-1} of the previous image of the current image is identical to a target object recognition result R_{X-2} of an image prior to the both images; if not, detecting the position of a target object in the current image by means of a first-level neural network, to acquire a position range C_X of the target object in the current image; if so, determining,

海淀区学院路蓟门里和景园A座1单元
102室, Beijing 100088 (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, 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, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, 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))。

according to the position range C_{X-1} of the target object in the previous image, the position range C_X of the target object in the current image; and performing, by means of a second-level neural network, target object recognition on the current image according to the position range C_X . Thus, the present invention can reduce the operation frequency of the first-level neural network for position detection, speed up the recognition and reduce the occupation of CPU and memory resources.

(57) 摘要: 本申请提出了一种用于实时视频的目标检测与识别的方法和系统。方法包括: 确定当前帧图像的前一帧图像的目标物体识别结果 R_{X-1} 和两帧之前的图像的目标物体识别结果 R_{X-2} 是否相同; 不相同, 通过第一级神经网络对当前帧图像进行目标物体位置检测, 获取当前帧图像中目标物体的位置范围 C_X ; 相同时, 根据前一帧图像中目标物体的位置范围 C_{X-1} 确定当前帧图像中目标物体的位置范围 C_X ; 以及通过第二级神经网络, 根据位置范围 C_X , 对当前帧图像进行目标物体识别。从而, 能够降低用于位置检测的第一级神经网络的运行频率, 加快识别速度并降低CPU和内存资源的占用。

实时视频的目标检测与识别方法、系统、存储介质及设备

本申请要求于 2018 年 07 月 13 日提交的申请号为 201810771226.X、发明名称为“一种实时视频的目标检测与识别方法和系统”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

本申请涉及人工智能和图像识别技术领域，尤其涉及一种实时视频的目标检测与识别方法、系统、存储介质及设备。

背景技术

随着识别精度地不断提高，用 CNN（Convolutional Neural Networks，卷积神经网络）来进行图像识别和分类的方法已逐渐被业界接受。在评价模型或者算法的性能时，则多采用 ImageNet（图像网络）数据集作为测试数据集。近年来，各种新的 CNN 网络结构被不断提出，从 2012 年的 AlexNet（Alex 网络）模型，到 2014 年的 VGG（VisualGeometryGroup，视觉几何组）和随后的 GoogLeNet（谷歌的 LeCun 网络）和 ResNet（Residual Network，残差网络）模型，CNN 在最初的 LeNet（LeCun 网络）基础上取得了长足进展，ImageNet 数据集的最佳定位和分类成绩也在不断刷新。ImageNet 大规模视觉识别挑战赛（ILSVRC）的历年成绩中，top-5 的错误率从 2012 年 AlexNet 的 16.4% 逐年下降到 2015 年 ResNet 的 3.57%，在 2016 年更进一步下降到 3% 以下。与模型精度的提高相对应的，是网络结构的深度也在不断增加，ILSVRC 2012 时，AlexNet 只有 8 层，到 ILSVRC 2015，ResNet 网络深度已经达到 152 层，当前 ResNet 的层数更是已经可以达到 10^3 的数量级。虽然比起 AlexNet 和 VGG，GoogLeNet 和 ResNet 的全连接层数和全连接层的节点数量都有所减少，但由于层数的大量增加，总的模型运算量还是远增。

在目标定位和分类联合应用的场景（例如，手势识别这样的典型场景），上述介绍的各种网络结构目前基本都只能在图像目标区域占图像比例大于 50% 的情况下才能取得较好的分类效果。以手势识别为例，对于手势区域占比较小的情况，就很难聚焦到手势区域学习到手势相关特征，无法识别出相应的手势。

为了解决这一问题，一种处理方式是使用级联的 CNN 网络，在两级 CNN 网络中依次进行物体定位和分类识别。为了取得更好的识别效果，随着神经网络的深度增加，运算量也逐步加大，两级神经网络的运算将占用大量 CPU 资源，对于实时视频播放过程中的同步目标分类和识别这样的应用场景，将会给 CPU 带来巨大负担，对 CPU 的运算能力提出了更高的要求。此外，在移动终端的应用中，CPU 高负荷运行产生的巨大热量也几乎是现有移动终端设备的硬件无法解决和承受的问题。

发明内容

本申请旨在至少在一定程度上解决上述相关技术中的技术问题之一。提供一种实时视频的目标检测与识别方法、系统、存储介质及设备。

为了达到上述目的，根据本申请第一方面的实施例提出了一种实时视频的目标检测与识别方法，该方法由计算设备执行，其包括：

确定当前帧图像的前一帧图像的目标物体识别结果 R_{X-1} 和两帧之前的图像的目标物体识别结果 R_{X-2} 是否相同；

当两个识别结果 R_{X-1} 和 R_{X-2} 不相同，通过第一级神经网络对当前帧图像进行目标物体位置检测，获取当前帧图像中目标物体的位置范围 C_X ；当两个识别结果 R_{X-1} 和 R_{X-2} 相同时，不经过第一级神经网络，根据前一帧图像中目标物体的位置范围 C_{X-1} 确定当前帧图像中目标物体的位置范围 C_X ；

通过第二级神经网络，根据当前帧图像中目标物体的位置范围 C_X ，对当前帧图像进行目标物体识别，得到当前帧图像的目标物体识别结果 R_X 。

在一些实施例中，通过第一级神经网络对当前帧图像进行目标物体位置检测，包括：

通过第一级神经网络判断当前帧图像中是否包含目标物体；

若当前帧图像包含目标物体，则通过第一级神经网络获取当前帧图像中目标物体的位置范围 C_X ；

若当前帧图像不含目标物体，则将判断帧图像中是否包含目标物体的间隔设置为每隔预定帧数执行一次，并通过第一级神经网络持续进行判断直至帧图像中包含目标物体。

在一些实施例中，通过第二级神经网络，根据当前帧图像中目标物体的位置范围 C_X ，对当前帧图像进行目标物体识别，得到当前帧图像的目标物体识别

结果 R_x ，包括：

将当前帧图像中目标物体的位置范围 C_x 扩展，得到面积增大预定义倍数的扩展位置范围；

通过第二级神经网络在所述扩展位置范围内进行目标物体识别，得到当前帧图像的目标物体识别结果 R_x 。

在一些实施例中，根据前一帧图像中目标物体的位置范围 C_{x-1} 确定当前帧图像中目标物体的位置范围 C_x ，包括：

将前一帧图像中目标物体的位置范围 C_{x-1} 作为当前帧图像中目标物体的位置范围 C_x ；或者

根据前一帧图像中目标物体的位置范围，以及之前预定义帧数的图像中目标物体的位置范围，预测当前帧图像中目标物体的位置范围 C_x 。

在一些实施例中，所述目标物体为人手；

所述通过第一级神经网络对当前帧图像进行目标物体位置检测，获取当前帧图像中目标物体的位置范围 C_x 包括：通过第一级神经网络识别人手在当前帧图像中的位置范围；

所述通过第二级神经网络，根据当前帧图像中目标物体的位置范围 C_x ，对当前帧图像进行目标物体识别，得到当前帧图像的目标物体识别结果 R_x ，包括：通过第二级神经网络，根据当前帧图像中目标物体的位置范围 C_x ，对当前帧图像中的人手进行手势识别，识别结果为手势的姿态。

使用本申请的实时视频的目标检测与识别方法，能够利用前帧图像的目标物体位置检测结果，降低用于位置检测的第一级神经网络的运行频率，加快整个目标物体检测和识别的速度并同时降低 CPU 和内存资源的占用，从整体上改善识别算法的性能。在两级神经网络分别采用相同结构的前提下，使用本申请的方法与每帧图像的识别均运行第一级神经网络的方案相比，可节约 CPU 资源 50% 以上。

根据本申请第二方面的实施例提供了一种实时视频的目标检测与识别系统，其包括：

第一级神经网络，用于对视频帧图像进行目标物体的位置检测，获取目标物体位置范围；

第二级神经网络，用于根据目标物体位置范围，在视频帧图像中进行目标物体的识别，得到识别结果；以及

神经网络运行控制模块，用于，

确定当前帧图像的前一帧图像的目标物体识别结果 R_{X-1} 和两帧之前的图像的目标物体识别结果 R_{X-2} 是否相同；

当两个识别结果 R_{X-1} 和 R_{X-2} 不相同，控制第一级神经网络对当前帧图像进行目标物体位置检测，获取当前帧图像中目标物体的位置范围 C_X ；

当两个识别结果 R_{X-1} 和 R_{X-2} 相同时，不经过第一级神经网络，根据前一帧图像中目标物体的位置范围 C_{X-1} 确定当前帧图像中目标物体的位置范围 C_X ；

控制第二级神经网络，根据当前帧图像中目标物体的位置范围 C_X ，对当前帧图像进行目标物体识别，得到当前帧图像的目标物体识别结果 R_X 。

在一些实施例中，在控制第一级神经网络对当前帧图像进行目标物体位置检测时，神经网络运行控制模块，用于，

控制所述第一级神经网络判断当前帧图像中是否包含目标物体；

若当前帧图像包含目标物体，则控制第一神经网络获取当前帧图像中目标物体的位置范围 C_X ；

若当前帧图像不含目标物体，则将判断帧图像中是否包含目标物体的间隔设置为每隔预定帧数执行一次，并控制第一级神经网络持续进行判断直至帧图像中包含目标物体。

在一些实施例中，在控制第二级神经网络根据当前帧图像中目标物体的位置范围 C_X ，对当前帧图像进行目标物体识别，当前帧图像的目标物体得到识别结果 R_X 时，神经网络运行控制模块，用于，

将当前帧图像中目标物体的位置范围 C_X 扩展，得到面积增大预定义倍数的扩展位置范围；

控制第二级神经网络在所述扩展位置范围内进行目标物体识别，得到当前帧图像的目标物体识别结果 R_X 。

在一些实施例中，在根据前一帧图像中目标物体的位置范围 C_{X-1} 确定当前帧图像中目标物体的位置范围 C_X 时，神经网络运行控制模块，用于，

将前一帧图像中目标物体的位置范围 C_{X-1} 作为当前帧图像中目标物体的位置范围 C_X ；或者

根据前一帧图像中目标物体的位置范围，以及之前预定义帧数的图像中目标物体的位置范围，预测当前帧图像中目标物体的位置范围 C_X 。

在一些实施例中，所述目标物体为人手；在控制第一级神经网络对当前帧

图像进行目标物体位置检测，获取当前帧图像中目标物体的位置范围 C_x 时，神经网络运行控制模块，用于控制所述第一级神经网络识别人手在当前帧图像中的位置范围；

在控制第二级神经网络，根据当前帧图像中目标物体的位置范围 C_x ，对当前帧图像进行目标物体识别，得到当前帧图像的目标物体识别结果 R_x 时，所述神经网络运行控制模块，用于控制所述第二级神经网络，根据所述当前帧图像中目标物体的位置范围 C_x ，对当前帧图像中的人手进行手势识别，识别结果为手势的姿态。

使用本申请的实时视频的目标检测与识别系统，能够利用前帧图像的目标物体位置检测结果，降低用于位置检测的第一级神经网络的运行频率，加快整个目标物体检测和识别的速度并同时降低 CPU 和内存资源的占用，从整体上改善识别算法的性能。在两级神经网络分别采用相同结构的前提下，使用本申请的方法与每帧图像的识别均运行第一级神经网络的方案相比，可节约 CPU 资源 50% 以上。

为了实现上述目的，本申请第三方面实施例提供了一种非临时性计算机可读存储介质，其上存储有计算机程序，该程序被处理器执行时实现根据本申请第一方面所述的实时视频的目标检测与识别方法。

为了实现上述目的，本申请第四方面实施例提供了一种计算机程序产品，当所述计算机程序产品中的指令由处理器执行时，实现根据本申请第一方面所述的实时视频的目标检测与识别方法。

为了实现上述目的，本申请第五方面实施例提供了一种计算设备，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序，所述处理器执行所述程序时，实现根据本申请第一方面所述的实时视频的目标检测与识别方法。

根据本申请第三到五方面的非临时性计算机可读存储介质，计算机程序产品和计算设备具有与根据本申请第一和第二方面的实时视频的目标检测与识别方法和系统类似的有益效果，在此不再赘述。

附图说明

本申请上述的和/或附加的方面和优点从下面结合附图对实施例的描述中将变得明显和容易理解，其中：

图 1 是根据本申请实施例的目标检测与识别的级联神经网络工作原理示意图；

图 2 是根据本申请实施例的实时视频的目标检测与识别方法的流程示意图；

图 3 是根据本申请实施例的第一神经网络进行目标物体位置检测的流程示意图；

图 4 是根据本申请的对目标物体位置范围进行扩展的示意图；

图 5 是根据本申请实施例的手势检测程序实现流程示意图；

图 6 是根据本申请实施例的实时视频的目标检测与识别系统的结构框图；

图 7 是根据本申请实施例的实时视频的目标检测与识别系统的神经网络运行控制模块结构框图；

图 8 是根据本申请实施例的计算设备的结构示意图。

具体实施方式

下面详细描述本申请的实施例，所述实施例的示例在附图中示出，其中自始至终相同或类似的标号表示相同或类似的元件或具有相同或类似功能的元件。下面通过参考附图描述的实施例是示例性的，旨在用于解释本申请，而不能理解为对本申请的限制。

在进行实时视频的目标检测与识别之时，需要将视频拆分成一帧帧的图像，在单帧图像的基础上进行目标检测与识别。从肉眼的极限分辨频率角度来说，视频的流速达到 30 帧/秒以上时人眼就已经无法进一步分辨，因此，目前视频的帧率一般都会设置在 30 帧/秒或以下。

以 30 帧/秒为例，经测试，按照每帧图像通过两级 CNN 神经网络进行目标检测和识别的方法，第一级神经网络使用 U-net（U 形网络，是全卷积神经网络 FCN 的一种，适于图像语义分割），第二级神经网络使用 VGG 结构为例，硬件设备选用 iPhoneX（属于现有移动终端中性能较佳的那一个级别），运行实时视频目标检测和识别时，CPU 的占用率可达 45% 左右。如果采用更复杂的网络结构，CPU 的占用率还会进一步增加。

对于边播放视频，边通过摄像头拍摄手势识别用户评价这样的应用场景，或者是视频直播，同时对手势、表情或者其他目标进行识别的场景，由于视频信息处理本身对 CPU 的使用就非常高，再加上目标物体识别所占用的，CPU 会一直处于一种负载非常高的状态。对于移动终端来说，CPU 持续高负荷运行会

使机身发热发烫，无论对于用户还是机器的硬件本身大部分时候都是无法接受的。

并且，为了使目标识别结果和视频帧同步，进行识别时也可能需要阻塞主线程。对于计算能力比较低的中低端设备，由于CPU的算力较差，处理目标识别时间较长，会导致视频渲染帧率下降，视频卡顿等现象，使得目标识别的神经网络在低端机型上无法使用。

为了解决相关技术中的目标识别方案CPU占用高，识别时间长的问题。本文提出一种高性能的目标检测与识别的方法。利用相邻视频帧中，目标物体的大范围位移概率较小的特点，在视频帧之间目标物体不发生位移或者位移不影响识别时，利用历史上的目标物体位置范围识别结果来确定当前帧的目标物体位置范围，减少用于位置检测的前级神经网络的运行，从而降低计算量。

下面参考附图对本申请实施例的方法和系统进行详细的说明。

为了更好地理解本申请，首先介绍用级联的两级神经网络进行目标检测与识别的工作原理。图1是根据本申请实施例的目标检测与识别的级联神经网络工作原理示意图。其中，用于目标检测与识别的神经网络可包括第一级神经网络110（也称为前级网络）和第二级神经网络120（也称为后级网络）。

第一级神经网络110，用于对视频帧图像进行目标物体的位置检测，获取目标物体位置范围，即目标物体在图中的区域。位置检测可以通过例如对图像进行语义分割等方式实现。其输出可以是包含目标物体的轮廓信息的位置检测识别结果位置范围，经过裁切和调整大小后得到包含目标物体的图像区域。例如，可以用来识别图像中的人手，并确定人手的位置范围（在图像中的所在区域）。或者是用来识别人脸，得到人脸的位置范围（在图像中的所在区域）。

在本申请实施例中，仅以目标物体和人手或人脸为例进行说明，除此之外，目标物体也可以是其它任意类型的物体，比如，目标物体可以是人脚等其它躯干，或者，目标物体是整个的人体，或者，目标物体也可以是特定的动物、植物、房屋等其它类型物体。对于目标物体的类型，本申请不做限定。

第二级神经网络120，用于根据目标物体位置范围（在图像中的所在区域），对视频帧图像进行目标物体的细节或者特征的识别，得到识别结果。例如，进行手势分类，得到手势类型。或者是用来识别人脸，得到人脸的表情等。

参见图1，其中，待识别的图像210为一幅包含人手的图像，图像210可以是一段人手动作的视频中的一帧。目标物体为人手，目的是识别人手的手势。

首先经过第一级神经网络 110，通过语义分割进行目标位置检测，得到图像中目标物体——人手的轮廓（对应图 1 中 220）。根据人手的轮廓可确定目标物体的位置范围。根据位置范围，获取图像 210 中对应目标物体位置范围的子图像，即可用于下一步的目标物体的进一步识别。例如，可对图像 210 进行裁切和调整大小（对应图 1 中 230），得到含有人手，且人手占据绝大部分画面面积的适于手势识别的子图像 240。之后，将子图像 240 输入到第二神经网络 120，进行手势识别，得到处理后的输出图像 250 和输出的对应手势类别的特征向量（以柱状图 260 表示），例如，在本实施例中，该柱状图 260 表示的是手势的含义是“好”或者“称赞”。

通过对第一级神经网络和第二级神经网络运行的分析，就目前常用的几种 CNN 结构来说，目标物体的位置检测任务由于可能涉及区分目标物体和背景，以及目标物体的边缘检测等，可能需要较多的卷积层，所需的计算量也相对较大。相对于目标检测，分类任务的特征更加容易提取，因此第二神经网络所需的网络结构相对没有那么复杂。以手势识别为例，实验也证明，在整个手势识别过程中，第一级神经网络的耗时和 CPU 占用约是第二级神经网络的五倍。也就是说，目标检测和识别过程中的主要消耗是在第一级神经网络上。

在实际场景中，例如利用移动终端进行直播时，用户在对着手机比手势或者做表情的时候，手势和表情可能会不断变换，但是手的位置或脸的位置大部分时间是变化很小的，基本不动或者缓慢移动，对于前级网络，可以参用上一次前级的结果对当前帧的手或脸的范围进行估计，并用后级网络计算出的结果对估计的范围进行有效性验证。

图 2 是根据本申请实施例的实时视频的目标检测与识别方法的流程示意图。其中，对各个视频帧中目标物体的检测与识别，可以包括步骤 S100 和 S200，其中步骤 S100 进一步包括子步骤 S110 到 S130。

第一级神经网络和第二级神经网络可以均使用卷积神经网络。并且预先进行训练，得到相应的模型参数后，在进行目标检测与识别的过程中直接根据训练好的模型参数进行计算。

其中，第一级神经网络可以使用各种适用于进行目标物体位置检测的网络模型，例如，全卷积神经网络（FCN）模型，本申请对此没有限制。在一些实施例中，第一级神经网络可以使用适于实现图像语义分割的 U-net 模型。

关于使用 FCN 以及 U-net 实现视频帧图像语义分割的具体方式，与相关技

术中的图像语义分割实现类似，在此不再赘述。

第二级神经网络可以使用各种适用于进行目标物体识别的网络模型，例如，AlexNet、VGG、GoogLeNet 和 ResNet 等，本申请对此没有限制。可以根据识别的精度需求和硬件计算能力选择相应的网络模型。

例如，对于手势识别，综合考虑模型复杂度和识别精度的需求，第二级神经网络可以使用 VGG 模型。相关技术中使用 VGG 进行目标物体特征识别的各种方法均可以用于本申请的第二神经网络，对此没有特殊的要求，也不再赘述细节。对于表情识别，随着深度学习技术的飞速发展，使用深度神经网络进行基于非显示特征的表情识别的精确度已经越来越高，直接使用级联的两级神经网络进行人脸的表情识别已展现出非常好的应用前景。AlexNet、VGG、GoogLeNet 和 ResNet 等网络结构也均可用于表情识别和手势识别。

第一级神经网络参数可按照目标位置检测目的进行模型训练得到，而第二级神经网络的参数可按照目标识别的目的训练，例如按照自适应分类，或者根据给定类别分类的标注数据进行模型训练得到。

在进行实时视频的目标检测与识别的过程中，对于拆分的每幅视频帧图像，首先，在步骤 S100，获取当前帧图像中目标物体的位置范围 C_X 。

所述目标物体的位置范围 C_X 可以是根据目标物体的轮廓确定的一个规则形状的范围，例如，图 1 中图像 240 所示的矩形范围，以便于后续的图像中目标物体的进一步识别。位置范围 C_X 可以用能够表征其的坐标范围来表示。例如，矩形的 4 个顶点，圆形的圆心和半径等，在图像处理中，以矩形范围的使用居多。根据目标物体的位置范围 C_X 即可以从当前帧图像中获得相应的适用于第二神经网络进一步识别的子图像。

由于相邻视频帧中，目标物体的位置较接近或者不变的概率很大，因此，可根据当前帧之前的历史帧图像的识别结果来确定是否需要运行第一神经网络进行位置检测，从而节约算力。步骤 S100 可以具体按照下述步骤 S110 到 S130 实现。

在步骤 S110，确定当前帧的前一帧图像的目标物体识别结果 R_{X-1} 和两帧之前的图像的目标物体识别结果 R_{X-2} 是否相同。

这一步骤主要为了估计将上一帧图像的位置检测结果用于本帧的可行性。前一帧和之前的一帧的识别结果 R_{X-1} 和 R_{X-2} 相同时，通常意味着目标物体处于较稳定的状态，可能没有发生移动和姿势变化。而 R_{X-1} 和 R_{X-2} 不相同，则目

标物体位置发生变化的概率较大。

以手势识别为例，当手势发生变化时，识别结果自然是不同的。除此之外，当目标人手位置发生了变化，而使得第二级网络的输入图像中不含人手时，可能由于背景物体的存在，仍旧可能得出识别结果，也就是说，位置范围 C_{X-1} 内不含“手”的情况，也可能反映在手势识别结果 R_{X-1} 的变化上。

因此，当两个识别结果 R_{X-1} 和 R_{X-2} 不相同，执行步骤 S130，通过第一级神经网络对当前帧图像进行目标物体位置检测，获取当前帧图像中目标物体的位置范围 C_X 。

参见图 3，图 3 是根据本申请实施例的第一种网络进行目标物体位置检测的流程示意图。通过第一级神经网络对当前帧图像进行目标物体位置检测，可包括：步骤 S131，首先通过第一级神经网络判断当前帧图像中是否包含目标物体；若当前帧图像包含目标物体，则进一步执行步骤 S133，通过第一级神经网络获取当前帧图像中目标物体的位置范围 C_X ；若当前帧图像不含目标物体，则显然无需进一步的识别。此时，执行步骤 S132，对于新的视频帧，可对于每帧或预定帧数重复执行判断是否有手的步骤。

考虑到对于手势识别或者表情识别等应用场景，人体的动作速度一般不会过高，按照 30 帧/秒的帧率，如果当前帧中不含目标物体（例如，人手或人脸），可能随后的若干帧中，仍然是不含目标物体的。因此，可将判断帧图像中是否包含目标物体的间隔设置为每隔预定帧数执行一次，并通过第一级神经网络持续进行判断直至帧图像中包含目标物体。例如，对于手势识别，可每隔 10-20 帧判断一次帧图像中是否有手。

至于的具体的位置检测过程，则可以使用任意一种相关技术中的位置检测算法，本申请对此没有限定。

当两个识别结果 R_{X-1} 和 R_{X-2} 相同时，执行步骤 S120，不经过第一级神经网络，根据前一帧图像中的目标物体位置范围 C_{X-1} 确定当前帧图像中的目标物体位置范围 C_X 。具体而言，可以直接将前一帧图像中目标物体的位置范围 C_{X-1} 作为当前帧图像中目标物体的位置范围 C_X 。此种方式比较适合目标物体静止或者运动幅度微小的情况。

考虑到有时目标物体可能是移动的，例如手势识别时，用户可能长时间手部保持同一位置会疲惫，可以通过改变手部位置缓解疲劳，此时手在帧图像中的位置则可能是存在较缓慢的变动。或者是，表情识别时，头部可能会转动移

动等，导致人脸位置的改变。因此，在一些实施例中，可根据上一帧图像中的目标物体位置范围，以及之前预定义帧数的图像中的目标物体位置范围，预测当前帧图像中目标物体的位置范围 C_X 。基于手部动作的连续性，简单的线性或者低级多项式形式的移动路径拟合算法，即可得到较好的预测效果，从而获取较准确的目标物体位置范围 C_X 。

在步骤 S140，通过第二级神经网络，根据当前帧图像中目标物体的位置范围 C_X ，对当前帧图像进行目标物体识别，得到当前帧图像的目标物体识别结果 R_X 。

考虑到目标物体在两帧之间可能发生的小范围移动，为了使利用前帧的位置检测检测结果来预测本帧中目标物体的位置范围的准确度更高，可以将每帧图像中目标物体位置范围进行扩展，以一个更大范围的图像来作为第二级神经网络的输入。

即，在通过第二级神经网络，根据当前帧图像中目标物体的位置范围 C_X ，对当前帧图像进行目标物体识别之前，将当前帧图像中目标物体的位置范围 C_X 扩展，得到面积增大预定义倍数的扩展位置范围；并通过第二级神经网络，在所述扩展位置范围内进行目标物体识别，得到当前帧图像的目标物体识别结果 R_X 。

参见图 4，其中内圈的方框 410 中示出的是当前帧图像中目标物体的位置范围 C_X ，外圈的方框内的范围表示扩展后用于第二级神经网络的新的位置范围。

在一些实施例中，本申请的方法被用于人的手势识别。其中，上述目标物体为人手；所述通过第一级神经网络对当前帧图像进行目标物体位置检测，获取当前帧图像中目标物体的位置范围 C_X 包括：通过第一级神经网络识别人手在当前帧图像中的位置范围；所述通过第二级神经网络，根据当前帧图像中目标物体的位置范围 C_X ，对当前帧图像进行目标物体识别，得到当前帧图像的目标物体识别结果 R_X ，包括：通过第二级神经网络，根据当前帧图像中目标物体的位置范围 C_X ，对当前帧图像中的人手进行手势识别，识别结果为手势的姿态。

下面结合图 5，对使用本申请的方法进行手势识别的实施例进行详细说明。图 5 是根据本申请实施例的手势检测程序实现流程示意图。其中，设置变量 $doFcn$ 用于判断是否需要运行第一级神经网络的标识， $doFcn$ 变量可设为二值变量，例如取值 “Yes”（是）和 “No”（否）。分别对应执行和不执行第一级神经网络的目标检测。

在步骤 S400, 开始执行。

步骤 S410, 在每帧检测开始前, 先对 doFcn 变量的值进行判断。如果 doFcn 值为 Yes 则执行步骤 S420, 如果 doFcn 值为 No 则执行步骤 S430。

一般来说, doFcn 的值可根据上一帧后级网络和上上一帧的第二级神经网络的识别结果而定。假设上一帧后级网络的识别结果为 R_{X-1} , 上上一帧的后级网络的识别 R_{X-2} , 如果 R_{X-1} 不等于 R_{X-2} , 则 doFcn = Yes ; 相等则 doFcn = No, 第一帧时 doFcn = Yes。

步骤 S420, doFcn 为 Yes, 先对此帧执行前级网络, 即使用第一神经网络计算, 进行帧图像中目标物体的位置检测。

首先, 在步骤 S422, 判断当前帧图像中是否有手。如果有手, 则执行步骤 S430, 无手则执行步骤 S424。

步骤 S424, 由于第一级神经网络判断当前帧中没手, 则将 doFcn 值为 Yes, 并将判定间隔设置为每 15 帧检测一次, 即直接跳过当前帧的后级检测, 并此跳过此帧过后的十四帧的前级检测和后级前级, 并直接反馈给用户图视频中没有手。第十五帧开始重新恢复检测。当然, 本实施例以每隔 15 帧为例, 其它的帧间隔也是可以的。

在步骤 S430, 第一级神经网络判断当前帧中有手, 则由第二级神经网络计算, 在检测到的目标位置范围内, 进行目标识别。利用上一帧中目标物体的位置范围, 对当前帧图像进行裁剪, 得到目标物体位置范围对应的子图像, 第二级神经网络根据子图像的区域进行计算, 并得出手势类型识别结果为 R_X 。

之后, 由于“手”已经进入了视频帧的范围内, 不必再以大的帧间隔进行检测。因此, 在接下来的步骤 S432, 将 doFcn 置为 No, 并将判定间隔设置为每帧都检测, 也就是将步骤 S422 的执行设置为每帧进行一次判断。

如果是在步骤 S410 中, doFcn 值为 No 而直接执行的步骤 S430, 使用第二级神经网络计算, 在检测到的目标位置范围内, 进行目标识别。则执行之后, 可以进一步执行 S440, 判断 doFcn 值是否需要修改。

在步骤 S440, 将当前帧的识别结果与上一帧的识别结果进行比较, 从而用于预测下一帧是否可以使用当前帧的目标位置范围和设置相应的 doFcn 值。在当前帧的识别结果与上一帧的识别结果相同时, 执行步骤 S444, 不同时执行步骤 S442。

在步骤 S442, 将 doFcn 值设置为 Yes, 并将步骤 S422 的执行设置为每帧进

行一次判断。

在步骤 S44, 将 doFcn 值设置为 No, 并将步骤 S422 的执行设置为每帧进行一次判断。

在步骤 S450, 完成对当前帧的处理, 判断是否有新的视频帧到达, 如果有, 则返回步骤 S410, 重新开始判断 doFcn 值。如果没有新的视频帧, 则执行步骤 S460, 结束识别过程。

通过上述步骤, 可以对手势进行快速准确的识别。进行人脸的表情识别的过程与此类似, 只是在将人手和手势替换成人脸和表情, 神经网络模型的结构和参数进行相应调整, 但是整体的思路是一致的, 因此不再赘述。

使用本申请的实时视频的目标检测与识别方法, 能够利用前帧图像的目标物体位置检测结果, 降低用于位置检测的第一级神经网络的运行频率, 加快整个目标物体检测和识别的速度并同时降低 CPU 和内存资源的占用, 从整体上改善识别算法的性能。在两级神经网络分别采用相同结构的前提下, 使用本申请的方法与每帧图像的识别均运行第一级神经网络的方案相比, 可节约 CPU 资源 50% 以上。

以手势识别为例, 按照每秒 30 帧的帧率, 设对于每帧图像, 第一级神经网络的执行耗时为 T_1 , 第二级神经网络的执行耗时为 T_2 , $T_1=5T_2$ 。则, 在出现视频帧内无手的情况时, 使用本申请方法与不使用的情况的耗时比为: $2*T_1/30*(T_1+T_2)=2*10*T_2/30*(5T_2+T_2)=11.11\%$, 相应地, 可视为其计算量和 CPU 占用上约为使用本方法之前的百分之 11.11。

对于视频帧内有手的情况, 经统计, 在手势变换和正常速度移动情况下, 手势位置的预测在每 3 帧中大约可以命中 2 帧, 使用本申请方法与不使用的情况的耗时比为: $(10*T_1+30*T_2)/30*(T_1+T_2)=60*T_2/30*(5T_2+T_2)=44.44\%$, 相应地, 可视为其计算量和 CPU 占用上约为使用本方法之前的百分之 44.44。

为了实现上述第一方面实施例中的方法, 本申请第二方面的实施例提出了一种实时视频的目标检测与识别系统。

所述实时视频的目标检测与识别系统的实现可包括一个或多个计算设备, 所述计算设备包括处理器和存储器, 所述存储器上存储有包括可在所述处理器上运行的计算机程序指令的应用程序。所述应用程序可以划分为多个程序模块, 用于系统各个组成部分的相应功能。其中, 程序的模块的划分是逻辑上的而非物理上的, 每个程序模块可以运行在一个或多个计算设备上, 一个计算设备上

也可以运行一个或一个以上的程序模块。以下对本申请的系统按照程序模块的功能逻辑划分进行详细说明。

图 6 是根据本申请实施例的实时视频的目标检测与识别系统结构框图。其中, 根据本申请的系统 100 可以包括第一级神经网络 110, 第二级神经网络 120 和神经网络运行控制模块 130。

所述实时视频的目标检测与识别系统 100 可以采用包括处理器和存储器的计算设备实现, 所述存储器中存储有可被所述处理器执行的程序模块, 各个程序模块被执行时, 控制所述计算设备实现相应的功能。例如, 第一级神经网络 110, 第二级神经网络 120 和神经网络运行控制模块 130 可以实现为这样的程序模块。

第一级神经网络 110, 用于对视频帧图像进行目标物体的位置检测, 获取目标物体位置范围。

第二级神经网络 120, 用于根据目标物体位置范围, 在视频帧图像中进行目标物体的识别, 得到识别结果。

神经网络运行控制模块 130, 用于根据过往视频帧图像的目标物体识别结果控制第一神经网络和第二神经网络工作。其中, 神经网络运行控制模块 130, 用于确定当前帧图像的前一帧图像的目标物体识别结果 R_{X-1} 和两帧之前的图像的目标物体识别结果 R_{X-2} 是否相同; 当两个识别结果 R_{X-1} 和 R_{X-2} 不相同, 控制第一级神经网络对当前帧图像进行目标物体位置检测, 获取当前帧图像中目标物体的位置范围 C_X ; 当两个识别结果 R_{X-1} 和 R_{X-2} 相同时, 不经过第一级神经网络 110, 根据前一帧图像中目标物体的位置范围 C_{X-1} 确定当前帧图像中目标物体的位置范围 C_X ; 以及控制第二级神经网络 120, 根据当前帧图像中目标物体的位置范围 C_X , 对当前帧图像进行目标物体识别, 得到当前帧图像的目标物体识别结果 R_X 。

参见图 7, 图 7 是根据本申请实施例的实时视频的目标检测与识别系统的神经网络运行控制模块结构框图。神经网络运行控制模块 130 可进一步包括: 识别结果比较单元 131、第一级神经网络运行控制单元 132、目标位置物体范围确定单元 133 和第二级神经网络运行控制单元 134。

识别结果比较单元 131, 用于确定当前帧图像的前一帧图像的目标物体识别结果 R_{X-1} 和两帧之前的图像的目标物体识别结果 R_{X-2} 是否相同。

第一级神经网络运行控制单元 132, 用于当两个识别结果 R_{X-1} 和 R_{X-2} 不相同

时，控制第一级神经网络对当前帧图像进行目标物体位置检测。

在一些实施例中，控制第一级神经网络 110 对当前帧的图像进行目标物体位置检测，可包括：控制第一级神经网络 110 判断当前帧图像中是否包含目标物体；若当前帧图像包含目标物体，则控制第一神经网络进一步获取当前帧图像中目标物体的位置范围 C_X ；若当前帧图像不含目标物体，则将判断帧图像中是否包含目标物体的间隔设置为每隔预定帧数执行一次，并控制第一级神经网络持续进行判断直至帧图像中包含目标物体。

目标位置物体范围确定单元 133，用于根据当前帧第一级神经网络的目标物体位置检测结果或者历史确定的目标物体范围，获取当前帧图像中的目标物体位置范围 C_X 。

在一些实施例中，在根据前一帧图像中目标物体的位置范围 C_{X-1} 确定当前帧图像中目标物体的位置范围 C_X 时，目标位置物体范围确定单元 133，用于将前一帧图像中目标物体的位置范围 C_{X-1} 作为当前帧图像中目标物体的位置范围 C_X 。或者，在另一些实施例中，可根据前一帧图像中目标物体的位置范围，以及之前预定义帧数的图像中目标物体的位置范围，预测当前帧图像中目标物体的位置范围 C_X 。

第二级神经网络运行控制单元 134，用于控制第二级神经网络 120，根据当前帧图像中目标物体的位置范围 C_X ，对当前帧图像进行目标物体识别，得到当前帧图像的目标物体识别结果 R_X 。

在一些实施例中，第二级神经网络运行控制单元 134，用于对当前帧图像进行目标物体识别之前，将当前帧图像中目标物体的位置范围 C_X 扩展，得到面积增大预定义倍数的扩展位置范围；并控制第二级神经网络在所述扩展位置范围内进行目标物体识别，得到当前帧图像的目标物体识别结果 R_X 。

特别是，在一些实施例中，所述目标物体可为人手；在控制第一级神经网络 110 对当前帧图像进行目标物体位置检测，获取当前帧图像中的目标物体位置范围 C_X 时，第一级神经网络运行控制单元 132，用于控制第一级神经网络识别人手在当前帧图像中的位置范围；在控制第二级神经网络 120，根据当前帧图像中目标物体的位置范围 C_X ，对当前帧图像进行目标物体识别，得到当前帧图像的目标物体识别结果 R_X 时，第二级神经网络运行控制单元 134，用于控制第二级神经网络，根据当前帧图像中目标物体的位置范围 C_X ，对当前帧图像中的人手进行手势识别，识别结果为手势的姿态。

本申请实时视频的目标检测与识别系统中的各个模块的功能和作用的实现过程具体详情可参见上述方法中对应步骤的实现过程。对于系统实施例而言，由于其基本对应于方法实施例，前述对本申请的方法实施例的解释说明也适用于本申请的系统实施例。为避免冗余，在系统实施例中将不会对所有细节进行重复，相关未尽之处可参见上述结合图 1 到图 5 对本申请实时视频的目标检测与识别方法实施例的相关描述。

使用本申请的实时视频的目标检测与识别方法，能够利用前帧图像的目标物体位置检测结果，降低用于位置检测的第一级神经网络的运行频率，加快整个目标物体检测和识别的速度并同时降低 CPU 和内存资源的占用，从整体上改善识别算法的性能。在两级神经网络分别采用相同结构的前提下，使用本申请的方法与每帧图像的识别均运行第一级神经网络的方案相比可节约 CPU 资源 50% 以上。

本申请上述实施例涉及的方法或系统，可以应用于 AI(Artificial Intelligence, 人工智能) 中，以实现基于目标检测与识别的 AI 应用场景，比如，实现通过手势或者表情与 AI 机器人进行交互的场景等。

本申请第三方面的实施例提出了一种非临时性计算机可读存储介质，其上存储有可执行指令，所述可执行指令在处理器上运行时，实现如本申请第一方面实施例所述的实时视频的目标检测与识别方法。该存储介质可以作为设备的一部分设置在其上；或者当该设备可以被服务器远程控制时，该存储介质可以设置在对设备进行控制的远程服务器上。

用于实现本申请方法的计算机指令的可以采用一个或多个计算机可读的介质的任意组合来承载。所谓非临时性计算机可读介质可以包括任何计算机可读介质，除了临时性地传播中的信号本身。计算机可读存储介质例如可以是——但不限于——电、磁、光、电磁、红外线、或半导体的系统、装置或器件，或者任意以上的组合。计算机可读存储介质的更具体的例子（非穷举的列表）包括：具有一个或多个导线的电连接、便携式计算机磁盘、硬盘、随机存取存储器（RAM）、只读存储器（ROM）、可擦式可编程只读存储器（EPROM 或闪存）、光纤、便携式紧凑磁盘只读存储器（CD-ROM）、光存储器件、磁存储器件、或者上述的任意合适的组合。在本文件中，计算机可读存储介质可以是任何包含或存储程序的有形介质，该程序可以被指令执行系统、装置或者器件使用或者与其结合使用。

本申请第四方面的实施例提供了一种计算机程序产品，当所述计算机程序产品中的指令由处理器执行时，实现根据本申请第一方面实施例所述的实时视频的目标检测与识别方法。

可以以一种或多种程序设计语言或其组合来编写用于执行本申请操作的计算机程序代码，所述程序设计语言包括面向对象的程序设计语言—诸如 Java、Smalltalk、C++，还包括常规的过程式程序设计语言—诸如“C”语言或类似的程序设计语言。程序代码可以完全地在用户计算机上执行、部分地在用户计算机上执行、作为一个独立的软件包执行、部分在用户计算机上部分在远程计算机上执行、或者完全在远程计算机或服务器上执行。在涉及远程计算机的情形中，远程计算机可以通过任意种类的网络——包括局域网(LAN)或广域网(WAN)—连接到用户计算机，或者，可以连接到外部计算机（例如利用因特网服务提供商来通过因特网连接）。

本申请第五方面实施例提供了一种计算设备，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序，所述处理器执行所述程序时，实现根据本申请第一方面所述的实时视频的目标检测与识别方法。

上述使用根据本申请的存储介质、计算机程序产品和计算设备，其相关部分的具体实施方式可以从相应的本申请的实时视频的目标检测与识别方法或系统的实施例中获得，并具有与相应的本申请的实时视频的目标检测与识别方法或系统相似的有益效果，在此不再赘述。

图 8 示出了适于用来实现本申请实施方式的示例性计算设备的框图。图 8 显示的计算设备 12 仅仅是一个示例，不应对本申请实施例的功能和使用范围带来任何限制。

如图 8 所示，计算设备 12 可以通用计算设备的形式实现。计算设备 12 的组件可以包括但不限于：一个或者多个处理器或者处理单元 16，系统存储器 28，连接不同系统组件（包括系统存储器 28 和处理单元 16）的总线 18。

总线 18 表示几类总线结构中的一种或多种，包括存储器总线或者存储器控制器，外围总线，图形加速端口，处理器或者使用多种总线结构中的任意总线结构的局域总线。举例来说，这些体系结构包括但不限于工业标准体系结构

（Industry Standard Architecture；以下简称：ISA）总线，微通道体系结构（Micro Channel Architecture；以下简称：MAC）总线，增强型 ISA 总线、视频电子标准协会（Video Electronics Standards Association；以下简称：VESA）局域总线以及

外围组件互连 (Peripheral Component Interconnection; 以下简称: PCI) 总线。

计算设备 12 典型地包括多种计算机系统可读介质。这些介质可以是任何能够被计算设备 12 访问的可用介质, 包括易失性和非易失性介质, 可移动的和不可移动的介质。

存储器 28 可以包括易失性存储器形式的计算机系统可读介质, 例如随机存取存储器 (Random Access Memory; 以下简称: RAM) 30 和/或高速缓存存储器 32。计算设备 12 可以进一步包括其它可移动/不可移动的、易失性/非易失性的计算机可读存储介质。仅作为举例, 存储系统 34 可以用于读写不可移动的、非易失性磁介质 (图中未显示, 通常称为“硬盘驱动器”)。尽管图 8 中未示出, 可以提供用于对可移动非易失性磁盘 (例如“软盘”) 读写的磁盘驱动器, 以及对可移动非易失性光盘 (例如: 光盘只读存储器 (Compact Disc Read Only Memory; 以下简称: CD-ROM)、数字多功能只读光盘 (Digital Video Disc Read Only Memory; 以下简称: DVD-ROM) 或者其它光介质) 读写的光盘驱动器。在这些情况下, 每个驱动器可以通过一个或者多个数据介质接口与总线 18 相连。存储器 28 可以包括至少一个程序产品, 该程序产品具有一组 (例如至少一个) 程序模块, 这些程序模块被配置以执行本申请各实施例的功能。

具有一组 (至少一个) 程序模块 42 的程序/实用工具 40, 可以存储在例如存储器 28 中, 这样的程序模块 42 包括但不限于操作系统、一个或者多个应用程序、其它程序模块以及程序数据, 这些示例中的每一个或某种组合中可能包括网络环境的实现。程序模块 42 通常执行本申请所描述的实施例中的功能和/或方法。

计算设备 12 也可以与一个或多个外部设备 14 (例如键盘、指向设备、显示器 24 等) 通信, 还可与一个或者多个使得用户能与该计算机系统/服务器 12 交互的设备通信, 和/或与使得该计算机系统/服务器 12 能与一个或多个其它计算设备进行通信的任何设备 (例如网卡, 调制解调器等等) 通信。这种通信可以通过输入/输出 (I/O) 接口 22 进行。并且, 计算设备 12 还可以通过网络适配器 20 与一个或者多个网络 (例如局域网 (Local Area Network; 以下简称: LAN), 广域网 (Wide Area Network; 以下简称: WAN) 和/或公共网络, 例如因特网) 通信。如图所示, 网络适配器 20 通过总线 18 与计算设备 12 的其它模块通信。要说明的是, 尽管图中未示出, 可以结合计算设备 12 使用其它硬件和/或软件模块, 包括但不限于: 微代码、设备驱动器、冗余处理单元、外部磁盘驱动阵列、

RAID 系统、磁带驱动器以及数据备份存储系统等。

处理单元 16 通过运行存储在系统存储器 28 中的程序，从而执行各种功能应用以及数据处理，例如实现前述实施例中提及的方法。

根据本申请第三到五方面的非临时性计算机可读存储介质，计算机程序产品和计算设备，可以参照根据本申请第一方面实施例具体描述的内容实现，并具有与根据本申请第一方面的实时视频的目标检测与识别方法类似的有益效果，在此不再赘述。

需要说明的是，在本说明书的描述中，参考术语“一个实施例”、“一些实施例”、“示例”、“具体示例”、或“一些示例”等的描述意指结合该实施例或示例描述的具体特征、结构、材料或者特点包含于本申请的至少一个实施例或示例中。在本说明书中，对上述术语的示意性表述不必须针对的是相同的实施例或示例。而且，描述的具体特征、结构、材料或者特点可以在任一个或多个实施例或示例中以合适的方式结合。此外，在不相互矛盾的情况下，本领域的技术人员可以将本说明书中描述的不同实施例或示例以及不同实施例或示例的特征进行结合和组合。

此外，术语“第一”、“第二”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”的特征可以明示或者隐含地包括至少一个该特征。在本申请的描述中，“多个”的含义是两个或两个以上，例如两个，三个等，除非另有明确具体的限定。

本技术领域的普通技术人员可以理解实现上述实施例的方法携带的全部或部分步骤是可以通过程序来指令相关的硬件完成，所述的程序可以存储于一种计算机可读存储介质中，该程序在执行时，包括方法实施例的步骤之一或其组合。

在本说明书的描述中，流程图中或在此以其他方式描述的任何过程或方法描述可以被理解为，表示包括一个或更多个用于实现特定逻辑功能或过程的步骤的可执行指令的代码的模块、片段或部分，并且本申请的优选实施方式的范围包括另外的实现，其中可以不按所示出或讨论的顺序，包括根据所涉及的功能按基本同时的方式或按相反的顺序，来执行功能，这应被本申请的实施例所属技术领域的技术人员所理解。

在流程图中表示或在此以其他方式描述的逻辑和/或步骤，例如，可以被认为用于实现逻辑功能的可执行指令的定序列表，可以具体实现在任何计算机

可读介质中，以供指令执行系统、装置或设备(如基于计算机的系统、包括处理器的系统或其他可以从指令执行系统、装置或设备取指令并执行指令的系统)使用，或结合这些指令执行系统、装置或设备而使用。就本说明书而言，"计算机可读介质"可以是任何可以包含、存储、通信、传播或传输程序以供指令执行系统、装置或设备或结合这些指令执行系统、装置或设备而使用的装置。

应当理解，本申请的各部分可以用硬件、软件、固件或它们的组合来实现。在上述实施方式中，多个步骤或方法可以用存储在存储器中且由合适的指令执行系统执行的软件或固件来实现。例如，如果用硬件来实现，和在另一个实施方式中一样，可用本领域公知的下列技术中的任一项或他们的组合来实现：具有用于对数据信号实现逻辑功能的逻辑门电路的离散逻辑电路，具有合适的组合逻辑门电路的专用集成电路，可编程门阵列(PGA)，现场可编程门阵列(FPGA)等。

尽管上面已经示出和描述了本申请的实施例，可以理解的是，上述实施例是示例性的，不能理解为对本申请的限制，本领域的普通技术人员在本申请的范围内可以对上述实施例进行变化、修改、替换和变型。

权 利 要 求 书

1、一种实时视频的目标检测与识别方法，其特征在于，所述方法由计算设备执行，所述方法包括：

确定当前帧图像的前一帧图像的目标物体识别结果 R_{X-1} 和两帧之前的图像的目标物体识别结果 R_{X-2} 是否相同；

当两个识别结果 R_{X-1} 和 R_{X-2} 不不同时，通过第一级神经网络对所述当前帧图像进行目标物体位置检测，获取所述当前帧图像中目标物体的位置范围 C_X ；
当两个识别结果 R_{X-1} 和 R_{X-2} 相同时，不经过所述第一级神经网络，根据所述前一帧图像中目标物体的位置范围 C_{X-1} 确定所述当前帧图像中目标物体的位置范围 C_X ；

通过第二级神经网络，根据所述当前帧图像中目标物体的位置范围 C_X ，对所述当前帧图像进行目标物体识别，得到所述当前帧图像的目标物体识别结果 R_X 。

2、根据权利要求 1 所述的实时视频的目标检测与识别方法，其特征在于，所述通过第一级神经网络对所述当前帧图像进行目标物体位置检测，包括：

通过所述第一级神经网络判断所述当前帧图像中是否包含目标物体；

若所述当前帧图像包含目标物体，则通过所述第一级神经网络获取所述当前帧图像中目标物体的位置范围 C_X ；

若所述当前帧图像不含目标物体，则将判断帧图像中是否包含目标物体的间隔设置为每隔预定帧数执行一次，并通过所述第一级神经网络持续进行判断直至帧图像中包含目标物体。

3、根据权利要求 1 所述的实时视频的目标检测与识别方法，其特征在于，所述通过第二级神经网络，根据所述当前帧图像中目标物体的位置范围 C_X ，对所述当前帧图像进行目标物体识别，得到所述当前帧图像的目标物体识别结果 R_X ，包括：

将所述当前帧图像中目标物体的位置范围 C_X 扩展，得到面积增大预定义倍数的扩展位置范围；

通过所述第二级神经网络在所述扩展位置范围内进行目标物体识别，得到

所述当前帧图像的目标物体识别结果 R_x 。

4、根据权利要求 1 所述的实时视频的目标检测与识别方法，其特征在于，所述根据所述前一帧图像中目标物体的位置范围 C_{x-1} 确定所述当前帧图像中目标物体的位置范围 C_x ，包括：

将所述前一帧图像中目标物体的位置范围 C_{x-1} 作为所述当前帧图像中目标物体的位置范围 C_x ；或者

根据所述前一帧图像中目标物体的位置范围，以及之前预定义帧数的图像中目标物体的位置范围，预测所述当前帧图像中目标物体的位置范围 C_x 。

5、根据权利要求 1-4 中任一项所述的实时视频的目标检测与识别方法，其特征在于，

所述目标物体为人手；

所述通过第一级神经网络对所述当前帧图像进行目标物体位置检测，获取所述当前帧图像中目标物体的位置范围 C_x ，包括：通过所述第一级神经网络识别人手在所述当前帧图像中的位置范围；

所述通过第二级神经网络，根据所述当前帧图像中目标物体的位置范围 C_x ，对所述当前帧图像进行目标物体识别，得到所述当前帧图像的目标物体识别结果 R_x ，包括：通过所述第二级神经网络，根据所述当前帧图像中目标物体的位置范围 C_x ，对所述当前帧图像中的人手进行手势识别，识别结果为手势的姿态。

6、一种实时视频的目标检测与识别系统，其特征在于，包括：

第一级神经网络，用于对视频帧图像进行目标物体的位置检测，获取目标物体位置范围；

第二级神经网络，用于根据目标物体位置范围，在视频帧图像中进行目标物体的识别，得到识别结果；以及

神经网络运行控制模块，用于，

确定当前帧图像的前一帧图像的目标物体识别结果 R_{x-1} 和两帧之前的图像的目标物体识别结果 R_{x-2} 是否相同；

当两个识别结果 R_{x-1} 和 R_{x-2} 不相同，控制第一级神经网络对所述当前帧

图像进行目标物体位置检测，获取所述当前帧图像中目标物体的位置范围 C_X ；当两个识别结果 R_{X-1} 和 R_{X-2} 相同时，不经过所述第一级神经网络，根据所述前一帧图像中目标物体的位置范围 C_{X-1} 确定所述当前帧图像中目标物体的位置范围 C_X ；

控制第二级神经网络，根据所述当前帧图像中目标物体的位置范围 C_X ，对所述当前帧图像进行目标物体识别，得到所述当前帧图像的目标物体识别结果 R_X 。

7、根据权利要求 6 所述的实时视频的目标检测与识别系统，其特征在于，在控制第一级神经网络对所述当前帧图像进行目标物体位置检测时，所述神经网络运行控制模块，用于，

控制所述第一级神经网络判断所述当前帧图像中是否包含目标物体；

若所述当前帧图像包含目标物体，则控制所述第一神经网络获取所述当前帧图像中目标物体的位置范围 C_X ；

若所述当前帧图像不含目标物体，则将判断帧图像中是否包含目标物体的间隔设置为每隔预定帧数执行一次，并控制所述第一级神经网络持续进行判断直至帧图像中包含目标物体。

8、根据权利要求 6 所述的实时视频的目标检测与识别系统，其特征在于，在控制第二级神经网络根据所述当前帧图像中目标物体的位置范围 C_X ，对所述当前帧图像进行目标物体识别，得到所述当前帧图像的目标物体识别结果 R_X 时，所述神经网络运行控制模块，用于，

将所述当前帧图像中目标物体的位置范围 C_X 扩展，得到面积增大预定义倍数的扩展位置范围；

控制所述第二级神经网络在所述扩展位置范围内进行目标物体识别，得到所述当前帧图像的目标物体识别结果 R_X 。

9、根据权利要求 6 所述的实时视频的目标检测与识别系统，其特征在于，在根据所述前一帧图像中目标物体的位置范围 C_{X-1} 确定所述当前帧图像中目标物体的位置范围 C_X 时，所述神经网络运行控制模块，用于，

将所述前一帧图像中目标物体的位置范围 C_{X-1} 作为所述当前帧图像中目标物体的位置范围 C_X ；或者

根据所述前一帧图像中目标物体的位置范围，以及之前预定义帧数的图像中目标物体的位置范围，预测所述当前帧图像中目标物体的位置范围 C_X 。

10、根据权利要求 6-9 中任一项所述的实时视频的目标检测与识别系统，其特征在于，

所述目标物体为人手；

在控制第一级神经网络对所述当前帧图像进行目标物体位置检测，获取所述当前帧图像中目标物体的位置范围 C_X 时，所述神经网络运行控制模块，用于控制所述第一级神经网络识别人手在所述当前帧图像中的位置范围；

在控制第二级神经网络，根据所述当前帧图像中目标物体的位置范围 C_X ，对所述当前帧图像进行目标物体识别，得到所述当前帧图像的目标物体识别结果 R_X 时，所述神经网络运行控制模块，用于控制所述第二级神经网络，根据所述当前帧图像中目标物体的位置范围 C_X ，对所述当前帧图像中的人手进行手势识别，识别结果为手势的姿态。

11、一种非临时性计算机可读存储介质，其上存储有计算机程序，其特征在于，该程序被处理器执行时，实现根据权利要求 1-5 中任一项所述的实时视频的目标检测与识别方法。

12、一种计算机程序产品，其特征在于，当所述计算机程序产品中的指令由处理器执行时，实现根据权利要求 1-5 中任一项所述的实时视频的目标检测与识别方法。

13、一种计算设备，其特征在于，包括存储器、处理器及存储在存储器上并可在处理器上运行的计算机程序，所述处理器执行所述程序时，实现根据权利要求 1-5 中任一项所述的实时视频的目标检测与识别方法。

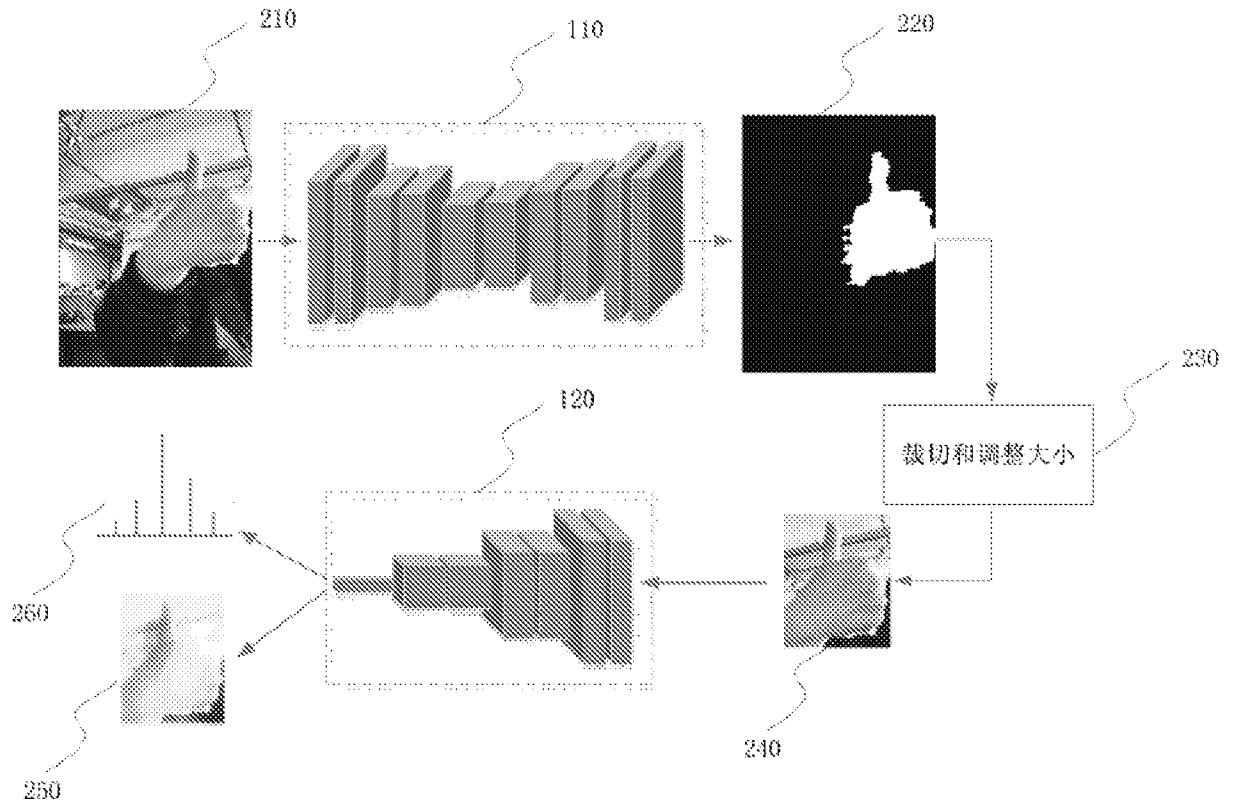


图 1

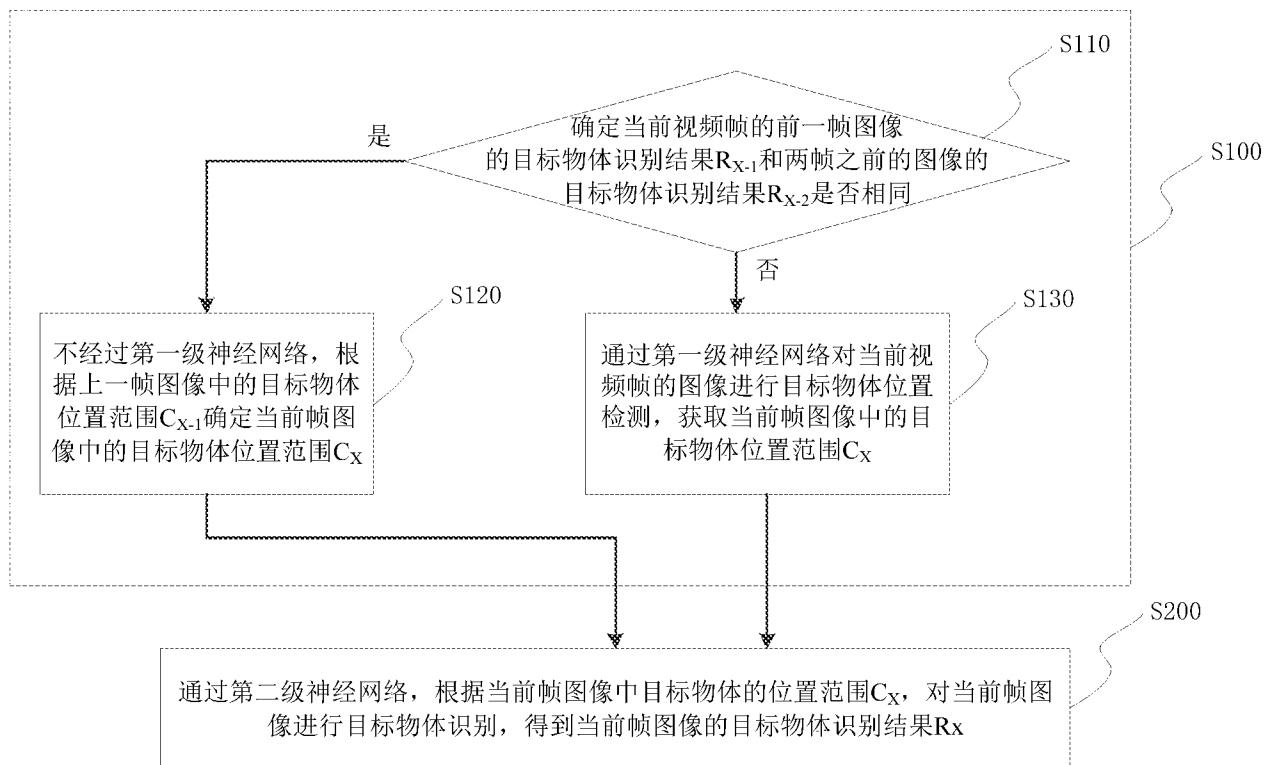


图 2

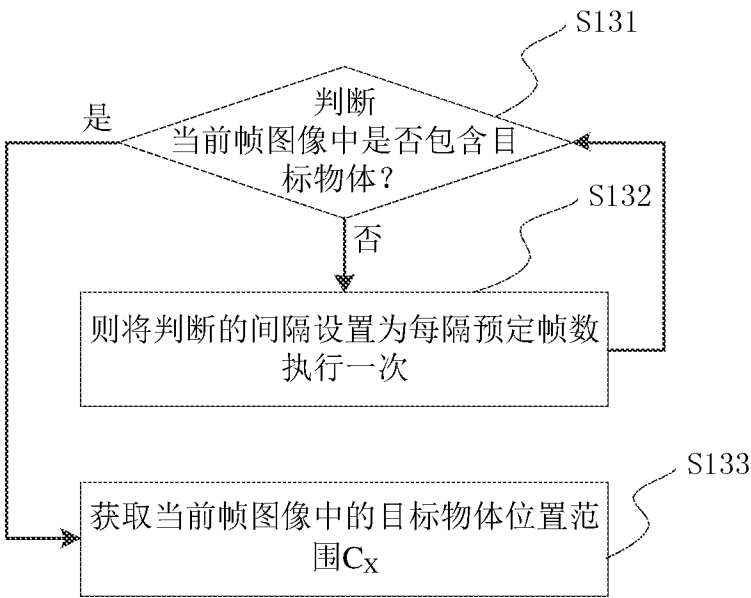


图 3

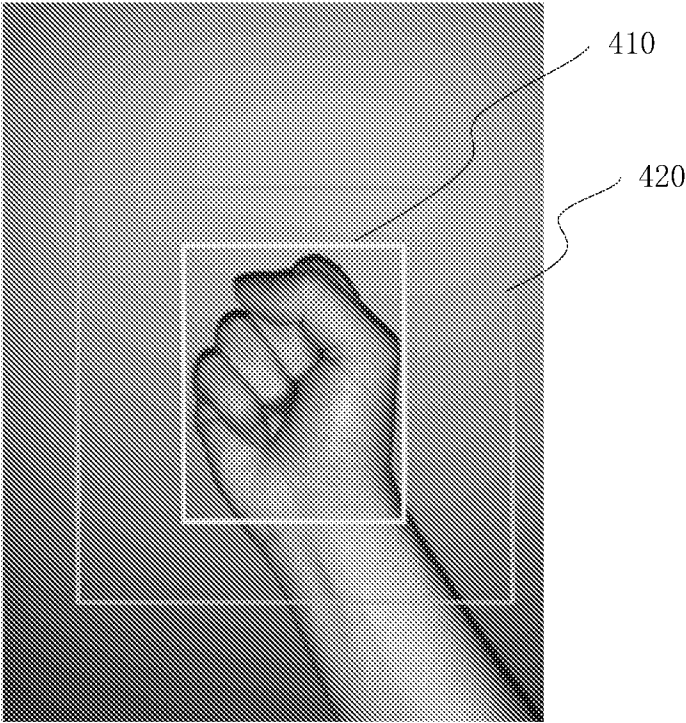


图 4

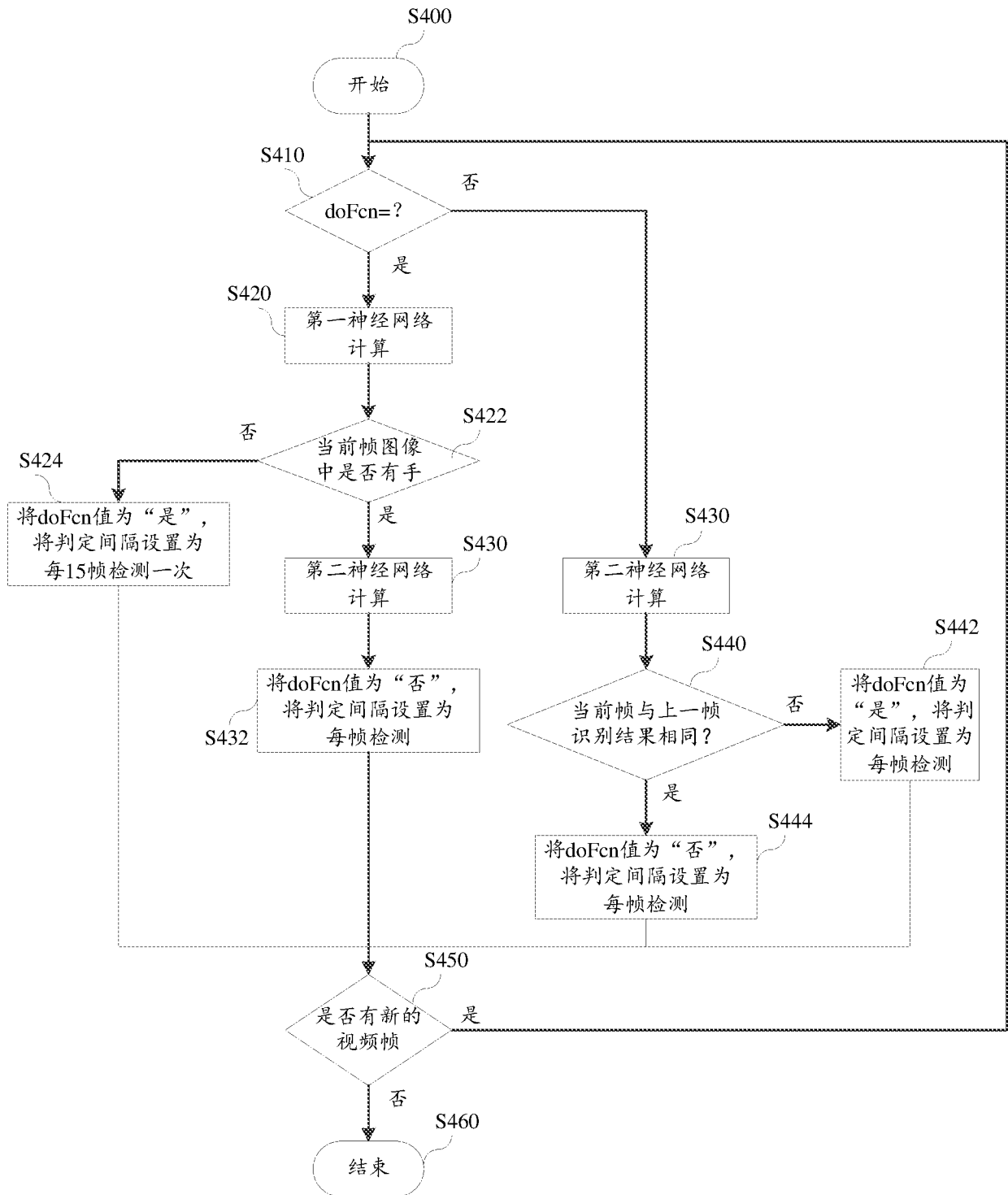


图 5

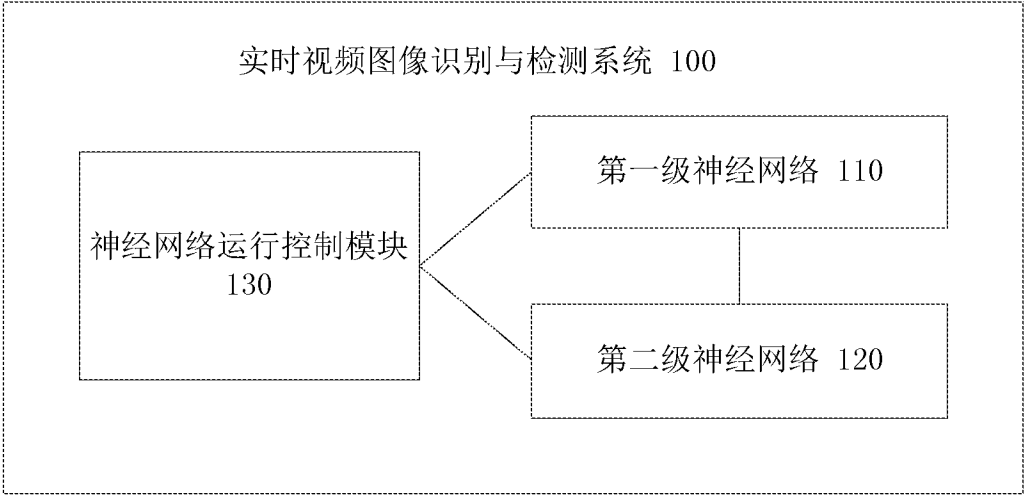


图 6

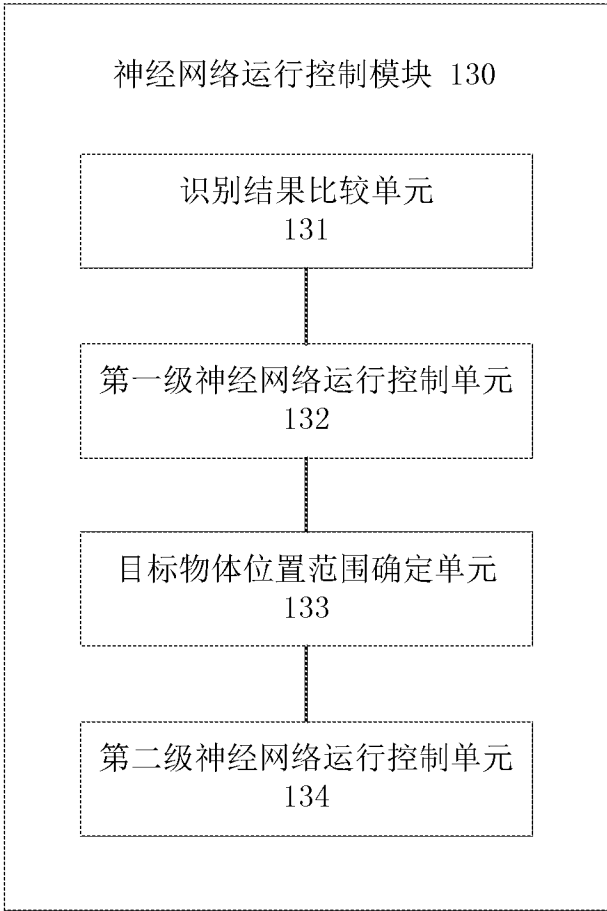


图 7

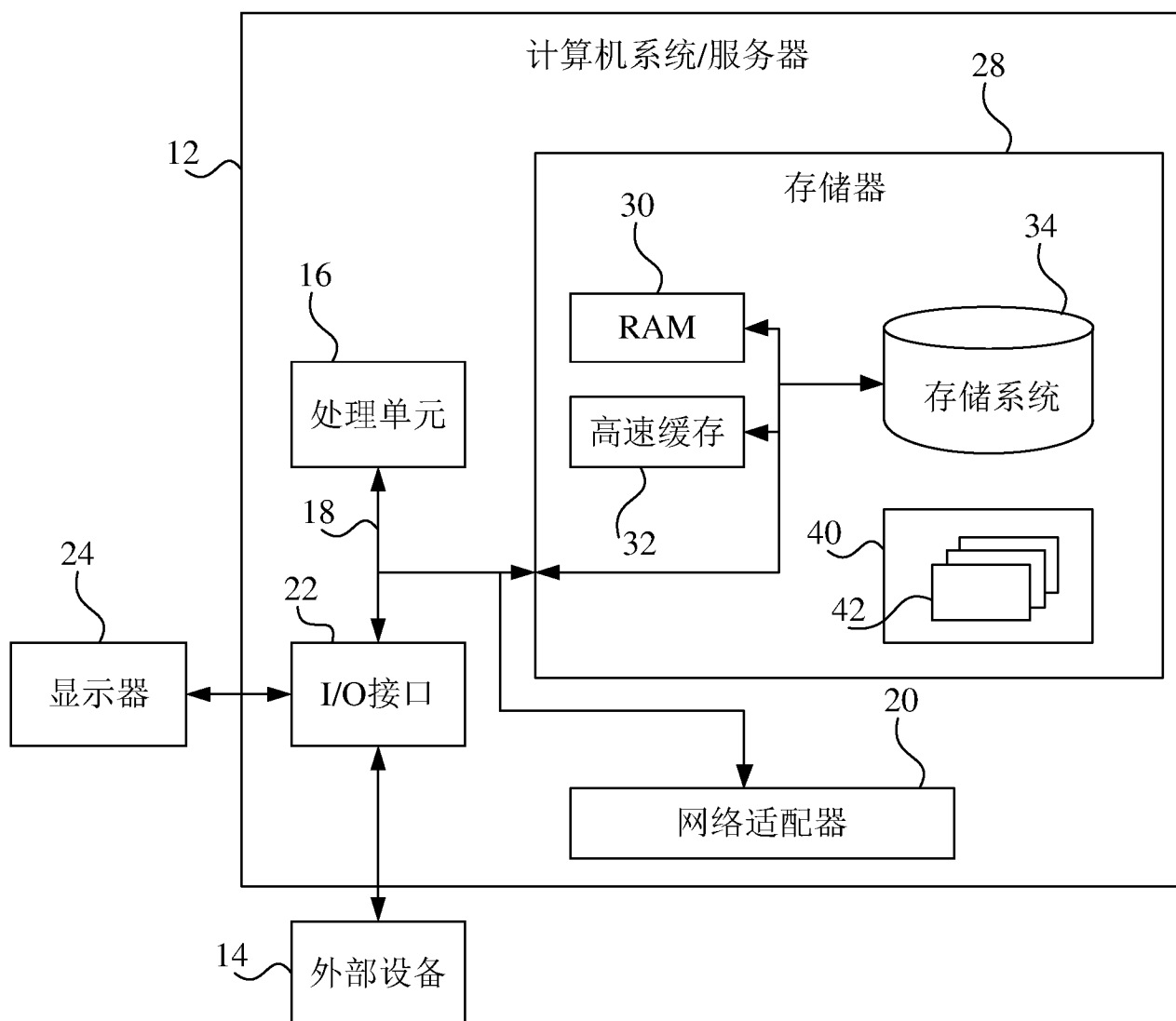


图 8

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/092791

A. CLASSIFICATION OF SUBJECT MATTER

G06K 9/00(2006.01)i; G06N 3/04(2006.01)n

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 9/-; G06N 3/-

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, CNKI, DWPI, SIPOABS: 目标, 手, 检测, 识别, 帧, 前, 位置, 范围, 神经网络, 第一, 第二, 级联, 两级, 二级, object, target, hand, detect+, recogni+, identifi+, frame, previous, before, location, position, range, neural network, first, second, cascade, two stage, two level

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	CN 108229277 A (BEIJING SENSETIME TECHNOLOGY DEVELOPMENT CO., LTD.) 29 June 2018 (2018-06-29) description, paragraphs [0073]-[0277]	1-13
A	CN 105787867 A (HUAWEI TECHNOLOGIES CO., LTD. ET AL.) 20 July 2016 (2016-07-20) entire document	1-13
A	CN 107492115 A (BEIJING XIAOMI MOBILE SOFTWARE CO., LTD.) 19 December 2017 (2017-12-19) entire document	1-13
A	CN 107944376 A (BEIJING QIHOO TECHNOLOGY CO., LTD.) 20 April 2018 (2018-04-20) entire document	1-13



Further documents are listed in the continuation of Box C.



See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“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)

“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

“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

“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

“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

“&” document member of the same patent family

Date of the actual completion of the international search

19 September 2019

Date of mailing of the international search report

26 September 2019

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.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/092791

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)			Publication date (day/month/year)
CN	108229277	A	29 June 2018	WO	2018177379	A1	04 October 2018
CN	105787867	A	20 July 2016	CN	105787867	B	05 March 2019
CN	107492115	A	19 December 2017	None			
CN	107944376	A	20 April 2018	None			

国际检索报告

国际申请号

PCT/CN2019/092791

A. 主题的分类 G06K 9/00(2006.01)i; G06N 3/04(2006.01)n 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 9/-; G06N 3/- 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, CNKI, DWPI, SIPOABS: 目标, 手, 检测, 识别, 帧, 前, 位置, 范围, 神经网络, 第一, 第二, 级联, 两级, 二级, object, target, hand, detect+, recogni+, identif+, frame, previous, before, location, position, range, neural network, first, second, cascade, two stage, two level																	
C. 相关文件 <table border="1"> <thead> <tr> <th>类 型*</th> <th>引用文件, 必要时, 指明相关段落</th> <th>相关的权利要求</th> </tr> </thead> <tbody> <tr> <td>X</td> <td>CN 108229277 A (北京市商汤科技开发有限公司) 2018年 6月 29日 (2018 - 06 - 29) 说明书第[0073]-[0277]段</td> <td>1-13</td> </tr> <tr> <td>A</td> <td>CN 105787867 A (华为技术有限公司等) 2016年 7月 20日 (2016 - 07 - 20) 全文</td> <td>1-13</td> </tr> <tr> <td>A</td> <td>CN 107492115 A (北京小米移动软件有限公司) 2017年 12月 19日 (2017 - 12 - 19) 全文</td> <td>1-13</td> </tr> <tr> <td>A</td> <td>CN 107944376 A (北京奇虎科技有限公司) 2018年 4月 20日 (2018 - 04 - 20) 全文</td> <td>1-13</td> </tr> </tbody> </table>			类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求	X	CN 108229277 A (北京市商汤科技开发有限公司) 2018年 6月 29日 (2018 - 06 - 29) 说明书第[0073]-[0277]段	1-13	A	CN 105787867 A (华为技术有限公司等) 2016年 7月 20日 (2016 - 07 - 20) 全文	1-13	A	CN 107492115 A (北京小米移动软件有限公司) 2017年 12月 19日 (2017 - 12 - 19) 全文	1-13	A	CN 107944376 A (北京奇虎科技有限公司) 2018年 4月 20日 (2018 - 04 - 20) 全文	1-13
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求															
X	CN 108229277 A (北京市商汤科技开发有限公司) 2018年 6月 29日 (2018 - 06 - 29) 说明书第[0073]-[0277]段	1-13															
A	CN 105787867 A (华为技术有限公司等) 2016年 7月 20日 (2016 - 07 - 20) 全文	1-13															
A	CN 107492115 A (北京小米移动软件有限公司) 2017年 12月 19日 (2017 - 12 - 19) 全文	1-13															
A	CN 107944376 A (北京奇虎科技有限公司) 2018年 4月 20日 (2018 - 04 - 20) 全文	1-13															
☐ 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, 但为了理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 2019年 9月 19日		国际检索报告邮寄日期 2019年 9月 26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		受权官员 朱晓莉 电话号码 (86-10)62411672															

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/092791

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	108229277	A	2018年 6月 29日	WO	2018177379	A1	2018年 10月 4日
CN	105787867	A	2016年 7月 20日	CN	105787867	B	2019年 3月 5日
CN	107492115	A	2017年 12月 19日	无			
CN	107944376	A	2018年 4月 20日	无			