

(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 6 月 4 日 (04.06.2020)



(10) 国际公布号
WO 2020/108362 A1

(51) 国际专利分类号:
G06K 9/00 (2006.01)

(21) 国际申请号: PCT/CN2019/119633

(22) 国际申请日: 2019 年 11 月 20 日 (20.11.2019)

(25) 申请语言: 中文

(26) 公布语言: 中文

(30) 优先权:
201811427578.X 2018 年 11 月 27 日 (27.11.2018) CN

(71) 申请人: 广州市百果园信息技术有限公司 (GUANGZHOU BAIGUOYUAN INFORMATION TECHNOLOGY CO., LTD.) [CN/CN]; 中国广东省广州市番禺区市桥街兴泰路 274 号 C 栋西塔 5-13 层, Guangdong 511402 (CN)。

(72) 发明人: 项伟 (XIANG, Wei); 中国广东省广州市番禺区市桥街兴泰路 274 号 C 栋西塔 5-13 层, Guangdong 511402 (CN)。王毅峰 (WANG, Yifeng); 中国广东省广州市番禺区市桥街兴泰路 274 号 C 栋西塔 5-13 层, Guangdong 511402 (CN)。黄秋实 (HUANG, Qiushi); 中国广东省广州市番禺区市桥街兴泰路 274 号 C 栋西塔 5-13 层, Guangdong 511402 (CN)。梁柱锦 (LIANG, Zhujin); 中国广东省广州市番禺区市桥街兴泰路 274 号 C 栋西塔 5-13 层, Guangdong 511402 (CN)。

(74) 代理人: 北京品源专利代理有限公司 (BEYOND ATTORNEYS AT LAW); 中国北京市海淀区莲花池东路 39 号西金大厦 6 层, Beijing 100036 (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,

(54) Title: BODY POSTURE DETECTION METHOD, APPARATUS AND DEVICE, AND STORAGE MEDIUM

(54) 发明名称: 人体姿态检测方法、装置、设备及存储介质

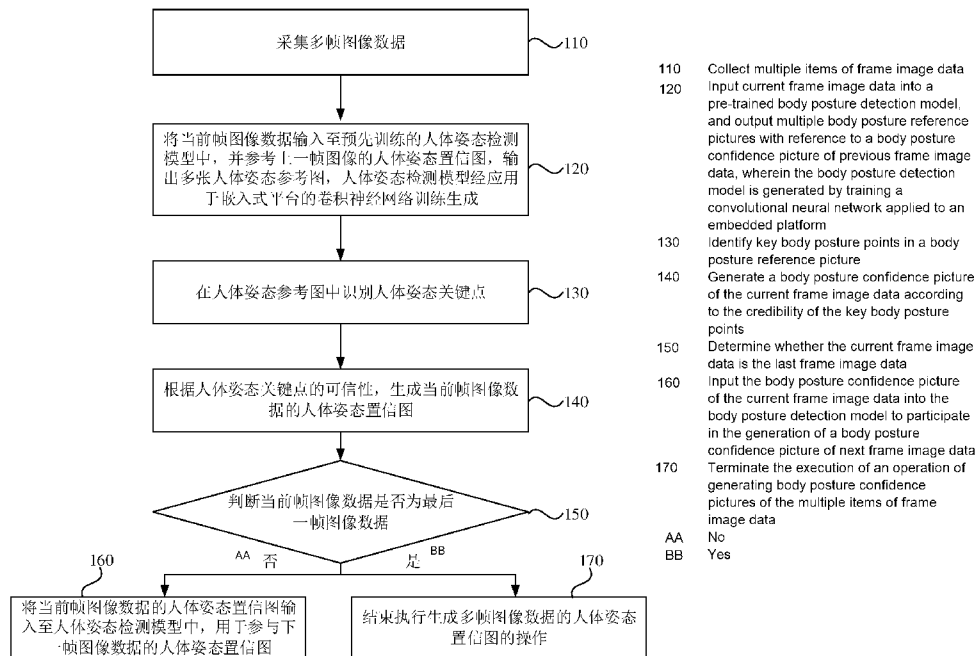


图 1

(57) Abstract: Disclosed are a body posture detection method, apparatus and device, and a storage medium. The method comprises: collecting multiple items of frame image data; inputting current frame image data into a pre-trained body posture detection model, and outputting multiple body posture reference pictures with reference to a body posture confidence picture of previous frame image data, wherein the body posture detection model is generated by training a convolutional neural network applied to an embedded platform; identifying key body posture points in each body posture reference picture; generating a body posture confidence picture of the current

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))。

frame image data according to the credibility of the key body posture points; determining whether the current frame image data is the last frame image data; when the current frame image data is not the last frame image data, inputting the body posture confidence picture of the current frame image data into the body posture detection model to participate in the generation of a body posture confidence picture of next frame image data; and when the current frame image data is the last frame image data, terminating the execution of an operation of generating body posture confidence pictures of the multiple items of frame image data.

(57) 摘要: 本申请公开了一种人体姿态检测方法、装置、设备及存储介质。该方法包括: 采集多帧图像数据; 将当前帧图像数据输入至预先训练的人体姿态检测模型中, 并参考上一帧图像数据的人体姿态置信图, 输出多张人体姿态参考图, 人体姿态检测模型经应用于嵌入式平台的卷积神经网络训练生成; 在每张人体姿态参考图中识别人体姿态关键点; 根据人体姿态关键点的可信性, 生成当前帧图像数据的人体姿态置信图; 判断当前帧图像数据是否为最后一帧图像数据; 在当前帧图像数据不是最后一帧图像数据的情况下, 将当前帧图像数据的人体姿态置信图输入至人体姿态检测模型中, 用于参与生成下一帧图像数据的人体姿态置信图; 在当前帧图像数据是最后一帧图像数据的情况下, 结束执行生成多帧图像数据的人体姿态置信图的操作。

人体姿态检测方法、装置、设备及存储介质

本申请要求在2018年11月27日提交中国专利局、申请号为201811427578.X的中国专利申请的优先权，该申请的全部内容通过引用结合在本申请中。

技术领域

本申请实施例涉及人体姿态检测技术，例如一种人体姿态检测方法、装置、设备及存储介质。

背景技术

人体姿态检测是计算机视觉领域中最具挑战性的研究方向，被广泛应用于人机交互、智能监控、虚拟现实和人体行为分析等领域。但是，由于组成人体姿态的各个关键点所在的局部图像特征呈多尺度的仿射变换，并且图像容易受目标人物着装、相机的拍摄角度、距离、光照变化和局部遮挡等因素影响，使得人体姿态检测研究进展缓慢。

相关技术中，通常采用基于卷积神经网络来进行人体姿态检测，同时，为了获得较高的识别精度，通常需要采集大量的训练样本对人体姿态检测模型进行长时间监督学习。

在实现本申请过程中，申请人发现相关技术中至少存在如下问题：由于嵌入式平台中没有图形处理器（Graphics Processing Unit, GPU）对卷积神经网络中计算量最大的卷积操作进行优化，因此，大量基于卷积神经网络的人体姿态检测方法无法应用于嵌入式平台。

发明内容

本申请实施例提供一种人体姿态检测方法、装置、设备及存储介质，以实现在嵌入式平台上的人体姿态检测。

第一方面，本申请实施例提供了一种人体姿态检测方法，该方法包括：采集多帧图像数据；将当前帧图像数据输入至预先训练的人体姿态检测模型中，并参考上一帧图像数据的人体姿态置信图，输出多张人体姿态参考图，所述人体检测模型经应用于嵌入式平台的卷积神经网络训练生成；在每张人体姿态参考图中识别人体姿态关键点；根据所述人体姿态关键点的可信性，生成当前帧图像数据的人体姿态置信图；判断当前帧图像数据是否为最后一帧图像数据；在当前帧图像数据不是最后一帧图像数据的情况下，将所述当前帧图像数据的人体姿态置信图输入至所述人体姿态检测模型中，用于参与生成下一帧图像数据的人体姿态置信图；在当前帧图像数据是最后一帧图像数据的情况下，结束

执行生成多帧图像数据的人体姿态置信图的操作。

第二方面，本申请实施例提供了一种人体姿态检测装置，该装置包括：图像数据采集模块，设置为采集多帧图像数据；人体姿态参考图输出模块，设置为将当前帧图像数据输入至预先训练的人体姿态检测模型中，并参考上一帧图像数据的人体姿态置信图，输出多张人体姿态参考图，所述人体姿态检测模型经应用于嵌入式平台的卷积神经网络训练生成；人体姿态关键点识别模块，设置为在每张人体姿态参考图中识别人体姿态关键点；人体姿态置信图生成模块，设置为根据所述人体姿态关键点的可信性，生成人体姿态置信图；判断模块，设置为判断当前帧图像数据是否为最后一帧图像数据；第一执行模块，设置为在当前帧图像数据不是最后一帧图像数据的情况下，将所述当前帧图像数据的人体姿态置信图输入至所述人体姿态检测模型中，用于参与生成下一帧图像数据的人体姿态置信图；第二执行模块，设置为在当前帧图像数据是最后一帧图像数据的情况下，结束执行生成多帧图像数据的人体姿态置信图的操作。

第三方面，本申请实施例还提供了一种设备，该设备包括：至少一个处理器；存储器，设置为存储至少一个程序；当所述至少一个程序被所述至少一个处理器执行，使得所述至少一个处理器实现如本申请实施例第一方面所述的方法。

第四方面，本申请实施例还提供了一种计算机可读存储介质，其上存储有计算机程序，所述计算机程序被处理器执行时实现如本申请实施例第一方面所述的方法。

附图说明

- 图 1 是本申请实施例中的一种人体姿态检测方法的流程图；
- 图 2 是本申请实施例中的一种卷积神经网络的应用示意图；
- 图 3 是本申请实施例中的另一种人体姿态检测方法的流程图；
- 图 4 是本申请实施例中的一种人体姿态检测装置的结构示意图；
- 图 5 是本申请实施例中的一种设备的结构示意图。

具体实施方式

所谓计算机视觉，就是让计算机模拟人的视觉功能，能够像人一样通过观察来理解客观世界。它研究的主要内容是：如何利用计算机视觉技术解决以人为中心的相关问题，包括物体识别、人脸识别、人体检测与跟踪、人体姿态检测和人体运动分析等。人体姿态检测是人体行为识别的重要组成部分，也是人体行为识别系统的重要研究内容，它最终的目的是输出人的整体或者局部肢体的结构参数，如人体轮廓、头部的位置与朝向、人体关键点的位置或者部位类

别。它在很多方面都有重要应用，示例性的，如运动员动作识别、动画人物制作以及基于内容的图像和视频检索等。

针对人体姿态检测来说，可将人体看成是由关键点相连接的不同部件组成，人体姿态检测可以通过获取各个关键点的位置信息来确定，其中，关键点的位置信息可以用一个平面二维坐标来表示。人体姿态检测通常需要获取人体的头部、脖子、左肩、右肩、左肘、右肘、左手腕、右手腕、左臀、右臀、左膝、右膝、左脚踝和右脚踝，共计 14 个关键点。

传统技术中，可以采用基于卷积神经网络的人体姿态检测方法来进行人体姿态检测，其中，卷积神经网络解决的核心问题就是如何自动提取并抽象特征，进而将特征映射到任务目标解决实际问题，一个卷积神经网络一般由以下三部分组成，第一部分是输入层，第二部分由卷积层、激励层和池化层（或下采样层）组合而成，第三部分由一个全连结的多层感知机分类器构成。卷积神经网络具有权值共享特性，权值共享即指可以通过一个卷积核的卷积操作提取整张图像不同位置的同一特征，换句话说，即是在一张图像数据中的不同位置的相同目标，它们的局部特征是基本相同的。可以理解到，使用一个卷积核只能得到一种特征，可以通过设置多核卷积，用每个卷积核来学习不同的特征来提取图像数据的特征。可以理解到，在图像处理中，卷积层的作用是将低层次的特征抽取并聚合为高层次特征，低层次的特征是基本特征，诸如纹理和边缘等局部特征，高层次特征如人脸和物体的形状等，更能表现样本的全局属性，这个过程就是卷积神经网络对目标物体层级概括性。

可以理解到，如果想实现基于卷积神经网络的人体姿态检测方法可以在嵌入式平台上运行，需要该卷积神经网络的计算量较小、运行速度快且预测精度满足实际需求。

为了避免基于卷积神经网络的人体姿态检测方法无法实现在嵌入式平台运行的情况，可考虑对卷积神经网络进行改进，例如可以采用轻量化卷积神经网络，本申请实施例所提供的卷积神经网络即指轻量化卷积神经网络。所谓轻量化卷积神经网络指的是可以应用于嵌入式平台的卷积神经网络。

下面将结合具体实施例对人体姿态检测方法进行说明。

图 1 为本申请实施例提供的一种人体姿态检测方法的流程图，本实施例可适用于检测人体姿态的情况，该方法可以由人体姿态检测装置来执行，该装置可以采用软件和硬件中至少之一的方式实现，该装置可以配置于设备中，例如典型的是计算机或移动终端等。如图 1 所示，该方法包括步骤 110 至步骤 170。

在步骤 110 中，采集多帧图像数据。

在本申请的实施例中，视频可以理解为是由至少一帧图像数据组成的，因此，为了对视频中的人体姿态进行识别，可以将视频划分为一帧一帧的图像数

据，分别对每帧图像数据进行分析。这里多帧图像数据表示的是同一视频中的图像数据，换句话说，该视频包括多帧图像数据。可以按照时间顺序对多帧图像数据进行命名。示例性的，如视频包括N帧图像数据， $N \geq 1$ ，此时，按照时间顺序可将上述N帧图像数据称为：第一帧图像数据、第二帧图像数据、.....、第N-1帧图像数据以及第N帧图像数据。

可以理解到，在将视频划分为多帧图像数据的情况下，可以按照时间顺序依次对每帧图像数据进行处理。同时，可将当前正在处理的某帧图像数据称为当前帧图像数据，将当前帧图像数据的上一帧图像数据称为上一帧图像数据，将当前帧图像数据的下一帧图像数据称为下一帧图像数据。当前可以理解到，如果当前帧数据为第一帧图像数据，则对该当前帧图像数据来说，其只有下一帧图像数据而没有上一帧图像数据；如果当前帧图像数据是最后一帧图像数据，则对当前帧图像数据来说，其只有上一帧图像数据而没有下一帧图像数据；如果当前帧图像数据既不是第一帧图像数据也不是最后一帧图像数据，则对当前帧图像数据来说，其有上一帧图像数据也有下一帧图像数据。

采用上述按照时间顺序依次对每帧图像数据进行处理的原因在于：对于人体姿态检测来说，相邻两帧图像数据之间可能存在一定的关联性，即如果根据上一帧图像数据识别出某关键点出现在了上一帧图像中的某个位置，则当前帧图像数据中该关键点也可能出现在当前帧图像数据中的同一位置附近。换句话说，如果上一帧图像数据的检测结果满足预设条件，则可以参考上一帧图像数据的检测结果，对当前帧图像数据进行处理。

在步骤120中，将当前帧图像数据输入至预先训练的人体姿态检测模型中，并参考上一帧图像数据的人体姿态置信图，输出多张人体姿态参考图，人体姿态检测模型经应用于嵌入式平台的卷积神经网络训练生成。

在本申请的实施例，人体姿态置信图可以指包括人体姿态关键点的图像，或者，人体姿态置信图可以理解为是基于人体姿态关键点所生成的图像，如以人体姿态关键点为中心生成的图像。这里所述的人体姿态关键点可以指前文所述的头部、脖子、左肩、右肩、左肘、右肘、左手腕、右手腕、左臀、右臀、左膝、右膝、左脚踝和右脚踝等14个关键点。

人体姿态参考图可以包括两方面的内容，即可能作为人体姿态关键点的多个点的位置信息以及所述位置信息对应的概率值，其中，可将可能作为人体姿态关键点的点称为候选点，相应的，人体姿态参考图可以包括多个候选点的位置信息以及该位置信息对应的概率值，即每个候选点对应一个概率值，位置信息可以以坐标形式表示。同时，可以根据多个候选点的位置信息对应的概率值确定将哪个候选点作为人体姿态关键点。示例性的，如选择多个候选点的位置信息对应的多个概率值中最大概率值对应的候选点作为人体姿态关键点。某人

体姿态参考图中包括候选点 A 的位置信息 (x_A, y_A) 以及对应的概率值 P_A ；候选点 B 的位置信息 (x_B, y_B) 以及对应的概率值 P_B ；候选点 C 的位置信息 (x_C, y_C) 以及对应的概率值 P_C ，其中， $P_A < P_B < P_C$ ，基于上述，确定将候选点 C 作为人体姿态关键点。

需要说明的是，每张人体姿态置信图对应一个人体姿态关键点，每张人体姿态参考图包括多个候选点，所述候选点是针对某个关键点的候选点，如某张人体姿态参考图包括多个候选点，所述候选点是针对左肘的候选点。再如某张人体姿态参考图也包括多个候选点，所述候选点是针对左膝的候选点。基于上述，可以理解到，针对某帧图像数据，需要从该帧图像数据中确定 N 个关键点，则对应存在 N 张人体姿态参考图以及 N 张人体姿态置信图。

预先训练的人体姿态检测模型可以由设定数量组的训练样本经应用于嵌入式平台的卷积神经网络训练生成，可应用于嵌入式平台的卷积神经网络即是轻量化卷积神经网络，人体姿态检测模型可以包括主路、第一支路、第二支路和第三支路；主路可以包括残差模块和上采样模块，第一支路可以包括提炼网络模块，第二支路可以包括反馈模块；残差模块可以包括第一残差单元、第二残差单元和第三残差单元。对于人体姿态检测模型的组成部分的详细说明可参见后文。

将当前帧图像数据输入至预先训练的人体姿态检测模型中，并参考上一帧图像数据的人体姿态置信图，输出多张人体姿态参考图，可分如下两种情况：

情况一、将当前帧图像数据作为输入变量输入至预先训练的人体姿态模型中，得到多张第一人体姿态参考图，并根据上一帧图像数据得到的多张人体姿态置信图，输出多张人体姿态参考图，其中，每张第一人体姿态参考图根据对应的上一帧图像数据得到的多张人体姿态置信图中的某张人体姿态置信图，输出当前帧图像数据的一张人体姿态参考图，上述的对应关系是基于关键点是否相同确定的。示例性的，如当前帧图像数据的某张第一人体姿态参考图针对的关键点是左肘，则其参考的是上一帧图像上数据中对应关键点是左肘的人体姿态置信图。

可以理解到，针对情况一，上一帧图像数据的人体姿态置信图并未作为输入变量，同当前帧图像数据一起输入至预先训练的人体姿态检测模型中，而是在当前帧图像数据输入至预先训练的人体姿态检测模型，得到多张第一人体姿态参考图后，根据上一帧图像数据的多张人体姿态置信图，依次确定每张第一人体姿态参考图是否可信，在该张第一人体姿态参考图可信的情况下，将该张第一人体姿态参考图作为该当前帧的人体姿态参考图；在该张第一人体姿态参考图不可信的情况下，将上一帧图像数据中对于该张第一人体姿态参考图的人体姿态置信图作为该当前帧的人体姿态参考图。

情况二、将当前帧图像数据和上一帧图像数据的人体姿态置信图作为输入变量输入至预先训练的人体姿态检测模型中，输出多张人体姿态参考图。

可以理解到，上述情况二中上一帧图像数据的人体姿态置信图也作为输入变量，同当前帧图像数据一起输入至预先训练的人体姿态检测模型中，针对视频来说，相邻两帧图像数据之间具有一定的关联性，将上一帧图像数据的结果作为反馈信息，输入至预先训练的人体姿态检测模型中，参与到预测当前帧图像数据的输出结果的进程中，可提高人体姿态检测模型的预测精度。

需要说明的是，针对第二种情况，为了提高人体姿态检测模型的预测精度，可采用如下方式：判断上一帧图像数据的人体姿态置信图是否可信；在上一帧图像数据的人体姿态置信图可信的情况下，将当前帧图像数据和上一帧图像数据的人体姿态置信图输入至预先训练的人体姿态检测模型中，输出多张人体姿态参考图；在上一帧图像数据的人体姿态置信图不可信的情况下，将当前帧图像数据和预设图像数据输入至预先训练的人体姿态检测模型中，输出多张人体姿态参考图；或者，在上一帧图像数据的人体姿态置信图不可信的情况下，将当前帧图像数据输入至预先训练的人体姿态检测模型中，输出多张人体姿态参考图。其中，预设图像数据指的是不包含先验知识的图像数据，如全黑图像，若以矩阵表形式表示，即为全零矩阵。对于当前帧图像数据的输出结果来说，上一帧图像数据的人体姿态置信图即是包含先验知识的图像数据；对于下一帧图像数据的输出结果来说，当前帧图像数据的人体姿态置信图即是包含先验知识的图像数据。

采用上述方式可提高人体姿态检测模型的预测精度的原因在于：如果上一帧图像数据的人体姿态置信图不可信，则可以说明上一帧图像数据的人体姿态置信图并不可靠，如果在上述情况下，仍然将其作为输入变量也输入至预先训练的人体姿态检测模型中，不但不会提高人体姿态检测模型的预测精度，反而可能降低人体姿态模型的预测精度。基于上述，需要确保作为输入变量输入至预先训练的人体姿态检测模型中的上一帧图像数据的人体姿态置信图是可信的，因此，在确定是否参考上一帧图像数据的人体姿态置信图之前，采用了判断上一帧图像数据的人体姿态置信图是否可信的方式来实现，在上一帧图像数据的人体姿态置信图可信的情况下，将上一帧图像数据的人体姿态置信图作为输入变量输入至预先训练的人体姿态检测模型中，相反的，在上一帧图像数据的人体姿态置信图不可信的情况下，不将其作为输入变量。可采用如下方式判断上一帧图像数据的人体姿态置信图是否可信：在上一帧的人体姿态参考图中识别人体姿态关键点，在人体关键点对应的概率值大于预设的阈值的情况下，以人体姿态关键点作为中心生成掩模图，作为上一帧的人体姿态置信图，并确定上一帧的人体姿态置信图可信；在人体关键点对应的概率值小于等于预设的阈值

的情况下，以预设图像数据作为人体姿态置信图，并确定上一帧的人体姿态置信图不可信。

还需要说明的是，上述所述的多张人体姿态参考图针对的是当前帧图像数据的输出结果，即当前帧图像数据对应多张人体姿态参考图，示例性的，在当前帧图像数据中确定 N 个关键点的情况下，对应输出 N 张人体姿态参考图。同时，作为参考的上一帧图像数据的人体姿态置信图也包括 N 张。

另需要说明的是，上述所述的判断上一帧图像数据的人体姿态置信图是否可信指的是分别判断上一帧图像数据的每张人体姿态置信图是否可信。还可以理解到，由于人体姿态置信图可以指包括关键点的图像，不同关键点对应不同的人体姿态置信图，因此，针对不同关键点，判断人体姿态置信图是否可信的条件可以相同，也可以不同，可根据实际情况进行确定，在此不作限定。

此外，如果当前帧图像数据是第一帧图像数据，即其不存在上一帧图像数据，则可将当前帧图像数据输入至预先训练的人体姿态检测模型中，或者，可将当前帧图像数据和预设图像数据输入至预先训练的人体姿态检测模型中。

在一些实施例，将当前帧图像数据输入至预先训练的人体姿态检测模型中，并参考上一帧图像数据的人体姿态置信图，输出多张人体姿态参考图，包括：判断上一帧图像数据的人体姿态置信图是否可信。在上一帧图像数据的人体姿态置信图可信的情况下，将当前帧图像数据和上一帧图像数据的人体姿态置信图输入至预先训练的人体姿态检测模型中，输出多张人体姿态参考图。在上一帧图像数据的人体姿态置信图不可信的情况下，将当前帧图像数据和预设图像数据输入至预先训练的人体姿态检测模型中，输出多张人体姿态参考图。

在本申请的实施例，为了提高人体姿态检测模型的预测精度，可考虑采用如下方式：判断上一帧图像数据的人体姿态置信图是否可信；在上一帧图像数据的人体姿态置信图可信的情况下，将当前帧图像数据和上一帧图像数据的人体姿态置信图输入至预先训练的人体姿态检测模型中，输出多张人体姿态参考图；在上一帧图像数据的人体姿态置信图不可信的情况下，则可将当前帧图像数据和预设图像数据输入至预先训练的人体姿态检测模型中，输出多张人体姿态参考图。

经过上述操作，确保作为输入变量输入至预先训练的人体姿态检测模型中的上一帧图像数据的人体姿态置信图是可信的，进而依据上一帧图像数据的人体姿态置信图所提供的先验知识，提高人体姿态检测模型对当前帧图像数据的输出结果的预测精度。

示例性的，如上一帧图像数据的人体姿态置信图有 N 张，分别判断 N 张人体姿态置信图是否可信，判断结果为 x 张人体姿态置信图可信， $(N-x)$ 张人体姿态置信图不可信，则可将 x 张可信的人体姿态置信图、 $(N-x)$ 张预设图像数

据和当前帧图像数据输入至人体姿态检测模型中，输出多张人体姿态参考图。

在一些实施例中，将当前帧图像数据输入至预先训练的人体姿态检测模型中，并参考上一帧图像数据的人体姿态置信图，输出多张人体姿态参考图之前，包括：分别对每帧图像数据进行预处理，得到处理后的图像数据。

在本申请的实施例中，预处理可以包括归一化和白化，其中，归一化是指通过一系列变换，即利用图像的不变矩寻找一组参数使其能够消除其他变换函数对图像变换的影响，将待处理的原始图像转换成相应的唯一标准形式，该标准形式图像对平移、旋转或缩放等仿射变换具有不变特性。通常归一化包括如下步骤：即坐标中心化、x-shearing 归一化、缩放归一化和旋转归一化。在将当前帧图像数据输入至预先训练的人体姿态检测模型之前，该人体姿态检测模型可以基于神经网络训练生成，将图像数据进行归一化所起到的作用是归纳统一样本的统计分布性，进而加快网络学习速度，保证输出数据中数值小的不被吞食。

由于图像数据中相邻像素之间具有很强的相关性，因此作为输入变量输入时是冗余的。白化的作用即是降低输入的冗余性，更准确的说，通过白化处理，使得输入变量具有如下性质：特征之间相关性较低；所有特征具有相同的方差，通常在图像处理中设置为单位方差。

可以理解到，在对图像数据进行预处理后，作为输入变量输入至预先训练的人体姿态检测模型中的当前帧图像数据便是经过处理后的图像数据。当然，上一帧图像数据也是经过处理后的图像数据。

在步骤 130 中，在每张人体姿态参考图中识别人体姿态关键点。

在本申请的实施例中，根据前文所述可知，人体姿态参考图可以包括两方面的内容，即可能作为人体姿态关键点的每个点的位置信息以及该位置信息对应的概率值，其中，人体姿态关键点可以指确定作为关键点的点，换句话说，人体姿态关键点即为关键点，同时，可将可能作为人体姿态关键点的点称为候选点。

基于上述，可以理解到，人体姿态参考图包括多个候选点的位置信息以及位置信息对应的概率值，可以根据多个候选点的位置信息对应的概率值确定将哪个候选点作为人体姿态关键点。示例性的，如选择多个候选点的位置信息对应的概率值中最大概率值对应的候选点作为人体姿态关键点。

在一些实施例中，所述每张人体姿态参考图包括人体姿态关键点的多个候选点，每个候选点的坐标位置对应一个概率值；在人体姿态参考图中识别人体姿态关键点，包括：在人体姿态参考图中确定所述多个候选点的坐标位置对应的多个概率值中最大概率值对应的坐标位置，将所述坐标位置对应的候选点作为人体姿态关键点。

在本申请的实施例中，由于人体姿态参考图包括可能作为人体姿态关键点的多个点的位置信息以及该位置信息对应的概率值，因此，可以根据多个点的位置信息所对应的概率值，确定将哪个点作为人体姿态关键点。示例性的，在人体姿态参考图中确定最大概率值的坐标位置，将坐标位置作为人体姿态关键点。

需要说明的是，针对每张人体姿态参考图来说，其只有一个人体姿态关键点。采用上述根据概率值的方式确定人体姿态关键点，可能存在如下情况，在人体姿态参考图中有至少两个概率值相等，且均大于其它概率值，则可以根据实际情况，如关节连接是否合理，确定将哪个概率值的坐标位置作为人体姿态关键点。示例性的，如人体姿态参考图中有两个概率值相等且均大于其它概率值，两个概率值的坐标位置分别为 A 和 B，分别将 A 和 B 作为人体姿态关键点进行关节连接是否合理的判断，判断结果为：在将 A 作为人体姿态关键点的情况下，关节连接不合理；在将 B 作为人体姿态关键点的情况下，关节连接合理。因此，确定 B 为人体姿态关键点。

在步骤 140 中，根据人体姿态关键点的可信性，生成人体姿态置信图。

在本申请的实施例中，可信性可以包括可信和不可信，确定可信和不可信的标准可以为：人体姿态关键点对应的概率值是否大于预设的阈值，即在人体姿态关键点对应的概率值大于预设的阈值的情况下，可以说明该人体姿态关键点可信；在人体姿态关键点对应的概率值小于等于预设的阈值的情况下，可以说明该人体姿态关键点不可信。

在此基础上，在人体姿态关键点可信的情况下，以人体姿态关键点为中心生成掩模图，作为人体姿态置信图；在人体姿态关键点不可信的情况下，可将预设图像数据作为人体姿态置信图。这里所述的预设图像数据与前文所述的预设图像数据相同，预设图像数据可以为全黑图像，在以矩阵表形式表示的情况下，为零矩阵。其中，可采用如下方式判断人体姿态关键点是否可信：判断人体姿态关键点的概率值是否大于预设的阈值。在人体姿态关键点的概率值大于预设的阈值的情况下，可以说明该人体姿态关键点可信；在人体姿态关键点的概率值小于等于预设的阈值的情况下，可以说明该人体姿态关键点不可信。

需要说明的是，在确定人体姿态关键点不可信的情况下，将上一帧图像数据中对应的人体姿态关键点作为当前帧人体姿态关键点，但是，针对不可信的人体姿态关键点来说，其人体姿态置信图并不是根据上一帧图像数据中对应的人体姿态关键点生成的，而是根据预设图像数据人体姿态置信图生成的。

在一些实施例中，根据人体姿态关键点的可信性，生成人体姿态置信图，包括：判断人体姿态关键点是否可信。在人体姿态关键点可信的情况下，以人体姿态关键点作为中心生成掩模图，作为人体姿态置信图。在人体姿态关键点

不可信的情况下，将预设图像数据作为人体姿态置信图。

在本申请的实施例中，掩模图即指对图像进行图像掩膜处理后得到的图像。其中，图像掩膜是指用选定的图像、图形或物体，对待处理的图像（全部或局部）进行遮挡来控制图像处理的区域或处理过程。其中，用于覆盖的特定图像或物体称为掩膜或模板。在数字图像处理中，掩膜可以为二维矩阵数组，也可以为多值图像，图像掩膜设置为：其一、提取感兴趣区域。即用预先制作的感兴趣区域掩膜与待处理图像相乘，得到感兴趣区域图像，感兴趣区域内图像值保持不变，而区域外图像值均为零；其二、屏蔽作用。即用掩膜对待处理图像上某些区域作屏蔽，使其不参与处理、不参与处理参数的计算或者仅对所掩区域作处理、统计；其三、结构特征提取。即用相似性模板或图像匹配方法检测和提取待处理图像中与掩膜相似的结构特征；其四、制作特殊形状的图像。

根据人体姿态关键点的可信性，将人体姿态参考图生成人体姿态置信图，包括：在人体姿态关键点可信的情况下，以人体姿态关键点作为中心生成掩模图，作为人体姿态置信图，示例性的，在人体姿态关键点可信的情况下，以人体姿态关键点作为中心，并使用高斯核生成掩模图，作为人体姿态置信图。需要说明的是，可以通过设置高斯核的参数来确定掩模图所影响区域，其中，高斯核的参数包括滤波窗口的宽度和高度，高斯核可以为二维高斯核。示例性的，如某高斯核为二维高斯核，该二维高斯核的参数为滤波窗口的宽度为7，高度为7，即掩模图所影响区域为 7×7 的方形区域。

需要说明的是，在人体姿态关键点不可信的情况下，可以将预设图像数据作为人体姿态置信图，也可将预设图像数据认为是一种掩模图。这里所述的预设图像数据与前文所述的预设图像数据相同，预设图像数据可以为全黑图像，在以矩阵表形式表示的情况下，为零矩阵。

在一些实施例中，判断人体姿态关键点是否可信，包括：判断人体关键点对应的概率值是否大于预设的阈值。在人体关键点对应的概率值大于预设的阈值的情况下，确定人体关键点可信。在人体关键点对应的概率值小于等于预设的阈值的情况下，确定人体关键点不可信。

在本申请的实施例中，需要说明的是，阈值可以根据实际情况进行设定，在此不作限定。此外，不同人体姿态关键点对应的阈值可以相同，也可以不同，也可以根据实际情况进行确定，在此不作限定，如对于重要的人体姿态关键点，可设置较大的阈值，对于不重要的人体姿态关键点，可设置较小的阈值。示例性的，如在人体姿态关键点为头顶的情况下，其对应的阈值为0.9，而在人体姿态关键点为左膝的情况下，其对应的阈值为0.5。

在步骤150中，判断当前帧图像数据是否为最后一帧图像数据；在当前帧图像数据不是最后一帧图像数据的情况下，执行步骤160；在当前帧图像数据是

最后一帧图像数据的情况下，执行步骤 170。

在步骤 160 中，将当前帧图像数据的人体姿态置信图输入至人体姿态检测模型中，用于参与生成下一帧图像数据的人体姿态置信图。

在步骤 170 中，结束执行生成多帧图像数据的人体姿态置信图的操作。

在本申请的实施例，判断当前帧图像数据是否为最后一帧图像数据，在当前帧图像数据不为最后一帧图像数据的情况下，可将当前帧图像数据的人体姿态置信图输入至人体姿态检测模型中，作为下一帧图像数据的输出结果的参考，以提高下一帧图像数据的输出结果的精度，即将下一帧图像数据输入至预先训练的人体姿态检测模型中，并参考当前帧图像数据的人体姿态置信图，输出下一帧图像数据的多张人体姿态参考图，在人体姿态参考图中识别人体姿态关键点，根据人体姿态关键点的可信性，生成人体姿态置信图。

需要说明的是，在当前帧图像数据为最后一帧图像数据的情况下，表明可以结束执行生成多帧图像数据的人体姿态置信图的操作，而无需再将得到的人体姿态置信图输入至人体姿态检测模型中。在此基础上，可以理解到，在当前帧图像数据为最后一帧图像数据的情况下，可以只执行步骤 120、步骤 130 以及判断人体姿态关键点是否可信，在人体姿态关键点不可信的情况下，将上一帧图像数据中对应的人体姿态关键点作为人体姿态关键点。当然可以理解到，每经过步骤 120、步骤 130 以及判断人体姿态关键点是否可信，在人体姿态关键点不可信的情况下，将上一帧图像数据中对应的人体姿态关键点作为人体姿态关键点，便可以得到当前帧图像数据对应的人体姿态关键点。

还需要说明的是，步骤 120-步骤 150，均是针对当前帧图像数据的处理过程，相应的，步骤 120 和步骤 130 中的人体姿态参考图指的是当前帧图像数据对应的人体姿态参考图，步骤 130 和步骤 140 中的人体姿态关键点指的是当前帧图像数据对应的人体姿态关键点，以及步骤 140 和步骤 150 中的人体姿态置信图指的是当前帧图像数据对应的人体姿态置信图。

基于上述，由于当前帧图像数据表示的是当前正在处理的某帧图像数据，因此，在第一帧图像数据为当前正在处理的某帧图像数据的情况下，可以将第一帧图像数据作为当前帧图像数据；在第二帧图像数据为当前正在处理的某帧图像数据的情况下，可以将第二帧图像数据作为当前帧图像数据，依此类推。换句话说，当前帧图像数据可以为第一帧图像数据、第二帧图像数据、第三帧图像数据、……、第 N-1 帧图像数据或第 N 帧图像数据。

假设视频包括 N 帧图像数据， $N \geq 1$ ，在确定当前帧图像数据不是第 N 帧图像数据的情况下，可以重复执行步骤 120-140，进而完成对第一帧图像数据至第 N-1 帧图像数据的处理操作；在确定当前帧图像数据为第 N 帧图像数据的情况下，可以执行步骤 120-步骤 130 以及在人体姿态关键点不可信的情况下，将上

一帧图像数据中对应的人体姿态关键点作为人体姿态关键点即可。

本实施例的技术方案，通过采集多帧图像数据，将当前帧图像数据输入至预先训练的人体姿态检测模型中，并参考上一帧图像数据的人体姿态置信图，输出多张人体姿态参考图，人体姿态检测模型经应用于嵌入式平台的卷积神经网络训练生成，在人体姿态参考图中识别人体姿态关键点，根据人体姿态关键点的可信性，生成人体姿态置信图，判断当前帧图像数据是否为最后一帧图像数据，在当前帧图像数据不是最后一帧图像数据的情况下，将当前帧图像数据的人体姿态置信图输入至人体姿态检测模型中，用于参与生成下一帧图像数据的人体姿态置信图，在当前帧图像数据是最后一帧图像数据的情况下，结束执行生成多帧图像数据的人体姿态置信图的操作，实现了在嵌入式平台上进行人体姿态检测，同时，将上一帧图像数据的输出结果引入对当前帧图像数据的输出结果的预测过程中，提高了预测精度。

在一些实施例中，人体姿态检测模型包括主路、第一支路和第二支路，主路包括残差模块和上采样模块，第一支路包括提炼网络模块，第二支路包括反馈模块。

将当前帧图像数据输入至预先训练的人体姿态检测模型中，并参考上一帧图像数据的人体姿态置信图，输出多张人体姿态参考图，包括：将当前帧图像数据输入至残差模块进行处理，并参考将上一帧图像数据的人体姿态置信图输入至反馈模块进行处理得到的结果进行处理，得到第一卷积结果。将残差模块输出的第一卷积结果输入至上采样模块进行处理，得到第二卷积结果，将残差模块输出的第一卷积结果输入至提炼网络模块进行处理，得到第三卷积结果。将第二卷积结果和第三卷积结果相加，输出多张人体姿态参考图。

在本申请的实施例中，残差模块设置为提取图像数据的边缘和轮廓等特征，而上采样模块设置为提取图像数据的上下文信息。提炼网络模块设置为对残差模块输出的第一卷积结果进行处理，可将第一卷积结果认为是网络中间层信息，即提炼网络模块利用了网络中间层信息，增加了其回传梯度，进而提高了卷积神经网络的预测精度。反馈模块设置为将上一帧图像数据的人体姿态置信图引入卷积神经网络中，提高当前帧图像数据输出结果的精度。

将当前帧图像数据输入至残差模块进行处理，并将上一帧图像数据的人体姿态置信图输入至反馈模块进行处理，得到第一卷积结果，可作如下理解：将所述当前帧图像数据输入至所述残差模块进行处理，并参考将所述上一帧图像数据的人体姿态置信图输入至所述反馈模块进行处理得到的结果进行处理，得到第一卷积结果。

将残差模块输出的第一卷积结果输入至上采样模块进行处理，得到第二卷积结果，将残差模块输出的第一卷积结果输入至提炼网络模块进行处理，得到

第三卷积结果，再将第二卷积结果和第三卷积结果相加，输出多张人体姿态参考图，其中，上采样模块可以采用最近邻插值方法，也可采用其它上采样方法，可以根据实际情况进行设定，在此不作限定。

通过提炼网络模块利用了网络中间层信息，增加了其回传梯度，进而提高了卷积神经网络的预测精度。通过反馈模块将上一帧图像数据的人体姿态置信图引入卷积神经网络中，参与人体姿态检测模型对当前帧图像数据的预测，也提高了卷积神经网络的预测精度。

在一些实施例中，残差模块包括第一残差单元、第二残差单元和第三残差单元。

将当前帧图像数据输入至残差模块进行处理，并参考将上一帧图像数据的人体姿态置信图输入至反馈模块进行处理得到的结果进行处理，得到第一卷积结果，包括：将当前帧图像数据输入至第一残差单元进行处理，得到第一中间结果。将第一中间结果输入至第二残差单元进行处理后的结果以及将上一帧图像数据的人体姿态置信图输入至反馈模块进行处理后的结果相加，得到第二中间结果。将第二中间结果输入至第三残差单元进行处理，得到第三中间结果，作为第一卷积结果。其中，第一中间结果、第二结果和第三结果的通道数依次增多。

在本申请的实施例中，残差模块包括第一残差单元、第二残差单元和第三残差单元，其中，每个残差单元均由 ShuffleNet 子单元和 ShuffleNet 下采样子单元组成，其中，ShuffleNet 子单元可以实现对任意尺寸的图像数据进行操作，其由两个参数控制，分别为输入深度和输出深度，其中，输入深度表示的是输入网络中间特征层的层数，输出深度指的是该子单元所输出中间特征层的层数，层数与通道数对应，ShuffleNet 子单元提取了较高层次的特征，同时保留了原有层次的信息，可以实现不改变图像数据的尺寸大小，只改变网络中间特征层的深度，可以将其看做一个保持尺寸大小不变的高级“卷积层”。其中，在卷积神经网络中，通道数即指每个卷积层中卷积核的个数。此外，需要说明的是，每个残差单元可以只包含一个 ShuffleNet 子单元，相比于原有的每个残差单元包括三个 ShuffleNet 子单元而言，简化了网络结构，相应的，也就减少了计算量，提升了处理效率。

通过第一残差单元、第二残差单元和第三残差单元中 ShuffleNet 下采样子单元的依次处理，使得第一中间结果、第二中间结果和第三中间结果的尺寸依次变小，同时，为了保持网络大小的不变，使第一中间结果的通道数、第二中间结果的通道数和第三中间结果的通道数依次增多。此外，每个通道对应一张特征图。

需要说明的是，中间结果可以用 $W \times H \times K$ 表示，其中， W 表示中间结果

的宽度, H 表示中间结果的长度, K 表示通道数, $W \times H$ 即表示中间结果的尺寸。针对输入图像数据来说, 其可以表示为 $W \times H \times D$, 其中, W 和 H 与前述含义相同, D 表示深度, 示例性的, 在输入图像数据是 RGB 图像的情况下, $D=3$, 在输入图像数据是灰度图像的情况下, $D=1$ 。

示例性的, 如第一中间结果、第二中间结果和第三中间结果用 $W \times H \times K$ 表示, W 、 H 和 K 的含义与前述相同, 第一中间结果为 $64 \times 32 \times 32$, 第二中间结果为 $32 \times 16 \times 64$, 第三中间结果为 $16 \times 8 \times 128$ 。基于上述可知, 第一中间结果的尺寸为 64×32 , 第二中间结果的尺寸为 32×16 , 第三中间结果的尺寸为 16×8 , 上述表明, 第一中间结果、第二中间结果和第三中间结果的尺寸依次变小。同时, 第一中间结果的通道数为 32, 第二中间结果的通道数为 64, 第三中间结果的通道数为 128, 上述表明, 第一中间结果、第二中间结果和第三中间结果的通道数依次增多。

在一些实施例中, 人体姿态检测模型包括第三支路。

将残差模块输出的第一卷积结果输入至上采样模块进行处理, 得到第二卷积结果, 将残差模块输出的第一卷积结果输入至提炼网络模块进行处理, 得到第三卷积结果, 包括: 将第一中间结果输入至第三支路进行处理, 得到第四中间结果。将第二中间结果输入至第三支路进行处理, 得到第五中间结果。将第三中间结果和第五中间结果输入至上采样模块进行处理, 得到第六中间结果。将第四中间结果和第六中间结果输入至上采样模块进行处理, 得到第七中间结果, 作为第二卷积结果。将残差模块输出的第一卷积结果输入至提炼网络模块进行处理, 得到第三卷积结果。其中, 第六中间结果和第七中间结果的通道数依次减少。

在本申请的一些实施例中, 人体姿态检测模型包括第三支路, 第三支路所起到的作用在于: 通过将第三支路, 实现了将跳转连接的卷积操作移至主路上, 从而提高了人体姿态检测模型的预测精度。第三支路包括 1×1 卷积核模块、批标准化模块和线性激活函数模块。其中, 1×1 卷积核可以起到如下作用:

情况一、针对单通道和单个卷积核来说, 1×1 的卷积核是对输入图像数据进行的比例缩放, 这是由于 1×1 卷积核只有一个参数, 这个卷积核在输入图像数据上滑动, 就相当于给输入图像数据乘以一个系数;

情况二、针对多通道和多个卷积核来说, 1×1 卷积核具有如下两方面的作用: 其一, 实现跨通道的交互和信息整合; 其二、进行降维和升维并减少网络参数, 这里所述的降维指的是减少通道数, 升维指的是增加通道数; 其三、在不损失分辨率的前提下大幅增加非线性特性。

批标准化模块设置为进行批标准化处理, 其中, 批标准化 (或称批量归一化) 是为了避免神经网络层数加深, 收敛速度变慢, 导致的梯度消失或者梯度

爆炸，可以通过采用批标准化来规范某些层或者所有层的输入，从而固定每层输入信号的均值与方差，使得每一层的输入有一个稳定的分布。示例性的，其一般用在激活函数之前，对 $x=W+b$ 进行规范化，使输出结果的均值为0，方差为1，其中， W 表示权值矩阵， b 表示偏置。可以理解到，在卷积神经网络中，权值矩阵即指卷积核，即 W 表示卷积核。

由于第七中间结果是将第六中间结果和第四中间结果输入至上采样模块后得到的，因此，第七中间结果的尺寸大于第六中间结果的尺寸，同时，为了保持网络大小不变，使第六中间结果的通道数和第七中间结果的通道数依次减少。

通过将第三支路，实现了将跳转连接的卷积操作移至主路上，从而提高了人体姿态检测模型的预测精度。此外，可将第一中间结果、第二中间结果和第三中间结果理解编码部分，将第六中间结果和第七中间结果理解为解码部分，为了保持网络大小不变，在编码部分，随着中间结果的尺寸减小，依次增加中间结果的通道数；在解码部分，随着中间结果的尺寸增大，依次减少中间结果的通道数。此外，可以理解到，本申请实施例所提供的卷积神经网络是一种非对称编码-解码结构。

在一些实施例中，将所述第二卷积结果和所述第三卷积结果相加，输出多张人体姿态参考图之后，还包括：将第一卷积结果和第二卷积结果相加，得到目标结果。将多张人体姿态参考图和目标结果相加，输出新的多张人体姿态参考图。其中，目标结果用于在对人体姿态检测模型进行训练时，提高人体姿态检测模型的精度。

在本申请的一些实施例中，为了提高人体姿态检测模型在训练阶段的精度，可以考虑增加中途监督，中途监督指的是在每个阶段的输出都计算损失，可以保证底层参数正常更新。

将第一卷积结果和第二卷积结果相加，得到目标结果，再将目标结果与多张人体姿态参考图相加，得到新的多张人体姿态参考图，上述目标结果即起到中途监督的作用，即目标结果也参与到损失的计算过程中。

需要说明的是，在预测阶段，可以不执行将第一卷积结果和第二卷积结果相加的操作，即在预测阶段，输出结果只包括多张人体姿态参考图。

还需要说明的是，本申请实施例所述的技术方案，在采集到多帧图像数据后，无需进行检测图像数据中是否有人脸，在存在人脸的情况下，检测人脸在图像数据中所在的位置，再将其提取等操作，不进行上述操作的原因在于：上述操作耗时较长，且检测结果误差较大。可以理解到，在不进行上述操作的情况下，可以大大提高数据处理效率。

另需要说明的是，由于第二残差单元和第三残差单元均由 ShuffleNet 子单元和 ShuffleNet 下采样子单元组成，在每次进行下采样之前，主路上保留原尺寸信

息,即在第二残差单元的 ShuffleNet 下采样子单元在进行下采样之前,将第一中间结果输入至第二残差单元;在第三残差单元的 ShuffleNet 下采样子单元在进行下采样之前,将第二中间结果输入至第三残差单元。两次下采样之间,使用一个 ShuffleNet 子单元提取特征,即第一残差单元和第二残差单元之间使用一个 ShuffleNet 子单元提取特征,该 ShuffleNet 子单元为第一残差单元的 ShuffleNet 子单元;第二残差单元和第三残差单元之间使用一个 ShuffleNet 子单元提取特征,即第二残差单元和第三残差单元之间使用一个 ShuffleNet 子单元提取特征,该 ShuffleNet 子单元为第二残差单元的 ShuffleNet 子单元。

本申请实施例所提供的卷积神经网络引入了提炼网络模块、反馈模块以及将跳转连接的卷积操作移至主路上,上述提高了卷积神经网络的预测精度。此外,采用非对称编码-解码结构,保证了网络大小基本不变,由于每个残差单元只包含一个 ShuffleNet 子单元,相比于原有的每个残差单元包括三个 ShuffleNet 子单元而言,简化了网络结构,相应的,也就减少了计算量,提升了处理效率。基于上述,使得基于卷积神经网络的人体姿态检测方法可以应用于嵌入式平台,如智能手机的嵌入式平台上,并且实时运行且预测精度可以满足要求。

为了更好的理解本申请实施例所提供的卷积神经网络,下面以具体示例进行说明:

如图 2 所示,为一种卷积神经网络的应用示意图,该卷积神经网络可以包括:主路、第一支路、第二支路和第三支路。其中,主路包括第一卷积模块 21、第一残差单元 22、第二残差单元 23、第三残差单元 24、第二卷积模块 25、上采样模块 26、按位加模块 27 和第三卷积模块 28。

其中,第一残差单元 22、第二残差单元 23 和第三残差单元 24 均包括 ShuffleNet 下采样子单元 221 和 ShuffleNet 子单元 222。第一支路包括提炼网络模块 29,其中,提炼网络模块 29 包括 ShuffleNet 子单元 222、上采样模块 26 及按位加模块 27;第二支路包括反馈模块 30;第三支路包括第二卷积模块 25。

需要说明的是,模块、单元或子单元上标注的 $W \times H \times K$,表示经过该模块、单元或子单元处理后得到的结果,其中, W 表示结果的宽度, H 表示结果的长度, K 表示通道数。

还需要说明的是,第一卷积模块 21 包括如下处理操作:第一步、卷积操作,所采用的卷积核的尺寸为 3×3 ;第二步、批标准化;第三步、线性激活函数。第二卷积模块 25 包括如下处理操作:第一步、卷积操作,采用所卷积核的尺寸为 1×1 ;第二步、批标准化;第三步、线性激活函数。第三卷积模块 26 包括如下处理操作:第一步、卷积操作,所采用的卷积核的尺寸为 1×1 ;第二步、批标准化;第三步、线性激活函数;第四步、卷积操作,所采用采用的卷积核的尺寸为 3×3 。

假设当前帧图像数据为 $256 \times 128 \times 3$ 的 RGB 图像, 将其作为输入变量输入至卷积神经网络中, 依次经过第一卷积模块 21 和第一残差单元 22 后, 得到第一中间结果, 第一中间结果为 $64 \times 32 \times 32$, 将第一中间结果和将上一帧图像数据的人体姿态置信图输入至反馈模块 30 进行处理后的结果, 共同输入至主路上的按位加模块 27 进行处理, 将主路上的按位加模块 27 处理后的结果输入至第二残差单元 23 进行处理, 得到第二中间结果, 第二中间结果为 $32 \times 16 \times 64$, 将第二中间结果输入至第三残差单元 24 进行处理, 得到第三中间结果, 将第三中间结果作为第一卷积结果, 第一卷积结果为 $16 \times 8 \times 128$ 。需要说明的是, 反馈模块 30 可以包括 1×1 卷积核, 设置为升维, 这是由于上一帧图像数据的人体姿态置信图为 $64 \times 32 \times 14$, 而第一中间结果为 $64 \times 32 \times 32$, 需要升维, 以保证两者输出通道数一致。

将第一中间结果输入至第三支路的第二卷积模块 25 进行处理, 得到第四中间结果, 第四中间结果为 $64 \times 32 \times 32$ 。

将第二中间结果输入至第三支路的第二卷积模块 25 进行处理, 得到第五中间结果, 第五中间结果为 $32 \times 16 \times 32$ 。

将第三中间结果输入至主路上的第二卷积模块 25 和上采样模块 26 进行处理后得到的结果和第五中间结果, 共同输入至主路上的按位加模块 27 进行处理, 得到第六中间结果, 将第六中间结果输入至主路上的上采样模块 26 进行处理得到的结果和第四中间结果, 共同输入至主路上的按位加模块 27 进行处理, 得到第七中间结果, 将第七中间结果作为第二卷积结果, 第二卷积结果为 $64 \times 32 \times 32$ 。

将第三中间结果输入至主路上的第二卷积模块 25 进行处理得到的结果, 再输入至第一支路上的 ShuffleNet 子单元 222 进行处理, 得到第八中间结果, 将第八中间结果输入至第一支路上的上采样模块 26 进行处理, 得到第九中间结果, 再将第九中间结果输入至第一支路上的 ShuffleNet 子单元 222 进行处理, 得到第十中间结果, 将第十中间结果输入至第一支路上的上采样模块 26 进行处理, 得到第十一中间结果。将第六中间结果输入至第一支路上的 ShuffleNet 子单元 222 进行处理, 得到第十二中间结果, 将第十二中间结果输入至第一支路上的上采样模块 26 进行处理, 得到第十三中间结果, 将第十一中间结果和第十三中间结果共同输入至第一支路上的按位加模块 27 进行处理, 得到第三卷积结果, 第三卷积结果为 $64 \times 32 \times 32$ 。

将第二卷积结果和第三卷积结果输入至主路上的按位加模块 27, 得到第十四中间结果, 将第十四中间结果输入至主路上的 ShuffleNet 子单元 222, 得到第十五中间结果, 第十五中间结果为 $64 \times 32 \times 32$, 将第十五中间结果输入至主路上的第三卷积模块 28, 输出多张人体姿态参考图。

将第一卷积结果和第二卷积结果相加，得到目标结果，目标结果为 $64 \times 32 \times 14$ 。将多张人体姿态参考图和目标结果相加，输出新的多张人体姿态参考图。其中，目标结果用于在对人体姿态检测模型进行训练时，提高人体姿态检测模型的精度。

需要说明的是，由于上一帧图像数据的人体姿态置信图并未在开始时，与当前帧图像数据作为输入变量输入至卷积神经网络中，而是在网络中间层与第一中间结果作为输入变量输入至卷积神经网络中，上述实现了减少数据处理量。

图 3 为本申请实施例提供的另一种人体姿态检测方法的流程图，本实施例可适用于检测人体姿态的情况，该方法可以由人体姿态检测装置来执行，该装置可以采用软件和硬件中至少之一的方式实现，该装置可以配置于设备中，例如典型的是计算机或移动终端等。如图 3 所示，该方法包括步骤 301 至步骤 311。

在步骤 301 中，采集多帧图像数据。

在步骤 302 中，判断上一帧图像数据的人体姿态置信图是否可信；在上一帧图像数据的人体姿态置信图可信的情况下，执行步骤 303；在上一帧图像数据的人体姿态置信图不可信的情况下，执行步骤 304。

在步骤 303 中，将当前帧图像数据和上一帧图像数据的人体姿态置信图输入至预先训练的人体姿态检测模型中，输出多张人体姿态参考图，并转入执行步骤 305。

在步骤 304 中，将当前帧图像数据和预设图像数据输入至预先训练的人体姿态检测模型中，输出多张人体姿态参考图，并转入执行步骤 305。

在步骤 305 中，每张人体姿态参考图包括人体姿态关键点的多个候选点，每个候选点的坐标位置对应一个概率值；在人体姿态参考图中确定多个候选点的坐标位置对应的多个概率值中最大概率值对应的坐标位置，将该坐标位置对应的候选点作为人体姿态关键点。

在步骤 306 中，判断人体姿态关键点对应的概率值是否大于预设的阈值；在人体姿态关键点对应的概率值大于预设的阈值的情况下，执行步骤 307；在人体姿态关键点对应的概率值小于等于预设的阈值的情况下，执行步骤 308。

在步骤 307 中，以人体姿态关键点为中心生成掩模图，作为当前帧图像数据的人体姿态置信图，并转入执行步骤 309。

在步骤 308 中，将预设图像数据作为当前帧图像数据的人体姿态置信图，并转入执行步骤 309。

在步骤 309 中，判断当前帧图像数据是否为最后一帧图像数据；在当前帧图像数据不是最后一帧图像数据的情况下，执行步骤 310；在当前帧图像数据是最后一帧图像数据的情况下，执行步骤 311。

在步骤 310 中，将当前帧图像数据的人体姿态置信图输入至人体姿态检测

模型中，用于参与生成下一帧图像数据的人体姿态置信图。

在步骤 311 中，结束执行生成多帧图像数据的人体姿态置信图的操作。

在本申请的实施例中，需要说明的是，本申请实施例所提供的人体姿态检测模型经应用于嵌入式平台的卷积神经网络训练生成。

本实施例的技术方案，通过采集多帧图像数据，将当前帧图像数据输入至预先训练的人体姿态检测模型中，并参考上一帧图像数据的人体姿态置信图，输出多张人体姿态参考图，人体姿态检测模型经应用于嵌入式平台的卷积神经网络训练生成，在人体姿态参考图中识别人体姿态关键点，根据人体姿态关键点的可信性，生成人体姿态置信图，判断当前帧图像数据是否为最后一帧图像数据，在当前帧图像数据不是最后一帧图像数据的情况下，将当前帧图像数据的人体姿态置信图输入至人体姿态检测模型中，用于参与生成下一帧图像数据的人体姿态置信图，在当前帧图像数据是最后一帧图像数据的情况下，结束执行生成多帧图像数据的人体姿态置信图的操作，实现了在嵌入式平台上进行人体姿态检测，同时，将上一帧图像数据的输出结果引入对当前帧图像数据的输出结果的预测过程中，提高了预测精度。

图 4 为本申请实施例提供的一种人体姿态检测装置的结构示意图，本实施例可适用于检测人体姿态的情况，该装置可以采用软件和硬件中至少之一的方式实现，该装置可以配置于设备中，例如典型的是计算机或移动终端等。如图 4 所示，该装置包括图像数据采集模块 410、人体姿态参考图输出模块 420、人体姿态关键点识别模块 430、人体姿态置信图生成模块 440、判断模块 450、第一执行模块 460 以及第二执行模块 470。

图像数据采集模块 410，设置为采集多帧图像数据。

人体姿态参考图输出模块 420，设置为将当前帧图像数据输入至预先训练的人体姿态检测模型中，并参考上一帧图像数据的人体姿态置信图，输出多张人体姿态参考图，人体姿态检测模型经应用于嵌入式平台的卷积神经网络训练生成。

人体姿态关键点识别模块 430，设置为在每张人体姿态参考图中识别人体姿态关键点。

人体姿态置信图生成模块 440，设置为根据人体姿态关键点的可信性，生成当前帧图像数据的人体姿态置信图。

判断模块 450，设置为判断当前帧图像数据是否为最后一帧图像数据。

第一执行模块 460，设置为在当前帧图像数据不是最后一帧图像数据的情况下，将当前帧图像数据的人体姿态置信图输入至人体姿态检测模型中，用于参与生成下一帧图像数据的人体姿态置信图。

第二执行模块 470，设置为在当前帧图像数据是最后一帧图像数据的情况下，

结束执行生成多帧图像数据的人体姿态置信图的操作。

本实施例的技术方案，通过采集多帧图像数据，将当前帧图像数据输入至预先训练的人体姿态检测模型中，并参考上一帧图像数据的人体姿态置信图，输出多张人体姿态参考图，人体姿态检测模型经应用于嵌入式平台的卷积神经网络训练生成，在每张人体姿态参考图中识别人体姿态关键点，根据人体姿态关键点的可信性，生成当前帧图像数据的人体姿态置信图，判断当前帧图像数据是否为最后一帧图像数据，在当前帧图像数据不是最后一帧图像数据的情况下，将当前帧图像数据的人体姿态置信图输入至人体姿态检测模型中，用于参与生成下一帧图像数据的人体姿态置信图，在当前帧图像数据是最后一帧图像数据的情况下，结束执行生成多帧图像数据的人体姿态置信图的操作，实现了在嵌入式平台上进行人体姿态检测，同时，将上一帧图像数据的输出结果引入对当前帧图像数据的输出结果的预测过程中，提高了预测精度。

在一些实施例中，人体姿态参考图输出模块 420，包括置信图可信性判断单元、第一人体姿态参考图输出单元以及第二人体姿态参考图输出单元。

置信图可信性判断单元，设置为判断上一帧图像数据的人体姿态置信图是否可信。

第一人体姿态参考图输出单元，设置为在上一帧图像数据的人体姿态置信图可信的情况下，将当前帧图像数据和上一帧图像数据的人体姿态置信图输入至预先训练的人体姿态检测模型中，输出多张人体姿态参考图。

第二人体姿态参考图输出单元，设置为在上一帧图像数据的人体姿态置信图不可信的情况下，将当前帧图像数据和预设图像数据输入至预先训练的人体姿态检测模型中，输出多张人体姿态参考图。

在一些实施例中，每张人体姿态参考图包括人体姿态关键点多个候选点，每个候选点的坐标位置对应一个概率值，人体姿态关键点识别模块 430，包括人体姿态关键点识别单元。

人体姿态关键点识别单元，设置为在人体姿态参考图中确定多个候选点的坐标位置对应的多个概率值中最大概率值对应的的坐标位置，将该坐标位置对应的候选点作为人体姿态关键点。

在一些实施例中，人体姿态置信图生成模块 440，包括人体姿态关键点可信性判断单元、第一人体姿态置信图生成单元以及第二人体姿态置信图生成单元。

人体姿态关键点可信性判断单元，设置为判断人体姿态关键点是否可信。

第一人体姿态置信图生成单元，设置为在人体姿态关键点可信的情况下，以人体姿态关键点作为中心生成掩模图，作为人体姿态置信图。

第二人体姿态置信图生成单元，设置为在人体姿态关键点不可信的情况下，将预设图像数据作为人体姿态置信图。

在一些实施例中，人体姿态关键点可信性判断单元，设置为：

判断人体关键点对应的概率值是否大于预设的阈值。

在人体关键点对应的概率值大于预设的阈值的情况下，确定人体姿态关键点可信。

在人体关键点对应的概率值小于等于预设的阈值的情况下，确定人体姿态关键点不可信。

在一些实施例中，人体姿态检测模型包括主路、第一支路和第二支路，主路包括残差模块和上采样模块，第一支路包括提炼网络模块，第二支路包括反馈模块。

将当前帧图像数据输入至预先训练的人体姿态检测模型中，并参考上一帧图像数据的人体姿态置信图，输出多张人体姿态参考图，包括：

将当前帧图像数据输入至残差模块进行处理，并参考将上一帧图像数据的人体姿态置信图输入至反馈模块进行处理得到的结果进行处理，得到第一卷积结果。

将残差模块输出的第一卷积结果输入至上采样模块进行处理，得到第二卷积结果，将残差模块输出的第一卷积结果输入至提炼网络模块进行处理，得到第三卷积结果。

将第二卷积结果和第三卷积结果相加，输出多张人体姿态参考图。

在一些实施例中，残差模块包括第一残差单元、第二残差单元和第三残差单元。

将当前帧图像数据输入至残差模块进行处理，并将上一帧图像数据的人体姿态置信图输入至反馈模块进行处理得到的结果进行处理，得到第一卷积结果，包括：

将当前帧图像数据输入至第一残差单元进行处理，得到第一中间结果。

将第一中间结果输入至第二残差单元进行处理以及将上一帧图像数据的人体姿态置信图输入至反馈模块进行处理后的结果相加，得到第二中间结果。

将第二中间结果输入至第三残差单元进行处理，得到第三中间结果，作为第一卷积结果。

其中，第一中间结果、第二中间结果和第三中间结果的通道数依次增多。

在一些实施例中，人体姿态检测模型包括第三支路。

将残差模块输出的第一卷积结果输入至上采样模块进行处理，得到第二卷积结果，将残差模块输出的第一卷积结果输入至提炼网络模块进行处理，得到第三卷积结果，包括：

将第一中间结果输入至第三支路进行处理，得到第四中间结果。

将第二中间结果输入至第三支路进行处理，得到第五中间结果。

将第三中间结果和第五中间结果输入至上采样模块进行处理，得到第六中间结果。

将第四中间结果和第六中间结果输入至上采样模块进行处理，得到第七中间结果，作为第二卷积结果。

将残差模块输出的第一卷积结果输入至提炼网络模块进行处理，得到第三卷积结果。

其中，第六中间结果和第七中间结果的通道数依次减少。

在一些实施例中，将所述第二卷积结果和所述第三卷积结果相加，输出多张人体姿态参考图之后，还包括：

将第一卷积结果和第二卷积结果相加，得到目标结果。

将多张人体姿态参考图和目标结果相加，输出新的多张人体姿态参考图。

其中，目标结果用于在对人体姿态检测模型进行训练时，提高人体姿态检测模型的精度。

本申请实施例所提供的人体姿态检测装置可执行本申请任意实施例所提供的人体姿态检测方法。

图 5 为本申请实施例提供的一种设备的结构示意图。图 5 示出了适于用来实现本申请实施方式的示例性设备 512 的框图。如图 5 所示，设备 512 的组件可以包括但不限于：至少一个处理器 516，系统存储器 528，连接于不同系统组件（包括系统存储器 528 和处理器 516）的总线 518。系统存储器 528 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, RAM) 530 和高速缓存存储器 532 中至少之一。设备 512 可以包括其它可移动/不可移动的、易失性/非易失性计算机系统存储介质。仅作为举例，存储系统 534 可以提供硬盘驱动器、磁盘驱动器，以及光盘驱动器。在这些情况下，每个驱动器可以通过至少一个数据介质接口与总线 518 相连。具有一组（至少一个）程序模块 542 的程序/实用工具 540，可以存储在例如存储器 528 中，程序模块 542 通常执行本申请所描述的实施例中的功能和/或方法。设备 512 也可以与至少一个外部设备 514（例如键盘、指向设备、显示器 524 等）通信，还可与至少一个使得用户能与该设备 512 交互的设备通信，和/或与使得该设备 512 能与至少一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（Input/Output, I/O）接口 522 进行。并且，设备 512 还可以通过网络适配器 520 与至少一个或网络（例如局域网(Local Area Network, LAN)，广域网(Wide Area Network, WAN)和/或公共网络，例如因特网)通信。处理器 516 通过运行存储在系统存储器 528 中的程序，从而执行各种功能应用以及数据处理，例如实现本申请任意实施例所提供的人体姿态检测方法。

本申请实施例还提供了一种计算机可读存储介质，其上存储有计算机程序，该计算机程序被处理器执行时实现如本申请任意实施例所提供的一种人体姿态检测方法。

1、一种人体姿态检测方法，包括：

采集多帧图像数据；

将当前帧图像数据输入至预先训练的人体姿态检测模型中，并参考上一帧图像数据的人体姿态置信图，输出多张人体姿态参考图，所述人体姿态检测模型经应用于嵌入式平台的卷积神经网络训练生成；

在每张人体姿态参考图中识别人体姿态关键点；

根据所述人体姿态关键点的可信性，生成当前帧图像数据的人体姿态置信图；

判断当前帧图像数据是否为最后一帧图像数据；

在当前帧图像数据不是最后一帧图像数据的情况下，将所述当前帧图像数据的人体姿态置信图输入至所述人体姿态检测模型中，用于参与生成下一帧图像数据的人体姿态置信图；

在当前帧图像数据是最后一帧图像数据的情况下，结束执行生成多帧图像数据的人体姿态置信图的操作。

2、根据权利要求1所述的方法，其中，所述将当前帧图像数据输入至预先训练的人体姿态检测模型中，并参考上一帧图像数据的人体姿态置信图，输出多张人体姿态参考图，包括：

判断上一帧图像数据的人体姿态置信图是否可信；

在上一帧图像数据的人体姿态置信图可信的情况下，将所述当前帧图像数据和所述上一帧图像数据的人体姿态置信图输入至预先训练的人体姿态检测模型中，输出多张人体姿态参考图；

在上一帧图像数据的人体姿态置信图不可信的情况下，将所述当前帧图像数据和预设图像数据输入至预先训练的人体姿态检测模型中，输出多张人体姿态参考图。

3、根据权利要求1所述的方法，其中，所述每张人体姿态参考图包括人体姿态关键点的多个候选点，每个候选点的坐标位置对应一个概率值；

所述在每张人体姿态参考图中识别人体姿态关键点，包括：

在所述每张人体姿态参考图中确定所述多个候选点的坐标位置对应的多个概率值中最大概率值对应的坐标位置，将所述坐标位置对应的候选点作为人体姿态关键点。

4、根据权利要求2所述的方法，其中，所述根据所述人体姿态关键点的可信性，生成当前帧图像数据的人体姿态置信图，包括：

判断所述人体姿态关键点是否可信；

在所述人体姿态关键点可信的情况下，以所述人体姿态关键点作为中心生成掩模图，作为人体姿态置信图；

在所述人体姿态关键点不可信的情况下，则将所述预设图像数据作为人体姿态置信图。

5、根据权利要求 4 所述的方法，其中，所述判断所述人体姿态关键点是否可信，包括：

判断所述人体关键点对应的概率值是否大于预设的阈值；

在所述人体关键点对应的概率值大于预设的阈值的情况下，确定所述人体姿态关键点可信；

在所述人体关键点对应的概率值小于或等于预设的阈值的情况下，确定所述人体姿态关键点不可信。

6、根据权利要求 1-5 任一所述的方法，其中，所述人体姿态检测模型包括主路、第一支路和第二支路，所述主路包括残差模块和上采样模块，所述第一支路包括提炼网络模块，所述第二支路包括反馈模块；

所述将当前帧图像数据输入至预先训练的人体姿态检测模型中，并参考上一帧图像数据的人体姿态置信图，输出多张人体姿态参考图，包括：

将当前帧图像数据输入至所述残差模块进行处理，并参考将上一帧图像数据的人体姿态置信图输入至所述反馈模块进行处理得到的结果进行处理，得到第一卷积结果；

将所述残差模块输出的第一卷积结果输入至所述上采样模块进行处理，得到第二卷积结果；将所述残差模块输出的第一卷积结果输入至所述提炼网络模块进行处理，得到第三卷积结果；

将所述第二卷积结果和所述第三卷积结果相加，输出多张人体姿态参考图。

7、根据权利要求 6 所述的方法，其中，所述残差模块包括第一残差单元、第二残差单元和第三残差单元；

所述将当前帧图像数据输入至所述残差模块进行处理，并参考将上一帧图像数据的人体姿态置信图输入至所述反馈模块进行处理的结果进行处理，得到第一卷积结果，包括：

将所述当前帧图像数据输入至所述第一残差单元进行处理，得到第一中间结果；

将所述第一中间结果输入至所述第二残差单元进行处理，并将所述上一帧图像数据的人体姿态置信图输入至所述反馈模块进行处理，将所述第二残差单元输出的结果和所述反馈模块输出的结果相加，得到第二中间结果；

将所述第二中间结果输入至所述第三残差单元进行处理，得到第三中间结果，作为所述第一卷积结果；

其中，所述第一中间结果、所述第二中间结果和所述第三中间结果的通道数依次增多。

8、根据权利要求 7 所述的方法，所述人体姿态检测模型还包括第三支路；

所述将所述残差模块输出的第一卷积结果输入至所述上采样模块进行处理，得到第二卷积结果；将所述残差模块输出的第一卷积结果输入至所述提炼网络模块进行处理，得到第三卷积结果，包括：

将所述第一中间结果输入至所述第三支路进行处理,得到第四中间结果;
将所述第二中间结果输入至所述第三支路进行处理,得到第五中间结果;
将所述第三中间结果和所述第五中间结果输入至所述上采样模块进行处理,得到第六中间结果;

将所述第四中间结果和所述第六中间结果输入至所述上采样模块进行处理,得到第七中间结果,作为所述第二卷积结果;

将所述残差模块输出的第一卷积结果输入至所述提炼网络模块进行处理,得到所述第三卷积结果;

其中,所述第六中间结果和所述第七中间结果的通道数依次减少。

9、根据权利要求6所述的方法,所述将所述第二卷积结果和所述第三卷积结果相加,输出多张人体姿态参考图之后,还包括:

将所述第一卷积结果和所述第二卷积结果相加,得到目标结果;

将所述多张人体姿态参考图和所述目标结果相加,输出新的所述多张人体姿态参考图。

10、一种人体姿态检测装置,包括:

图像数据采集模块,设置为采集多帧图像数据;

人体姿态参考图输出模块,设置为将当前帧图像数据输入至预先训练的人体姿态检测模型中,并参考上一帧图像数据的人体姿态置信图,输出多张人体姿态参考图,所述人体姿态检测模型经应用于嵌入式平台的卷积神经网络训练生成;

人体姿态关键点识别模块,设置为在每张人体姿态参考图中识别人体姿态关键点;

人体姿态置信图生成模块,设置为根据所述人体姿态关键点的可信性,生成当前帧图像数据的人体姿态置信图;

判断模块,设置为判断当前帧图像数据是否为最后一帧图像数据;

第一执行模块,设置为在当前帧图像数据不是最后一帧图像数据的情况下,将所述当前帧图像数据的人体姿态置信图输入至所述人体姿态检测模型中,用于参与生成下一帧图像数据的人体姿态置信图;

第二执行模块,设置为在当前帧图像数据是最后一帧图像数据的情况下,结束执行生成多帧图像数据的人体姿态置信图的操作。

11、一种设备,包括:

至少一个处理器;

存储器,设置为存储至少一个程序;

当所述至少一个程序被所述至少一个处理器执行,使得所述至少一个处理器实现如权利要求1-9任一项所述的方法。

12、一种计算机可读存储介质,其上存储有计算机程序,所述计算机程序被处理器执行时实现如权利要求1-9任一项所述的方法。

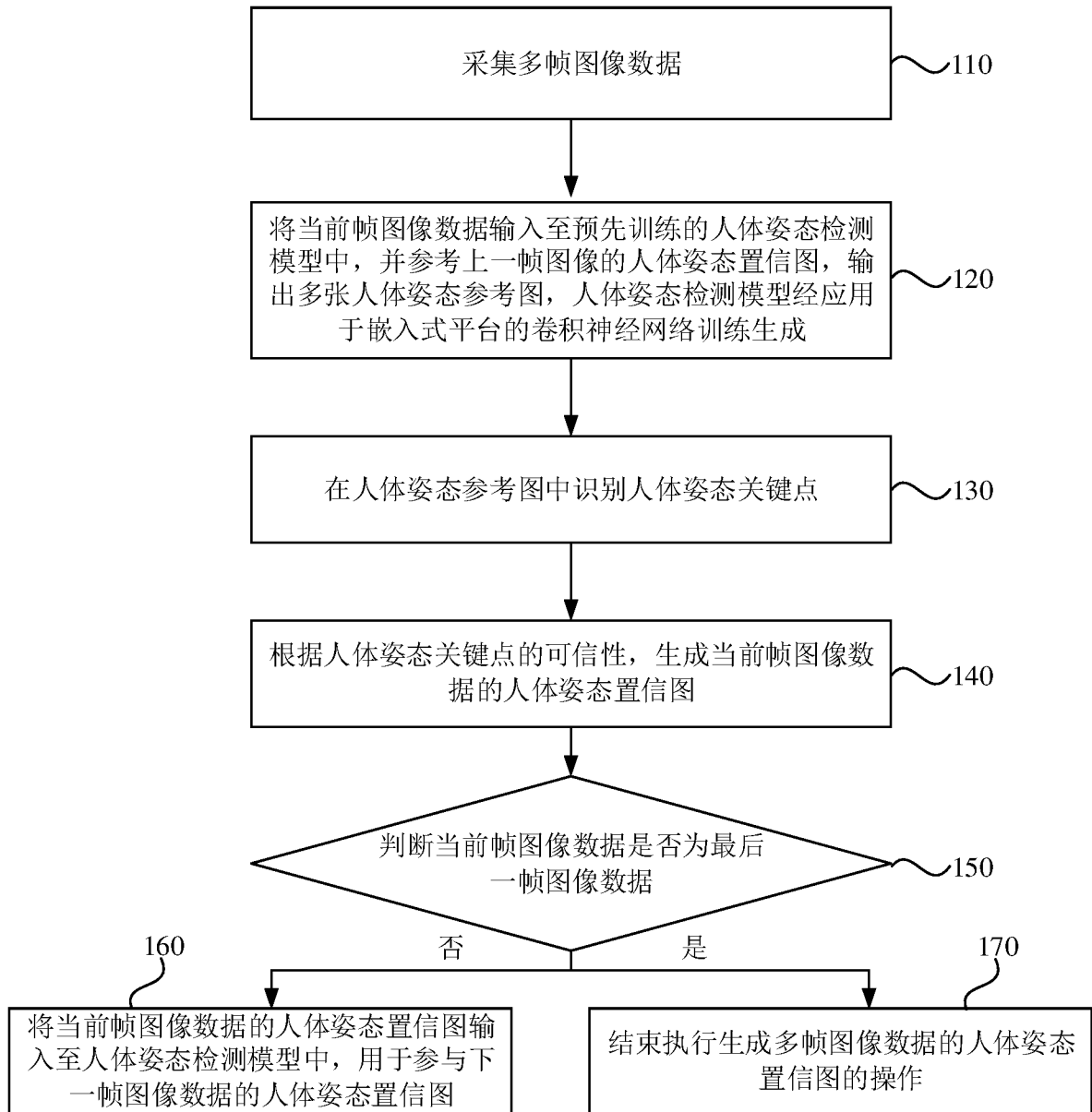


图 1

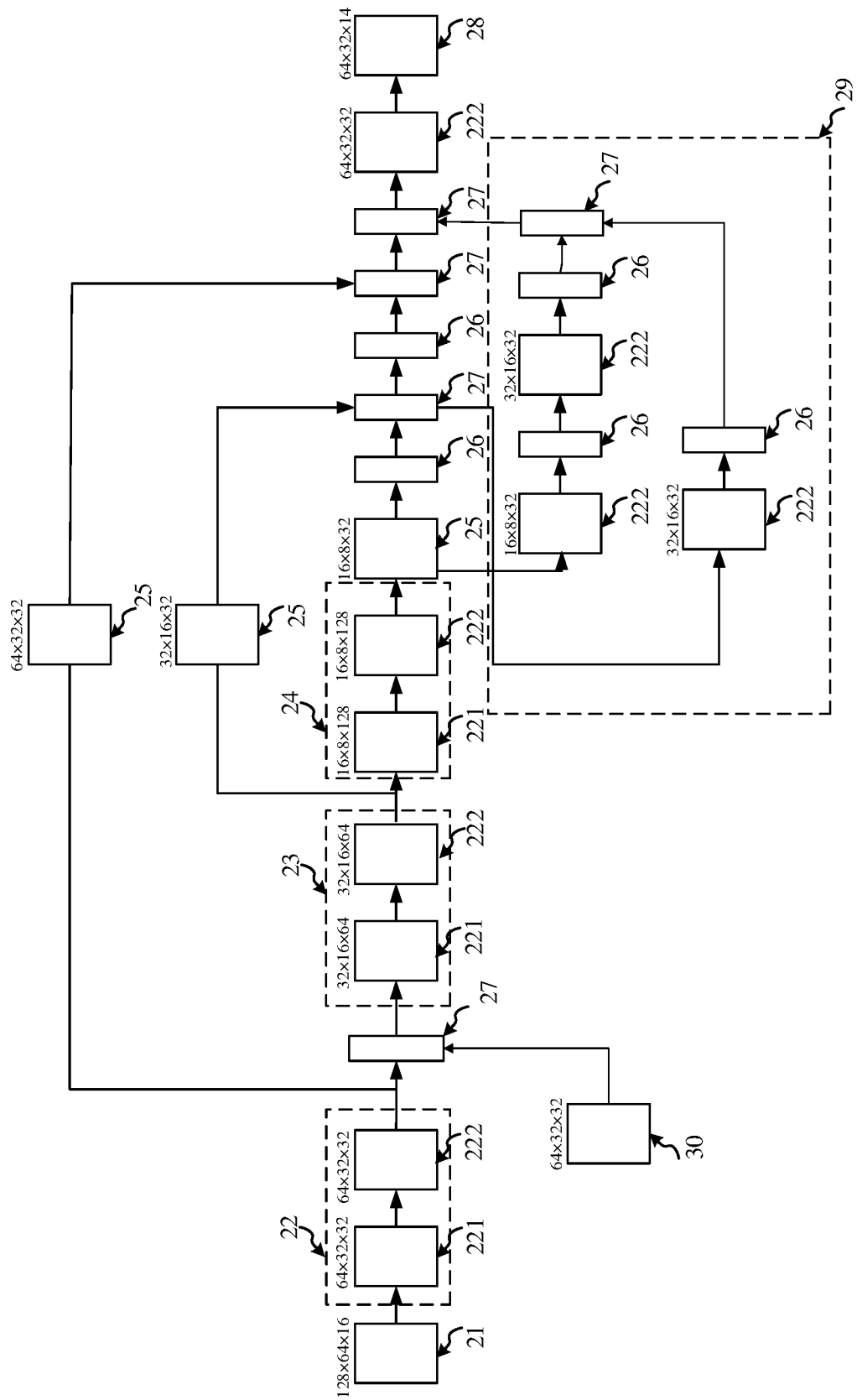


图 2

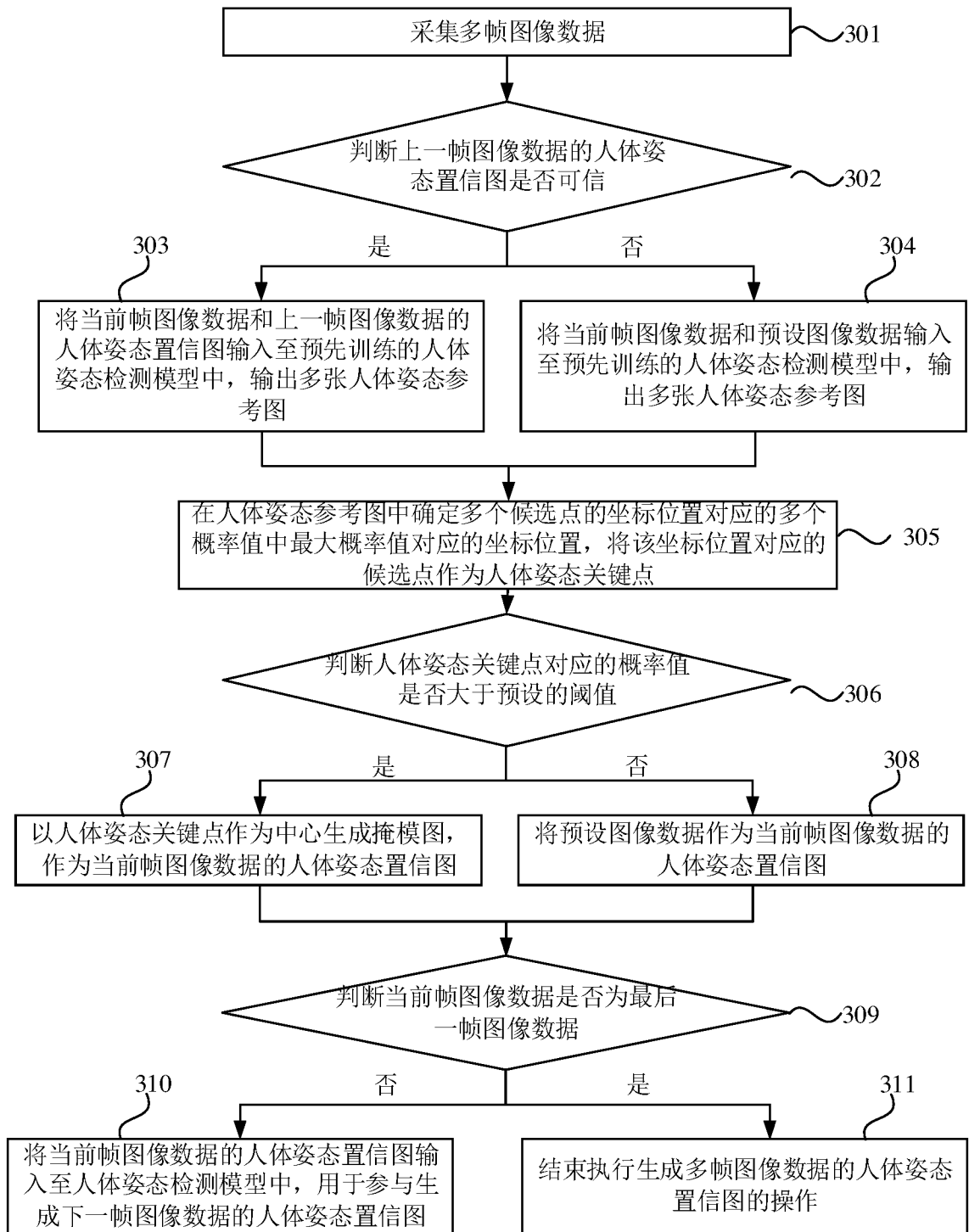


图 3

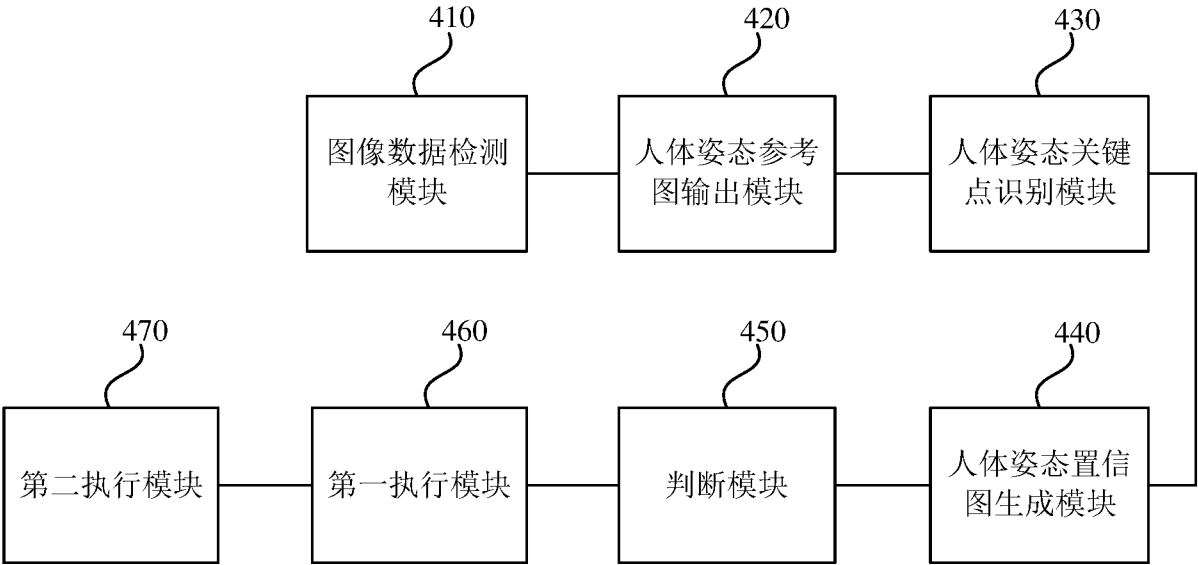


图 4

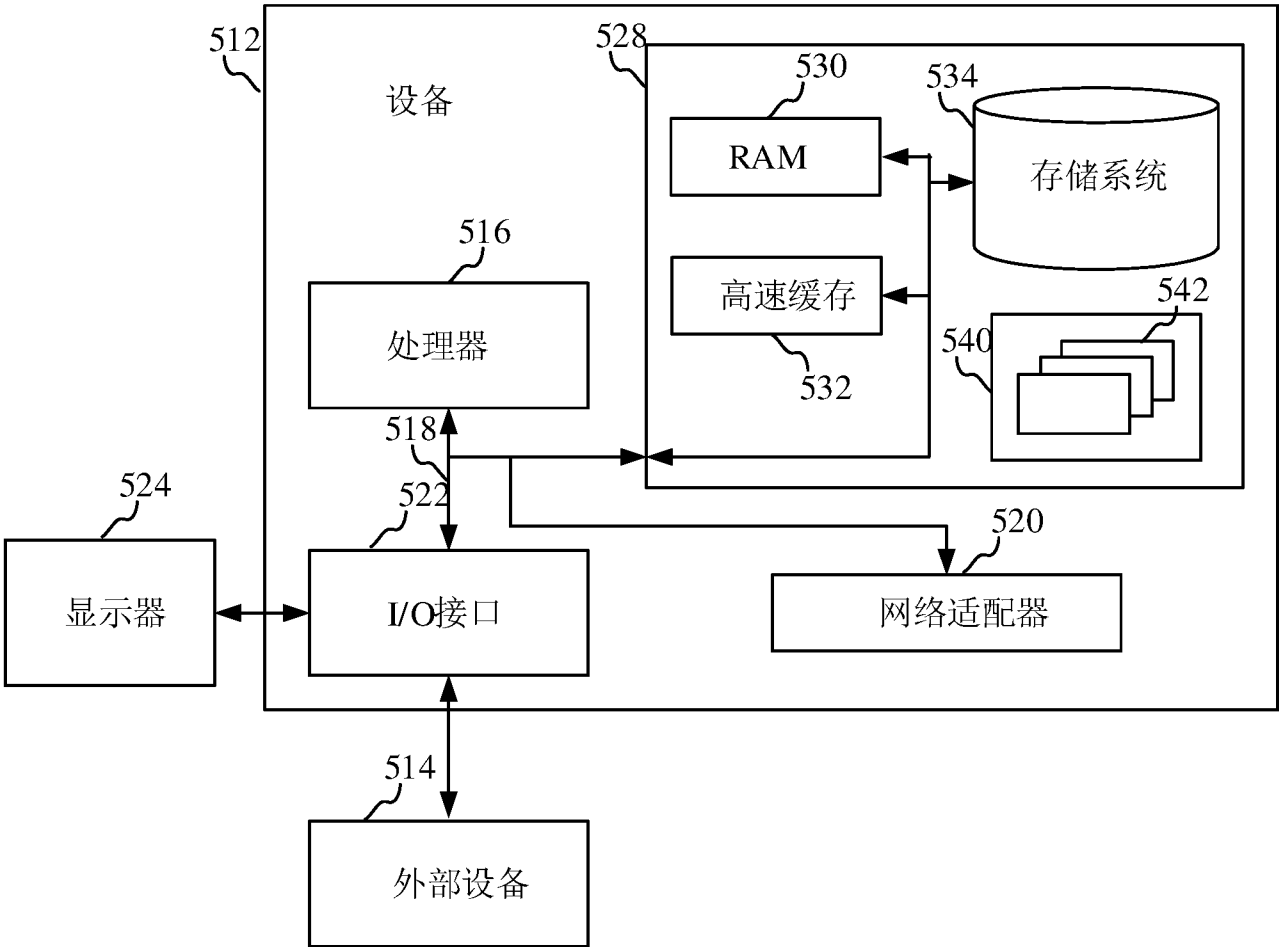


图 5

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/119633

A. CLASSIFICATION OF SUBJECT MATTER G06K 9/00(2006.01)i According to International Patent Classification (IPC) or to both national classification and IPC																				
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) G06K; G06F 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) CNPAT; CNKI; WPI; EPODOC; IEEE: 姿态, 姿势, 动作, 行为, 手势, 检测, 识别, 置信, 关键, 重点, 定位, 卷积, 神经, 网络, 训练, 模型, 前一帧, 上一帧, 之前, 前序, 在前, 帧, 当前, 现在, 最后, 结束, pose, act, behavior, gesture, detect, recognition, recognize, identify, confidence, key, position, convolution, neural, network, train, model, previous frame, present, current, before, preorder, frame, last																				
C. DOCUMENTS CONSIDERED TO BE RELEVANT <table border="1"> <thead> <tr> <th>Category*</th> <th>Citation of document, with indication, where appropriate, of the relevant passages</th> <th>Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td>PX</td> <td>CN 109558832 A (GUANGZHOU BAIGUOYUAN INFORMATION TECHNOLOGY CO., LTD.) 02 April 2019 (2019-04-02) claims 1-12</td> <td>1-12</td> </tr> <tr> <td>PX</td> <td>CN 109344755 A (GUANGZHOU BAIGUOYUAN INFORMATION TECHNOLOGY CO., LTD.) 15 February 2019 (2019-02-15) description, paragraphs [0064]-[0159]</td> <td>1-12</td> </tr> <tr> <td>A</td> <td>CN 108846365 A (SHENZHEN ZHONGYUE TECHNOLOGY CO., LTD.) 20 November 2018 (2018-11-20) description, paragraphs [0006] and [0025]-[0058], and figures 1 and 4</td> <td>1-12</td> </tr> <tr> <td>A</td> <td>CN 107798313 A (YANG, Xiaoyan) 13 March 2018 (2018-03-13) None</td> <td>1-12</td> </tr> <tr> <td>A</td> <td>CN 107832708 A (YUNDING NETWORK TECHNOLOGY BEIJING CO., LTD.) 23 March 2018 (2018-03-23) entire document</td> <td>1-12</td> </tr> </tbody> </table>			Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	PX	CN 109558832 A (GUANGZHOU BAIGUOYUAN INFORMATION TECHNOLOGY CO., LTD.) 02 April 2019 (2019-04-02) claims 1-12	1-12	PX	CN 109344755 A (GUANGZHOU BAIGUOYUAN INFORMATION TECHNOLOGY CO., LTD.) 15 February 2019 (2019-02-15) description, paragraphs [0064]-[0159]	1-12	A	CN 108846365 A (SHENZHEN ZHONGYUE TECHNOLOGY CO., LTD.) 20 November 2018 (2018-11-20) description, paragraphs [0006] and [0025]-[0058], and figures 1 and 4	1-12	A	CN 107798313 A (YANG, Xiaoyan) 13 March 2018 (2018-03-13) None	1-12	A	CN 107832708 A (YUNDING NETWORK TECHNOLOGY BEIJING CO., LTD.) 23 March 2018 (2018-03-23) entire document	1-12
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.																		
PX	CN 109558832 A (GUANGZHOU BAIGUOYUAN INFORMATION TECHNOLOGY CO., LTD.) 02 April 2019 (2019-04-02) claims 1-12	1-12																		
PX	CN 109344755 A (GUANGZHOU BAIGUOYUAN INFORMATION TECHNOLOGY CO., LTD.) 15 February 2019 (2019-02-15) description, paragraphs [0064]-[0159]	1-12																		
A	CN 108846365 A (SHENZHEN ZHONGYUE TECHNOLOGY CO., LTD.) 20 November 2018 (2018-11-20) description, paragraphs [0006] and [0025]-[0058], and figures 1 and 4	1-12																		
A	CN 107798313 A (YANG, Xiaoyan) 13 March 2018 (2018-03-13) None	1-12																		
A	CN 107832708 A (YUNDING NETWORK TECHNOLOGY BEIJING CO., LTD.) 23 March 2018 (2018-03-23) entire document	1-12																		
☒ 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 15 February 2020		Date of mailing of the international search report 20 February 2020																		
Name and mailing address of the ISA/CN China National Intellectual Property Administration No. 6, Xitucheng Road, Jimenqiao Haidian District, Beijing 100088 China Facsimile No. (86-10)62019451		Authorized officer Telephone No.																		

INTERNATIONAL SEARCH REPORT

International application No.

PCT/CN2019/119633

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2012057761 A1 (SONY CORPORATION) 08 March 2012 (2012-03-08) entire document	1-12

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/CN2019/119633

Patent document cited in search report			Publication date (day/month/year)	Patent family member(s)	Publication date (day/month/year)
CN	109558832	A	02 April 2019	None	
CN	109344755	A	15 February 2019	None	
CN	108846365	A	20 November 2018	None	
CN	107798313	A	13 March 2018	None	
CN	107832708	A	23 March 2018	None	
US	2012057761	A1	08 March 2012	CN 102385695	A 21 March 2012

国际检索报告

国际申请号

PCT/CN2019/119633

A. 主题的分类 G06K 9/00 (2006.01) i 按照国际专利分类 (IPC) 或者同时按照国家分类和 IPC 两种分类																							
B. 检索领域 检索的最低限度文献 (标明分类系统和分类号) G06K; G06F 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)) CNPAT; CNKI; WPI; EPDOC; IEEE: 姿态, 姿势, 动作, 行为, 手势, 检测, 识别, 置信, 关键, 重点, 定位, 卷积, 神经, 网络, 训练, 模型, 前一帧, 上一帧, 之前, 前序, 在前, 帧, 当前, 现在, 最后, 结束, pose, act, behavior, gesture, detect, recognition, recognize, identify, confidence, key, position, convolution, neural, network, train, model, previous frame, present, current, before, preorder, frame, last																							
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>PX</td> <td>CN 109558832 A (广州市百果园信息技术有限公司) 2019年 4月 2日 (2019 - 04 - 02) 权利要求1-12</td> <td>1-12</td> </tr> <tr> <td>PX</td> <td>CN 109344755 A (广州市百果园信息技术有限公司) 2019年 2月 15日 (2019 - 02 - 15) 说明书第[0064]-[0159]段</td> <td>1-12</td> </tr> <tr> <td>A</td> <td>CN 108846365 A (深圳市中悦科技有限公司) 2018年 11月 20日 (2018 - 11 - 20) 说明书第[0006]段, 第[0025]-[0058]段, 附图1, 4</td> <td>1-12</td> </tr> <tr> <td>A</td> <td>CN 107798313 A (杨晓艳) 2018年 3月 13日 (2018 - 03 - 13) 无</td> <td>1-12</td> </tr> <tr> <td>A</td> <td>CN 107832708 A (云丁网络技术北京有限公司) 2018年 3月 23日 (2018 - 03 - 23) 全文</td> <td>1-12</td> </tr> <tr> <td>A</td> <td>US 2012057761 A1 (SONY CORPORATION) 2012年 3月 8日 (2012 - 03 - 08) 全文</td> <td>1-12</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	PX	CN 109558832 A (广州市百果园信息技术有限公司) 2019年 4月 2日 (2019 - 04 - 02) 权利要求1-12	1-12	PX	CN 109344755 A (广州市百果园信息技术有限公司) 2019年 2月 15日 (2019 - 02 - 15) 说明书第[0064]-[0159]段	1-12	A	CN 108846365 A (深圳市中悦科技有限公司) 2018年 11月 20日 (2018 - 11 - 20) 说明书第[0006]段, 第[0025]-[0058]段, 附图1, 4	1-12	A	CN 107798313 A (杨晓艳) 2018年 3月 13日 (2018 - 03 - 13) 无	1-12	A	CN 107832708 A (云丁网络技术北京有限公司) 2018年 3月 23日 (2018 - 03 - 23) 全文	1-12	A	US 2012057761 A1 (SONY CORPORATION) 2012年 3月 8日 (2012 - 03 - 08) 全文	1-12
类 型*	引用文件, 必要时, 指明相关段落	相关的权利要求																					
PX	CN 109558832 A (广州市百果园信息技术有限公司) 2019年 4月 2日 (2019 - 04 - 02) 权利要求1-12	1-12																					
PX	CN 109344755 A (广州市百果园信息技术有限公司) 2019年 2月 15日 (2019 - 02 - 15) 说明书第[0064]-[0159]段	1-12																					
A	CN 108846365 A (深圳市中悦科技有限公司) 2018年 11月 20日 (2018 - 11 - 20) 说明书第[0006]段, 第[0025]-[0058]段, 附图1, 4	1-12																					
A	CN 107798313 A (杨晓艳) 2018年 3月 13日 (2018 - 03 - 13) 无	1-12																					
A	CN 107832708 A (云丁网络技术北京有限公司) 2018年 3月 23日 (2018 - 03 - 23) 全文	1-12																					
A	US 2012057761 A1 (SONY CORPORATION) 2012年 3月 8日 (2012 - 03 - 08) 全文	1-12																					
☐ 其余文件在C栏的续页中列出。 ☒ 见同族专利附件。																							
<table border="0"> <tr> <td> * 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 </td> <td> “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 </td> </tr> </table>			* 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																			
* 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 2020年 2月 15日		国际检索报告邮寄日期 2020年 2月 20日																					
ISA/CN的名称和邮寄地址 中国国家知识产权局 (ISA/CN) 中国北京市海淀区蓟门桥西土城路6号 100088 传真号 (86-10)62019451		授权官员 芦霞 电话号码 (86-10)53961568																					

国际检索报告
关于同族专利的信息

国际申请号

PCT/CN2019/119633

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	109558832	A	2019年 4月 2日	无	
CN	109344755	A	2019年 2月 15日	无	
CN	108846365	A	2018年 11月 20日	无	
CN	107798313	A	2018年 3月 13日	无	
CN	107832708	A	2018年 3月 23日	无	
US	2012057761	A1	2012年 3月 8日	CN 102385695 A	2012年 3月 21日