机械臂控制方法、装置、电子设备和计算机可读存储介质

相关申请的交叉引用

本申请要求在 2021 年 06 月 04 日在中国提交的中国专利申请号 202110626122.1 的优先权，其全部内容通过引用并入本文。

技术领域

本公开涉及人工智能技术领域，具体涉及计算机视觉技术领域，更具体涉及一种机械臂控制方法、装置、电子设备、计算机可读存储介质、计算机程序产品和计算机程序。

背景技术

机器人领域在分类机器人、服务机器人、人机交互等场景中有着重要的应用，近年来受到了越来越多的关注，然而，机器人手臂自动抓取紧密堆积的物体仍然是一个挑战。因为直接抓取堆叠紧密的物体一方面会导致算法无法找到一个合适的抓取位姿进行抓取，另一方面会造成机械臂夹爪与物体的碰撞，导致抓取的失败。

发明内容

本公开提供了一种提升抓取成功率的机械臂控制方法、装置、电子设备、计算机可读存储介质、计算机程序产品和计算机程序。

根据本公开的一方面实施例，提供了一种机械臂控制方法，包括：

获取目标对象的第一深度图和第一色彩图；

根据所述第一深度图和所述第一色彩图，预测得到机械臂各候选动作的第一预测值和第二预测值，其中，所述第一预测值是所述机械臂执行对应的候选动作使得所述目标对象与相邻对象分离开的概率；所述第二预测值是所述机械臂执行对应的候选动作成功抓取所述目标对象的概率；

根据各所述候选动作的第一预测值和所述第二预测值，确定目标动作；和

控制所述机械臂执行所述目标动作。

根据本公开的另一方面实施例，提供了一种机械臂控制装置，包括：

获取模块，用于获取目标对象的第一深度图和第一色彩图；

预测模块，用于根据所述第一深度图和所述第一色彩图，预测得到机械臂各候选动作的第一预测值和第二预测值，其中，所述第一预测值是所述机械臂执行对应的候选动作使得所述目标对象与相邻对象分离开的概率；所述第二预测值是所述机械臂执行对应的候选动作成功抓取所述目标对象的概率；

确定模块，用于根据各所述候选动作的第一预测值和所述第二预测值，确定目标动作；和

控制模块，用于控制所述机械臂执行所述目标动作。

根据本公开的另一方面实施例，提供了一种计算机设备，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计算机程序时，实现如前述一方面实施例所述的机械臂控制方法。

根据本公开的另一方面实施例，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，该计算机程序被处理器执行时实现如前述一方面实施例所述的机械臂控制方法。

根据本公开的另一方面实施例，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，执行如前述一方面实施例所述的机械臂控制方法。

根据本公开的另一方面实施例，提供了一种计算机程序，包括计算机程序代码，当所述计算机程序代码在计算机上运行时，使得计算机执行如前述一方面实施例所述的机械臂

控制方法。

本公开实施例所提供的技术方案可以包含如下的有益效果：获取目标对象的第一深度图和第一色彩图，根据第一深度图和第一色彩图，预测得到机械臂各候选动作的第一预测值和第二预测值，其中，第一预测值是机械臂执行对应的候选动作使得目标对象与相邻对象分离开的概率，第二预测值是机械臂执行对应的候选动作成功抓取目标对象的概率，根据各候选动作的第一预测值和第二预测值，确定目标动作，控制机械臂执行目标动作，本公开中通过预测各候选动作的第一预测值和第二预测值，实现了将目标对象与相邻对象分离的策略和抓取目标对象的策略协同分析，以从中选择出最大的预测值对应的动作，作为目标动作，提高了目标动作确定的准确性，避免了机械臂的碰撞，提高了机械臂动作执行的可靠性。

本公开附加的方面和优点将在下面的描述中部分给出，部分将从下面的描述中变得明显，或通过本公开的实践了解到。

附图说明

本公开上述的和/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 为本公开实施例所提供的一种机械臂控制方法的流程示意图；

图 2 为本公开实施例提供的另一种机械臂控制方法的流程示意图；

图 3 为本公开实施例提供的一种预测网络的结构示意图；

图 4 为本公开实施例提供的另一种机械臂控制方法的流程示意图；

图 5 为本公开实施例提供的另一种机械臂控制方法的流程示意图；

图 6 为本公开实施例提供的一种分类网络的结构示意图；

图 7 为本公开实施例提供的一种机械臂控制装置的结构示意图；

图 8 示出了适于用来实现本公开实施方式的示例性计算机设备的框图。

具体实施方式

下面详细描述本公开的实施例，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，旨在用于解释本公开，而不能理解为对本公开的限制。

下面参考附图描述本公开实施例的机械臂控制方法、装置、电子设备和存储介质。

图 1 为本公开实施例所提供的一种机械臂控制方法的流程示意图。

如图 1 所示，该方法包括以下步骤 101 至步骤 104。

步骤 101，获取目标对象的第一深度图和第一色彩图。

其中，第一深度图和第一色彩图，是机械臂坐标系下的深度图和色彩图。

本实施例中，通过获取目标对象的第一深度图和第一色彩图，可确定目标对象当前的深度信息和色彩信息。

在本公开实施例的一种实现方式中，获取采集的原始深度图和原始色彩图，将原始深度图和原始色彩图转换至机械臂坐标系下，得到机械臂坐标系下原始深度图和原始色彩图中各像素点对应的三维坐标位置。

具体来说，使用 RGB-D 相机采集原始深度图 I_{depth} 和原始色彩图 I_{color}，其中，原始深度图和原始色彩图的分辨率为 M*N，其中，M 和 N 分别为原始深度图和原始色彩图的宽和高，将原始深度图和原始色彩图中各个像素点 P_{pix} 转换到机械臂坐标系下，得到机械臂坐标系下原始深度图

和原始色彩图中各像素点对应的三维坐标位置，记为P_w:

$$P_w = R \times (K^{-1} \times z \times P_{\text{pix}}) + T$$

其中，K^{-1}表示内参矩阵的逆矩阵，R表示外参旋转矩阵，T表示外参平移矩阵，z表示相机坐标系下的深度。

进而，对各三维坐标位置投影至设定二维平面，得到各三维坐标位置对应的二维坐标点。作为一种实现方式，将各三维坐标位置中 X 轴和 Y 轴坐标(x_w, y_w)映射到尺寸大小为$H \times W$的二维平面，记为(x_s, y_s)：

$$(x_s, y_s) = (floor((x_w - x_l)/res), floor((y_w - y_l)/res))$$

其中，x_l表示机械臂坐标系X轴方向上工作空间的最小值，y_l表示机械臂坐标系Y轴方向上工作空间的最小值，res表示映射后每个像素代表的实际尺寸，$floor(\cdot)$表示向下取整操作。

在本实施例的一种实现方式中，设定二维平面是根据机械臂坐标系下沿设定方向上的最小工作距离确定的。

进一步，根据各二维坐标点在原始深度图中对应像素点的深度，生成第一深度图，根据各二维坐标点在原始色彩图中对应像素点的色彩，生成第一色彩图，具体地，将各二维坐标点在原始深度图中对应像素点的深度 Z 传递到对应的坐标(x_s, y_s)上，得到尺寸大小为$H \times W$的深度状态图I_{depth_map}；将各二维坐标点在原始色彩图中对应像素点的色彩信息(r, g, b)传递到对应的坐标(x_s, y_s)上，得到尺寸大小为$H \times W$的颜色状态图I_{color_map}。

步骤 102，根据第一深度图和第一色彩图，预测得到机械臂各候选动作的第一预测值和第二预测值，其中，第一预测值是机械臂执行对应的候选动作使得目标对象与相邻对象分离开的概率，第二预测值是机械臂执行对应的候选动作成功抓取目标对象的概率。

其中，候选动作包含用于使得目标对象和相邻对象分离开的动作，以及抓取目标对象的动作。

本实施例中，根据第一深度图和第一色彩图中携带的深度信息和色彩信息，提取对应的第一深度图的特征和第一色彩图的特征，基于提取到的第一深度图的特征和第一色彩图的特征预测得到机械臂各候选动作的第一预测值和第二预测值。

步骤 103，根据各候选动作的第一预测值和第二预测值，确定目标动作。

本实施例中，对比各候选动作的第一预测值和第二预测值，将预测值较大的候选动作作为目标动作，其中，目标动作为用于使得目标对象和相邻对象分离开的动作，或者是抓取目标对象的动作。

步骤 104，控制机械臂执行目标动作。

进而，根据确定的目标动作，控制机械臂执行目标动作，本公开实施例中，通过预测各候选动作的第一预测值和第二预测值，实现了将目标对象与相邻对象分离的策略和抓取目标对象的策略协同分析，以从中选择出最大的预测值对应的动作，作为目标动作，提高了目标动作确定的准确性，避免了机械臂的碰撞，提高了机械臂动作执行的可靠性。

本实施例的机械臂控制方法中，获取目标对象的第一深度图和第一色彩图，根据第一深度图和第一色彩图，预测得到机械臂各候选动作的第一预测值和第二预测值，其中，第一预测值是机械臂执行对应的候选动作使得目标对象与相邻对象分离开的概率，第二预测值是机械臂执行对应的候选动作成功抓取目标对象的概率，根据各候选动作的第一预测值和第二预测值，确定目标动作，控制机械臂执行目标动作，本公开中通过预测各候选动作的第一预测值和第二预测值，实现了将目标对象与相邻对象分离的策略和抓取目标对象的策略协同分析，以从中选择出最大的预测值对应的动作，作为目标动作，提高了目标动作确定的准确性，降低机械臂碰撞的概率，提高了机械臂动作执行的可靠性，进而，提升了复杂环境中物体抓取的成功率。

基于上一实施例，本实施例提供了另一种机械臂控制方法，图 2 为本公开实施例提供的另一种机械臂控制方法的流程示意图。

如图 2 所示，步骤 102 可以包括以下步骤 201 至步骤 204。

步骤 201，对第一深度图沿设定转动方向旋转多个设定角度，得到多个输入深度图。

本实施例中，为了构造不同场景下的深度图，以获取更多深度图的特征，以和后下述步骤得到的多个色彩图构成强化学习算法的状态空间，采用将第一深度图沿设定转动方向旋转多个设定角度，得到多个输入深度图。

例如，将第一深度图，在360°圆周内以$\Delta\Theta$为旋转间隔进行进行设定方向旋转，例如逆时针旋转，则$d = 360^{\circ} / \Delta\Theta$次，得到$d$组具有不同旋转角度的深度图$I^{d}_{depth_map}$，例如，得到的多个深度图的个数 d 为 16。

步骤 202，对第一色彩图沿设定转动方向旋转多个设定角度，得到多个输入色彩图。

本实施例中，为了构造不同场景下的色彩图，以获取更多色彩图的特征，以和上述得到的多个深度图构成强化学习算法的状态空间，采用将第一色彩图沿设定转动方向旋转多个设定角度，得到多个输入色彩图。

例如，将第一色彩图，在360°圆周内以$\Delta\Theta$为旋转间隔进行设定方向旋转，例如逆时针旋转，则$d = 360^{\circ} / \Delta\Theta$次，得到$d$组具有不同旋转角度的色彩图$I^{d}_{color_map}$，例如，得到的多个色彩图的个数 d 为 16。

步骤 203，将多个输入深度图和多个输入色彩图输入第一预测网络预测得到各候选动作的第一预测值。

其中，第一预测网络由特征提取层、特征融合层、预测层和动态优化层组成。

例如，如图 3 所示，其中，特征提取层为 DenseNet 网络的卷积层，例如，经过 ImageNet 预训练的 DenseNet-121 网络。特征融合层由 Batch Normalization 归一化层、Rectified Linear Unit 激活层和一个卷积核大小为 3×3 的卷积层组成。预测层为上采样 upstate 层。

在本实施例的一种实现方式中，采用第一预测网络的特征提取层，对多个输入深度图和多个输入色彩图进行特征提取，并将多个输入深度图的特征与对应的输入色彩图的特征进行融合，得到多个第一融合特征图，对各第一融合特征图沿设定转动方向的反向旋转，使得反向旋转后的第一融合特征图和第一深度图或第一色彩图的方向一致。进而，采用第一预测网络的预测层，对反向旋转后的各第一融合特征图进行动作预测，得到各候选动作的第一预测值。

例如，如图 3 所示，本实施例中以多个深度图和多个色彩图为 d=16 组为例进行说明，其中，每一组中包含一个深度图和一个色彩图。为了便于说明，以第d_i组为例，其中，深

度图$I_{color_map}^{d_i}$和色彩图$I_{color_map}^{d_i}$各自通过一个经过 ImageNet 预训练的 DenseNet-121 网络的卷积层部分进行特征提取操作，得到色彩特征图$I_{cfeat_push}^{d_i}$和深度特征图$I_{dfeat_push}^{d_i}$。进而，将色彩特征图$I_{cfeat_push}^{d_i}$和深度特征图$I_{dfeat_push}^{d_i}$进行通道拼接操作得到初始融合特征图$I_{init_fusion_push}^{d_i}$，然后将$I_{init_fusion_push}^{d_i}$经过两个结构相同的卷积组，对特征进行深度融合后，得到第一融合特征图$I_{fusion_push}^{d_i}$。进一步，将第一融合特征图$I_{fusion_push}^{d_i}$进行顺时针旋转，使得旋转后与色彩图I_{color_map}一样的角度方向，得到反向旋转后的第二融合特征图；同理，可得到其他组对应的反向旋转后的第二融合特征图。进而，采用第一预测网络的预测层，对反向旋转后的第一融合特征图进行动作预测，得到各候选动作的第一预测值。作为一种可能的实现方式，在第一预测网络的预测层，将反向旋转后的各第一融合特征图通过通道拼接操作，得到各候选动作的d维尺寸为$H\times W$的第一预测值$Q_1(s_t,a;\theta_1)$，其中，将采集得到的当前时间机械臂工作区间环境下的d组色彩图$I_{color_map}^{d}$和深度图$I_{depth_map}^{d}$记为状态s_t，θ_1表示第一预测网络参数；a表示动作空间，该动作空间由机械臂的候选动作类型，机械臂的执行位置(x_w,y_w,z_w)，夹爪旋转角度Θ和夹爪推动长度L四部分组成。

步骤 204，将多个输入深度图和多个输入色彩图输入第二预测网络预测得到各候选动作的第二预测值。

其中，第一预测网络由特征提取层、特征融合层、预测层和动态优化层组成。

例如，如图 3 所示，其中，特征提取层为 DenseNet 网络的卷积层，例如，经过 ImageNet 预训练的 DenseNet-121 网络。特征融合层由 Batch Normalization 归一化层、Rectified Linear Unit 激活层和一个卷积核大小为 3×3 的卷积层组成。预测层为 upstate 层。

在本实施例的一种实现方式中，采用第二预测网络的特征提取层，对多个输入深度图和多个输入色彩图进行特征提取，并将多个输入深度图的特征与对应的输入色彩图的特征进行融合，得到多个第二融合特征图，对各第二融合特征图沿设定转动方向的反向旋转，采用第二预测网络的预测层，对反向旋转后的各第二融合特征图进行动作预测，得到各候选动作的第二预测值。

例如，如图 3 所示，本实施例中以多个深度图和多个色彩图为 d=16 组为例进行说明，其中，每一组中包含一个深度图和一个色彩图。为了便于说明，以第d_i组为例，其中，深度图$I_{color_map}^{d_i}$和色彩图$I_{color_map}^{d_i}$各自通过一个经过 ImageNet 预训练的 DenseNet-121 网络的卷积层部分进行特征提取操作，得到色彩特征图$I_{cfeat_grasp}^{d_i}$和深度特征图$I_{dfeat_grasp}^{d_i}$。进而，将色彩特征图$I_{cfeat_grasp}^{d_i}$和深度特征图$I_{dfeat_grasp}^{d_i}$进行通道拼接操作得到初始融合特征图$I_{init_fusion_grasp}^{d_i}$，然后将$I_{init_fusion_grasp}^{d_i}$经过两个结构相同的卷积组，对特征进行深度融合后，得到对应d_i组的第二融合特征图$I_{fusion_grasp}^{d_i}$。进一步，将第二融合特征图$I_{fusion_grasp}^{d_i}$进行顺时针

旋转，使得旋转后的第二融合特征图与颜色状态图I_{color_map}一样的角度方向；同理，可得到其他组对应的反向旋转后的第二融合特征图。进而，采用第二预测网络的预测层，对反向旋转后的各第二融合特征图进行动作预测，得到各候选动作的第二预测值。作为一种可能的实现方式，在第二预测网络的预测层，将反向旋转后的各第二融合特征图通过通道拼接操作，得到各候选动作的d维尺寸为$H \times W$的第二预测值$Q_2(s_t, a; \theta_2)$，其中，将采集得到的当前机械臂工作区间环境下的d组色彩图$I^d_{color_map}$和深度图$I^d_{depth_map}$记为状态s_t，θ_2表示第二预测网络参数；a表示动作空间，该动作空间由机械臂的候选动作类型，机械臂的执行位置(x_w, y_w, z_w)，夹爪旋转角度Θ和夹爪推动长度L四部分组成。

本实施例的机械臂控制方法中，通过将第一色彩图和第一深度图沿设定转动方向旋转多个设定角度，得到多个输入色彩图和多个输入深度图，实现了构造不同场景下的色彩图和深度图，以获取更多色彩图和深度图的特征，以构成强化学习算法的状态空间，进而，基于多组深度图和色彩图组成的状态空间，分别采用第一预测网络预测得到各候选动作的第一预测值和各候选动作的第二预测值，进而基于各候选动作的第一预测值和第二预测值，实现了将目标对象与相邻对象分离的策略和抓取目标对象的策略协同分析，以从中选择出最大的预测值对应的动作，作为目标动作，提高了目标动作确定的准确性，降低机械臂碰撞的概率，提高了机械臂动作执行的可靠性，进而，提升了复杂环境中物体抓取的成功率。

基于上述实施例，本实施例提供了一种实现方式，说明了通过对预测得到的各候选动作的第一预测值和各候选动作的第二预测值进行修正，以提高第一预测值和第二预测值的准确性，进而提高目标动作确定的准确性。

图 4 为本公开实施例提供的另一种机械臂控制方法的流程示意图，如图 4 所示，步骤 103 包含以下步骤 401 至步骤 403。

步骤 401，根据第一深度图指示的目标对象的轮廓，对各候选动作的第一预测值进行修正。

作为一种实现方式，根据第一深度图指示的目标对象的轮廓，计算得到动态掩码，将动态掩码与各候选动作的第一预测值相乘，得到各候选动作修正后的第一预测值。

步骤 402，根据第一深度图指示的目标对象的中心位置，对各候选动作的第二预测值进行修正。

作为一种实现方式，根据第一深度图指示的目标对象的中心位置，计算得到动态掩码，将动态掩码与各候选动作的第二预测值相乘，得到各候选动作修正后的第二预测值。

步骤 403，根据修正后的第一预测值和修正后的第二预测值，从各候选动作中选取目标动作。

本实施例的一种实现方式中，根据各候选动作修正后的第一预测值，确定修正后的第一预测值中的最大值，以及根据各候选动作修正后的第二预测值，确定修正后的第二预测值中的最大值，进而，比较修正后的第一预测值的最大值和第二预测值的最大值，以确定最大的预测值对应的候选动作，将最大的预测值对应的候选动作作为目标动作。

在一种场景下，候选动作为“推动作”和“抓取动作”，确定“推动作”修正后的第一预测值和“抓取动作”修正后的第一预测值，进而，确定修正后的第一预测值中的最大值，例如，最大值为“推动作”的修正后的第一预测值，以及确定“推动作”修正后的第二预测值和“抓取动作”修正后的第二预测值，进而，确定修正后的第二预测值中的最大值，例如，最大值为“抓取动作”的修正后的第二预测值。进一步，比较“推动作”的修正后的第一预测值和“抓取动作”的修正后的第二预测值，选择较大的预测值，例如为“推动作”修正后的第一阈值值，则将“推动作”作为机械臂的目标动作。

本实施例的机械臂的控制方法中，通过对各候选动作的第一预测值和各候选动作的第二预测值进行修正，提高了各候选动作对应的第一预测值和第二预测值的准确性，进而根据修正后的各候选动作的第一预测值和第二预测值进行目标动作的确定，实现了将目标对象与相邻对象分离的策略和抓取目标对象的策略协同分析，以从中选择出最大的预测值对应的动作，作为目标动作，提高了目标动作确定的准确性，降低机械臂碰撞的概率，提高了机械臂动作执行的可靠性，进而，提升了复杂环境中物体抓取的成功率。

基于上述实施例，本实施例提供了一种可能的实现方式，具体说明了如何对第一预测网络和第二预测网络进行训练。

图 5 为本公开实施例提供的另一种机械臂控制方法的流程示意图，如图 5 所示，步骤 104 之后包含以下步骤 501 至步骤 504。

步骤 501，采集目标动作执行之后目标对象的第二深度图。

其中，控制机械臂执行目标动作之后，目标对象的位置分布会发生变化，则获取采集目标动作执行之后目标对象的第二深度图，其中，第二深度图的获取方法可参照前述实施例中的说明，本实施例中不再限定。

步骤 502，根据第二深度图和第一深度图，采用分类网络，确定目标动作的第一奖励值，其中，第一奖励值用于指示机械臂执行目标动作使得目标对象与相邻对象分离开的有效程度。

其中，第一深度度即为未控制机械臂执行目标动作前获取到的深度图。

如图 6 所示，图 6 为本公开实施例提供的一种分类网络的结构示意图，如图 6 所示，将第一深度图和第二深度图作为分类网络的输入，各自通过一个经过预训练的 VGG16 网络的卷积层进行特征提取，并对提取到的第一深度图的特征和第二深度图的特征进行通道拼接操作得到融合特征图，进而，融合特征图经过卷积核大小为 1×1 的卷积层得到卷积特征图 1，进而，将卷积特征图 1 输入卷积核大小为 3×3 的卷积层得到卷积特征图 2；最后，卷积特征图 2 经过三个全连接层，确定目标动作使得目标对象分离的程度，根据分离的程度确定目标动作的第一奖励值。

例如，第一奖励值为 r_p，

$$r_p = \begin{cases} -0.5 & \text{if output} = 0 \\ 0.5 & \text{if output} = 1 \end{cases}$$

其中，$output = 0$ 表示目标动作能够将目标物体进行聚合，这不利于目标物体的成功抓取，确定这类目标动作的第一奖励值，即为 -0.5 的惩罚值；$output = 1$ 表示目标动作能够将物体进行分离，这有利于目标物体的成功抓取，确定这类目标动作的第一奖励值，即为正 0.5 的奖励值。

步骤 503，根据机械臂对目标对象是否抓取成功，确定目标动作的第二奖励值。

本实施例的一种实现方式中，第二奖励值 r_g 定义为：$r_g = \begin{cases} 0 & \text{抓取失败} \\ 1.5 & \text{抓取成功} \end{cases}$

如果机械臂执行目标动作后，对目标对象抓取成功，确定目标动作的第二奖励值为 1.5，如果机械臂执行目标动作后，对目标对象抓取成功，确定目标动作的第二奖励值为 0。

步骤 504，根据第一奖励值训练第一预测网络，以及根据第二奖励值训练第二预测网络。

在本实施例的一种实现方式中，第一预测网络的损失函数为：

$$Loss_{Q_1}=\left[Q_1(s_t,a_t;\theta_1)-(r_p+\gamma Q_1(s_{t+1},\arg\max_{a_{t+1}}Q(s_{t+1},a;\theta_1);\theta_{1target}))\right]^2$$

其中，r_p 是第一奖励值，s_t 是时刻 t 时对应的多个深度图和多个色彩图组成的状态空间，$Q_1(s_t,a_t;\theta)$ 是 t 时刻第一预测网络的目标动作的预测值的价值函数，θ_1 是当前时刻 t 的网络参数，$\theta_{1target}$ 是第一预测网络的目标网络参数，$Q_1(s_{t+1},\arg\max_{a_{t+1}}Q(s_{t+1},a;\theta_1);\theta_{1target})$ 是 t+1 时刻第一预测网络的目标动作的预测值的价值函数，γ 表示衰减因子。

根据第一奖励值确定的损失函数，不断调整第一预测网络的参数，以对第一预测网络进行训练。

在本实施例的一种实现方式中，第二预测网络的损失函数为：

$$Loss_{Q_2}=\left[Q_2(s_t,a_t;\theta_2)-(r_g+\gamma Q_2(s_{t+1},\arg\max_{a_{t+1}}Q_2(s_{t+1},a;\theta_2);\theta_{2target}))\right]^2$$

其中，rg 是第二奖励值，s_t 是时刻 t 时对应的多个深度图和多个色彩图组成的状态空间，$Q_2(s_t,a_t;\theta_2)$ 是 t 时刻第二预测网络的目标动作的预测值的价值函数，θ_2 是当前时刻 t 的网络参数，$\theta_{2target}$ 是第二预测网络的目标网络参数，$Q_2(s_{t+1},\arg\max_{a_{t+1}}Q_2(s_{t+1},a;\theta_2);\theta_{2target})$ 是 t+1 时刻第二预测网络的目标动作的预测值的价值函数，γ 表示衰减因子。

根据第二奖励值确定的损失函数，不断调整第二预测网络的参数，以对第二预测网络进行训练。

本实施例的机械臂控制方法中，通过确定第一奖励值和第二奖励值，采用第一奖励值对第一预测网络进行训练，使得训练得到的第一预测网络学习得到多个深度提和多个彩色图与各候选动作的第一预测值的对应关系，并采用第二奖励值对第二预测网络进行训练，使得训练得到的第二预测网络学习得到多个深度提和多个彩色图与各候选动作的第二预测值的对应关系。

为了实现上述实施例，本公开还提出一种机械臂控制装置。

图 7 为本公开实施例提供的一种机械臂控制装置的结构示意图。

如图 7 所示，该装置包括获取模块 71、预测模块 72、确定模块 73 和控制模块 74。

获取模块 71，用于获取目标对象的第一深度图和第一色彩图。

预测模块 72，用于根据所述第一深度图和所述第一色彩图，预测得到机械臂各候选动作的第一预测值和第二预测值，其中，所述第一预测值是所述机械臂执行对应的候选动作使得所述目标对象与相邻对象分离开的概率；所述第二预测值是所述机械臂执行对应的候选动作成功抓取所述目标对象的概率。

确定模块 73，用于根据各所述候选动作的第一预测值和所述第二预测值，确定目标动作。

控制模块 74，用于控制所述机械臂执行所述目标动作。

进一步地，在本公开实施例的一种可能的实现方式中，预测模块 72，用于：

对所述第一深度图沿设定转动方向旋转多个设定角度，得到多个输入深度图；

对所述第一色彩图沿所述设定转动方向旋转所述多个设定角度，得到多个输入色彩图；

将所述多个输入深度图和所述多个输入色彩图输入第一预测网络预测得到各所述候选动作的所述第一预测值；

将所述多个输入深度图和所述多个输入色彩图输入第二预测网络预测得到各所述候选动作的所述第二预测值。

在本公开实施例的一种可能的实现方式中，所述装置还包括：

采集模块，用于采集所述目标动作执行之后所述目标对象的第二深度图；

处理模块，用于根据所述第二深度图和所述第一深度图，采用分类网络，确定所述目标动作的第一奖励值，其中，所述第一奖励值用于指示所述机械臂执行所述目标动作使得所述目标对象与相邻对象分离开的有效程度；

所述确定模块，用于根据所述机械臂对所述目标对象是否抓取成功，确定所述目标动作的第二奖励值；

训练模块，用于根据所述第一奖励值训练所述第一预测网络，以及根据所述第二奖励值训练所述第二预测网络。

在本公开实施例的一种可能的实现方式中，上述处理模块，用于：

采用分类网络的特征提取层，分别对所述第一深度图和所述第二深度图进行特征提取；

将所述第一深度图的特征与所述第二深度图的特征进行融合，得到第一融合特征；

采用所述分类网络的分类层对所述第一融合特征进行分类预测，得到所述目标动作的第一奖励值。

在本公开实施例的一种可能的实现方式中，预测模块 72，具体用于：

采用所述第一预测网络的特征提取层，对所述多个输入深度图和所述多个输入色彩图进行特征提取，并将所述多个输入深度图的特征与对应的输入色彩图的特征进行融合，得到多个第一融合特征图；

对各所述第一融合特征图沿所述设定转动方向的反向旋转；

采用所述第一预测网络的预测层，对反向旋转后的各所述第一融合特征图进行动作预测，得到各所述候选动作的第一预测值。

在本公开实施例的一种可能的实现方式中，预测模块 72，具体用于：

采用所述第二预测网络的特征提取层，对所述多个输入深度图和所述多个输入色彩图进行特征提取，并将所述多个输入深度图的特征与对应的输入色彩图的特征进行融合，得到多个第二融合特征图；

对各所述第二融合特征图沿所述设定转动方向的反向旋转；

采用所述第二预测网络的预测层，对反向旋转后的各所述第二融合特征图进行动作预测，得到各所述候选动作的第二预测值。

在本公开实施例的一种可能的实现方式中，确定模块 73，具体用于：

根据所述第一深度图指示的所述目标对象的轮廓，对各所述候选动作的第一预测值进行修正；

根据所述第一深度图指示的所述目标对象的中心位置，对各所述候选动作的第二预测值进行修正；

根据修正后的第一预测值和修正后的第二预测值，从各所述候选动作中选取所述目标动作。

在本公开实施例的一种可能的实现方式中，获取模块 71，具体用于：

获取采集的原始深度图和原始色彩图；

将所述原始深度图和所述原始色彩图转换至所述机械臂坐标系下，得到所述机械臂坐标系下所述原始深度图和所述原始色彩图中各像素点对应的三维坐标位置；

对各所述三维坐标位置投影至设定二维平面，得到各三维坐标位置对应的二维坐标点；

根据各所述二维坐标点在所述原始深度图中对应像素点的深度，生成所述第一深度图；

根据各所述二维坐标点在所述原始色彩图中对应像素点的色彩，生成所述第一色彩图。

在本公开实施例的一种可能的实现方式中，所述设定二维平面是根据所述机械臂坐标系下沿设定方向上的最小工作距离确定的。

需要说明的是，前述对方法实施例的解释说明也适用于该实施例的装置，此处不再赘述。

本实施例的机械臂控制装置中，获取目标对象的第一深度图和第一色彩图，根据第一深度图和第一色彩图，预测得到机械臂各候选动作的第一预测值和第二预测值，其中，第一预测值是机械臂执行对应的候选动作使得目标对象与相邻对象分离开的概率，第二预测值是机械臂执行对应的候选动作成功抓取目标对象的概率，根据各候选动作的第一预测值和第二预测值，确定目标动作，控制机械臂执行目标动作，本公开中通过预测各候选动作的第一预测值和第二预测值，实现了将目标对象与相邻对象分离的策略和抓取目标对象的策略协同分析，以从中选择出最大的预测值对应的动作，作为目标动作，提高了目标动作确定的准确性，降低机械臂碰撞的概率，提高了机械臂动作执行的可靠性，进而，提升了复杂环境中物体抓取的成功率。

为了实现上述实施例，本公开实施例提供了一种计算机设备，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计算机程序时，实现如前述方法实施例所述的机械臂控制方法。

为了实现上述实施例，本公开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，该计算机程序被处理器执行时实现如前述方法实施例所述的机械臂控制方法。

为了实现上述实施例，本公开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，执行如前述方法实施例所述的机械臂控制方法。

为了实现上述实施例，本公开实施例提供了一种计算机程序，包括计算机程序代码，当所述计算机程序代码在计算机上运行时，使得计算机执行如前述方法实施例所述的机械臂控制方法。

需要说明的是，前述对方法实施例的解释说明也适用于上述实施例的计算机设备、非临时性计算机可读存储介质、计算机程序产品和计算机程序，此处不再赘述。

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

如图 8 所示，计算机设备 12 以通用计算设备的形式表现。计算机设备 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，存储器 28，连接不同系统组件（包括存储器 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。

计算机设备 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机设备 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。计算机设备 12 可以进一步包括其它可移动/不可移动的、易失性/非易失性计算机系统存储介质。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 8 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公开各实施例的功能。

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 通常执行本公开所描述的实施例中的功能和/或方法。

计算机设备 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设备 12 交互的设备通信，和/或与使得该计算设备 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，计算机设备 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 与计算机设备 12 的其它模块通信。应当明白，尽管图中未示出，可以结合计算机设备 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。

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公开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。在本公开的描述中，“多个”的含义是至少两个，例如两个，三个等，除非另有明确具体的限定。

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定制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，并且本公开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，这应被本公开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，可以被认为是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。计算机可读介质的更具体的示例（非穷尽性列表）包括以下：具有一个或多个布

线的电连接部（电子装置），便携式计算机盘盒（磁装置），随机存取存储器（RAM），只读存储器（ROM），可擦除可编辑只读存储器（EPROM 或闪速存储器），光纤装置，以及便携式光盘只读存储器（CDROM）。另外，计算机可读介质甚至可以是可在其上打印所述程序的纸或其他合适的介质，因为可以例如通过对纸或其他介质进行光学扫描，接着进行编辑、解译或必要时以其他合适方式进行处理来以电子方式获得所述程序，然后将其存储在计算机存储器中。

应当理解，本公开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。如，如果用硬件来实现和在另一实施方式中一样，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。

本技术领域的普通技术人员可以理解实现上述实施例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。

此外，在本公开各个实施例中的各功能单元可以集成在一个处理模块中，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模块中。上述集成的模块既可以采用硬件的形式实现，也可以采用软件功能模块的形式实现。所述集成的模块如果以软件功能模块的形式实现并作为独立的产品销售或使用时，也可以存储在一个计算机可读取存储介质中。

上述提到的存储介质可以是只读存储器，磁盘或光盘等。尽管上面已经示出和描述了本公开的实施例，可以理解的是，上述实施例是示例性的，不能理解为对本公开的限制，本领域的普通技术人员在本公开的范围内可以对上述实施例进行变化、修改、替换和变型。

本公开所有实施例均可以单独被执行，也可以与其他实施例相结合被执行，均视为本公开要求的保护范围。

权利要求书

1、一种机械臂控制方法，包括以下步骤：

获取目标对象的第一深度图和第一色彩图；

根据所述第一深度图和所述第一色彩图，预测得到机械臂各候选动作的第一预测值和第二预测值，其中，所述第一预测值是所述机械臂执行对应的候选动作使得所述目标对象与相邻对象分离开的概率；所述第二预测值是所述机械臂执行对应的候选动作成功抓取所述目标对象的概率；

根据各所述候选动作的第一预测值和所述第二预测值，确定目标动作；和

控制所述机械臂执行所述目标动作。

2、根据权利要求 1 所述的机械臂控制方法，其中所述根据所述第一深度图和所述第一色彩图，预测得到机械臂各候选动作的第一预测值和第二预测值，包括：

对所述第一深度图沿设定转动方向旋转多个设定角度，得到多个输入深度图；

对所述第一色彩图沿所述设定转动方向旋转所述多个设定角度，得到多个输入色彩图；

将所述多个输入深度图和所述多个输入色彩图输入第一预测网络预测得到各所述候选动作的所述第一预测值；和

将所述多个输入深度图和所述多个输入色彩图输入第二预测网络预测得到各所述候选动作的所述第二预测值。

3、根据权利要求 1 或 2 所述的机械臂控制方法，还包括：

采集所述目标动作执行之后所述目标对象的第二深度图；

根据所述第二深度图和所述第一深度图，采用分类网络，确定所述目标动作的第一奖励值，其中，所述第一奖励值用于指示所述机械臂执行所述目标动作使得所述目标对象与相邻对象分离开的有效程度；

根据所述机械臂对所述目标对象是否抓取成功，确定所述目标动作的第二奖励值；和

根据所述第一奖励值训练所述第一预测网络，以及根据所述第二奖励值训练所述第二预测网络。

4、根据权利要求 3 所述的机械臂控制方法，其中所述根据所述第二深度图和所述第一深度图，采用分类网络，确定所述目标动作的第一奖励值，包括：

采用分类网络的特征提取层，分别对所述第一深度图和所述第二深度图进行特征提取；

将所述第一深度图的特征与所述第二深度图的特征进行融合，得到第一融合特征；和

采用所述分类网络的分类层对所述第一融合特征进行分类预测，得到所述目标动作的第一奖励值。

5、根据权利要求 2-4 中任一项所述的机械臂控制方法，其中所述将所述多个输入深度图和所述多个输入色彩图输入第一预测网络预测得到各所述候选动作的所述第一预测值，包括：

采用所述第一预测网络的特征提取层，对所述多个输入深度图和所述多个输入色彩图进行特征提取，并将所述多个输入深度图的特征与对应的输入色彩图的特征进行融合，得到多个第一融合特征图；

对各所述第一融合特征图沿所述设定转动方向的反向旋转；和

采用所述第一预测网络的预测层，对反向旋转后的各所述第一融合特征图进行动作预测，得到各所述候选动作的第一预测值。

6、根据权利要求 2-5 中任一项所述的机械臂控制方法，其中所述将所述多个输入深度图和所述多个输入色彩图输入第二预测网络预测得到各所述候选动作的所述第二预测值，包括：

采用所述第二预测网络的特征提取层，对所述多个输入深度图和所述多个输入色彩图进行特征提取，并将所述多个输入深度图的特征与对应的输入色彩图的特征进行融合，得到多个第二融合特征图；

对各所述第二融合特征图沿所述设定转动方向的反向旋转；和

采用所述第二预测网络的预测层，对反向旋转后的各所述第二融合特征图进行动作预测，得到各所述候选动作的第二预测值。

7、根据权利要求 1-6 任一项所述的机械臂控制方法，其中所述根据各所述候选动作的第一预测值和所述第二预测值，确定目标动作，包括：

根据所述第一深度图指示的所述目标对象的轮廓，对各所述候选动作的第一预测值进行修正；

根据所述第一深度图指示的所述目标对象的中心位置，对各所述候选动作的第二预测值进行修正；和

根据修正后的第一预测值和修正后的第二预测值，从各所述候选动作中选取所述目标动作。

8、根据权利要求 1-7 任一项所述的机械臂控制方法，其中所述获取目标对象的第一深度图和第一色彩图，包括：

获取采集的原始深度图和原始色彩图；

将所述原始深度图和所述原始色彩图转换至所述机械臂坐标系下，得到所述机械臂坐标系下所述原始深度图和所述原始色彩图中各像素点对应的三维坐标位置；

对各所述三维坐标位置投影至设定二维平面，得到各三维坐标位置对应的二维坐标点；

根据各所述二维坐标点在所述原始深度图中对应像素点的深度，生成所述第一深度图；和

根据各所述二维坐标点在所述原始色彩图中对应像素点的色彩，生成所述第一色彩图。

9、根据权利要求 8 所述的机械臂控制方法，其中所述设定二维平面是根据所述机械臂坐标系下沿设定方向上的最小工作距离确定的。

10、一种机械臂控制装置，包括：

获取模块，用于获取目标对象的第一深度图和第一色彩图；

预测模块，用于根据所述第一深度图和所述第一色彩图，预测得到机械臂各候选动作的第一预测值和第二预测值，其中，所述第一预测值是所述机械臂执行对应的候选动作使得所述目标对象与相邻对象分离开的概率；所述第二预测值是所述机械臂执行对应的候选动作成功抓取所述目标对象的概率；

确定模块，用于根据各所述候选动作的第一预测值和所述第二预测值，确定目标动作；和

控制模块，用于控制所述机械臂执行所述目标动作。

11、一种计算机设备，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计算机程序时，实现如权利要求 1-9 中任一项所述的机械臂控制方法。

12、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，其中所述计算机程序被处理器执行时实现如权利要求 1-9 中任一项所述的机械臂控制方法。

13、一种计算机程序产品，其中当所述计算机程序产品中的指令由处理器执行时，执行如权利要求 1-9 中任一项所述的机械臂控制方法。

14、一种计算机程序，包括计算机程序代码，当所述计算机程序代码在计算机上运行时，使得计算机执行如权利要求 1-9 中任一项所述的机械臂控制方法。