

虚拟角色的动作决策方法、装置、设备及存储介质

本申请要求于2023年09月15日提交的申请号为202311198980.6、发明名称为“虚拟角色的动作决策方法、装置、设备及存储介质”的中国专利申请的优先权，其全部内容通过引用结合在本申请中。

技术领域

本申请实施例涉及人工智能技术领域，特别涉及一种虚拟角色的动作决策方法、装置、设备及存储介质。

背景技术

现如今，在策划电子游戏过程中，通常希望NPC（Non-Player Character，非玩家角色）的行为逻辑能够与真人玩家的行为逻辑尽可能一致，即使得游戏中的NPC具有高拟人的特性。

相关技术中，在基于神经网络模型确定虚拟角色的行为逻辑时，通常会根据当前虚拟角色在虚拟环境中的属性信息以及虚拟环境信息预测虚拟角色行为。其中，可以通过对当前虚拟环境的原始像素图像进行图像识别的方式，确定当前虚拟角色所处虚拟环境的虚拟环境信息，之后将虚拟环境信息输入至神经网络模型，并控制虚拟角色执行神经网络模型输出的动作。

然而，有些虚拟环境的规模较大且其中的场景内容较为复杂，如虚拟环境中存在各种移动的障碍物，例如飞机、汽车以及虚拟角色等等，若基于对当前虚拟环境的图像进行图像识别的方式确定虚拟环境信息，对虚拟环境的图像进行采集时需要消耗的资源较多，使得确定虚拟角色动作的复杂度较高，且人力成本较高。

发明内容

本申请实施例提供了一种虚拟角色的虚拟角色的动作决策方法、装置、设备及存储介质。所述技术方案如下：

一方面，本申请实施例提供了一种虚拟角色的动作决策方法，由计算机设备执行，该方法包括：

获取第一虚拟角色的角色状态信息，以及获取所述第一虚拟角色对当前所处虚拟环境的环境感知信息，其中，所述环境感知信息包括多个模态的感知信息，所述多个模态分别代表一个信息维度；

融合所述角色状态信息和所述环境感知信息分别对应的编码结果，得到融合特征，其中，所述编码结果是通过不同编码方式对所述角色状态信息以及多个感知信息分别进行特征编码得到的结果；

基于所述融合特征对所述第一虚拟角色进行动作决策，得到角色动作；

控制所述第一虚拟角色执行所述角色动作。

另一方面，本实施例提供了一种虚拟角色的动作决策装置，装置包括：

获取模块，用于获取第一虚拟角色的角色状态信息，以及获取所述第一虚拟角色对当前所处虚拟环境的环境感知信息，其中，所述环境感知信息包括多个模态的感知信息，所述多个模态分别代表一个信息维度；

编码模块，用于融合所述角色状态信息和所述环境感知信息分别对应的编码结果，得到融合特征，其中，所述编码结果是通过不同编码方式对所述角色状态信息以及多个感知信息分别进行特征编码得到的结果；

决策模块，用于基于所述融合特征对所述第一虚拟角色进行动作决策，得到角色动作；

控制模块，用于控制所述第一虚拟角色执行所述角色动作。

另一方面，本申请实施例提供了一种计算机设备，所述计算机设备包括处理器和存储器，

所述存储器中存储有至少一条指令、至少一段程序、代码集或指令集，所述至少一条指令、所述至少一段程序、所述代码集或指令集由所述处理器加载并执行以实现如上述方面所述的虚拟角色的动作决策方法。

另一方面，提供了一种计算机可读存储介质，所述可读存储介质中存储有至少一条指令、至少一段程序、代码集或指令集，所述至少一条指令、所述至少一段程序、所述代码集或指令集由处理器加载并执行以实现如上述方面所述的虚拟角色的动作决策方法。

另一方面，本申请实施例提供了一种计算机程序产品或计算机程序，该计算机程序产品或计算机程序包括计算机指令，该计算机指令存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该计算机指令，处理器执行该计算机指令，使得该计算机设备执行上述方面提供的虚拟角色的动作决策方法。

本申请实施例提供的技术方案带来的有益效果至少包括：

本申请实施例中，在虚拟环境中需要对第一虚拟角色进行动作决策时，获取第一虚拟角色的角色状态信息以及多模态的环境感知信息，从而对角色状态信息以及多模态的环境感知信息进行特征编码融合，最终基于融合特征进行动作决策。在进行动作决策时获取第一虚拟角色感知到的不同信息维度的环境感知信息，有助于获取到第一虚拟角色周围更加全面、细致的三维虚拟环境，从而能够进行更加合理的动作决策。采用不同的编码方式对角色状态信息以及多个不同模态的感知信息进行特征编码，能够使得动作决策模型对当前场景下的各类信息拟合更加精准。此外，相较于相关技术中采用图像识别的方式获取环境特征的方式，通过对量化的多模态的环境感知信息进行特征提取进而融合角色状态信息的编码结果，有助于在量化环境感知信息的同时更丰富地考虑第一虚拟角色本身的角色状态信息，在节省计算资源的同时提高决策的角色动作的准确性，使得第一虚拟角色执行角色动作的效果更加逼真。

附图说明

图1示出了本申请一个示例性实施例提供的实施环境的示意图；

图2示出了本申请一个示例性实施例提供的虚拟角色的动作决策方法的流程图；

图3示出了本申请一个示例性实施例提供的射线探测方案的示意图；

图4示出了本申请一个示例性实施例提供的第二模态感知信息的示意图；

图5示出了本申请一个示例性实施例提供的占栅网格结构的示意图；

图6示出了本申请一个示例性实施例提供的生成角色动作的过程的流程图；

图7示出了本申请一个示例性实施例提供的对环境感知信息进行一维卷积的过程示意图；

图8示出了本申请一个示例性实施例提供的用于提供虚拟环境的应用程序的用户界面示意图；

图9示出了本申请一个示例性实施例提供的移动方向以及描述动作的子动作组成示意图；

图10示出了本申请一个示例性实施例提供的n个动作输出头的自回归结构的示意图；

图11示出了本申请一个示例性实施例提供的可视情况感知的示意图；

图12示出了本申请一个示例性实施例提供的确定竖直方向有效瞄准范围的示意图；

图13示出了本申请一个示例性实施例提供的水平方向有效瞄准范围的示意图；

图14示出了本申请一个示例性实施例提供的动作决策模型的示意图；

图15示出了本申请一个示例性实施例提供的动作决策模型训练过程的流程图；

图16示出了本申请一个示例性实施例提供的对深度图进行区域划分的示意图；

图17示出了本申请一个示例性实施例提供的角色动作中存在转向子动作的示意图；

图18示出了本申请一个示例性实施例提供的虚拟角色的动作决策装置的结构示意图；

图19示出了本申请一个示例性实施例提供的计算机设备的结构示意图。

具体实施方式

为使本申请的目的、技术方案和优点更加清楚，下面将结合附图对本申请实施方式作进一步地详细描述。

本申请实施例涉及人工智能（Artificial Intelligence, AI）和机器学习技术，基于人工智能中的机器学习（Machine Learning, ML）而设计。

人工智能是利用数字计算机或者数字计算机控制的机器模拟、延伸和扩展人的智能，感知环境、获取知识并使用知识获得最佳结果的理论、方法、技术及应用系统。强化学习、深度学习等人工智能技术在多种领域具有广泛的应用。示意性的，本申请涉及机器学习中的强化学习技术。

强化学习（Reinforcement Learning, RL）是机器学习的一个分支，agent（智能体）通过与 env（环境）交互来进行学习。这是一个以结果为导向的学习过程，agent 不被告知要采取什么 action（行为）；相反，agent 从其 action 的结果中学习。

智能体能通过传感器感知环境和通过执行器作用与环境，对于每种可能的感知序列，智能体应当选择一个动作来执行，该动作能够在它具有指示和感知序列提供的证据条件下，使它的性能达到期望最大值。本申请实施例中，第一虚拟角色作为智能体感知虚拟环境，获取多模态的环境感知信息，并执行动作决策模型确定的角色动作。

虚拟环境是应用程序在计算机设备上运行时显示（或提供）的虚拟环境。该虚拟环境可以是对真实世界的仿真环境，也可以是半仿真半虚构的环境，还可以是纯虚构的环境。虚拟环境可以是二维虚拟环境、2.5 维虚拟环境和三维虚拟环境中的任意一种，本申请对此不加以限定。下述实施例以虚拟环境是三维虚拟环境来举例说明。

虚拟角色是指在虚拟环境中的可活动对象。该可活动对象可以是虚拟角色、虚拟动物、动漫人物中的至少一种。可选地，当虚拟环境为三维虚拟环境时，虚拟角色可以是三维虚拟模型，每个虚拟角色在三维虚拟环境中具有自身的形状和体积，占据三维虚拟环境中的一部分空间。可选地，虚拟角色是基于三维人体骨骼技术构建的三维角色，该虚拟角色通过穿戴不同的皮肤来实现不同的外在形象。在一些实现方式中，虚拟角色也可以采用 2.5 维或 2 维模型来实现，本申请实施例对此不加以限定。

相关技术中，通过使用原始像素的方式作为环境感知的方式，基于原始像素信息做图像识别的方式，需要从客户端采集图像数据再进行图像识别的方式得到虚拟环境的状态。然而，有些虚拟环境的规模较大且其中的场景内容较为复杂，如虚拟环境中存在各种移动的障碍物，例如飞机、汽车以及虚拟角色等等，若基于对当前虚拟环境的图像进行图像识别的方式确定虚拟环境信息，对虚拟环境的图像进行采集时需要消耗的资源较多，使得确定虚拟角色动作的复杂度较高，且人力成本较高。

此外，在相关技术中，可能会在虚拟角色所处位置的导航网格中发射多条射线，射线碰到导航网络边缘时停止，再基于场景查询系统筛选出若干寻路点，从而保证虚拟角色的可通行性。然而，采用查询离散点的方式仅仅为第一虚拟角色提供了若干达到目的地的寻路点，并不能够清晰的表达周围环境特征，不利于深度神经网络获取周围环境特征。且当深度神经网络进行动作决策的帧率较高的情况下，虚拟角色会在不同寻路点之间拉扯摆动，出现虚拟角色抽搐的现象。因此，本申请实施例提供了一种虚拟角色的动作决策方法，通过获取第一虚拟角色感知的多模态的环境感知特征，能够获取到更精细的环境特征，从而实现对周围环境以及第一虚拟角色的状态特征的融合，得到融合特征。进而根据融合特征进行动作决策，控制第一虚拟角色的动作。

本申请中的计算机设备可以为台式计算机、膝上编写计算机、手机、平板电脑、台式电脑、电子书阅读器、MP3（Moving Picture Experts Group Audio Layer III，动态影像专家压缩标准音频层面 3）播放器、MP4（Moving Picture Experts Group Audio Layer IV，动态影像专家压缩标准音频层面 4）播放器、智能语音交互设备、智能家电以及车载终端等等。该计算机设备中安装和运行有支持虚拟环境的应用程序，比如支持三维虚拟环境的应用程序。该应用

程序可以是虚拟现实应用程序、三维地图程序、TPS (Third-Person Shooter, 第三人称射击) 游戏、FPS (First-Person Shooter, 第一人称射击) 游戏、MOBA (Multiplayer Online Battle Arena, 多人在线战术竞技) 游戏中的任意一种。可选地, 该应用程序可以是单机版的应用程序, 比如单机版的 3D 游戏程序, 也可以是网络联机版的应用程序。下述实施例是以在游戏中的应用来举例说明。

基于虚拟环境的游戏往往由一个或多个游戏世界的地图构成, 游戏中的虚拟环境模拟现实世界的场景, 第一虚拟角色在虚拟环境中可以进行行走、跑步、跳跃、射击、格斗、驾驶、攀爬、滑翔、切换使用虚拟道具、使用虚拟道具攻击其他虚拟角色等动作。

请参考图 1, 其示出了本申请一个示例性实施例提供的实施环境的示意图。该实施环境可以包括: 终端 110、第一服务器 120 以及第二服务器 130。

在本申请实施例中, 终端 110 包括但不限于手机、平板电脑、笔记本电脑、台式电脑、电子书阅读器、智能语音交互设备、智能家电以及车载终端等设备。终端 110 中运行有支持虚拟环境的应用程序 111, 该应用程序可以是 TPS (Third-Person Shooter, 第三人称射击) 游戏或者 FPS (First-Person Shooter, 第一人称射击) 游戏。当终端 110 运行应用程序 111 时, 终端 110 的屏幕上显示应用程序 111 的用户界面。用户使用终端 110 控制位于虚拟环境中的虚拟角色进行活动, 或者由终端控制位于虚拟环境中的 NPC (Non-Player Character, 非玩家控制角色) 进行活动。虚拟角色的活动包括但不限于: 调整身体姿态、爬行、步行、奔跑、骑行、飞行、跳跃、驾驶、拾取、射击、攻击、投掷、释放技能中的至少一种。

第一服务器 120 可以是独立的物理服务器, 也可以是多个物理服务器构成的服务器集群或者分布式系统, 还可以是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、内容分发网络 (Content Delivery Network, CDN)、以及大数据和人工智能平台等基础云计算服务的云服务器。本申请实施例中, 第一服务器 120 用于为支持三维虚拟环境的应用程序提供后台服务, 可以为 DS (Dedicated Server, 专用服务器)。可选地, 第一服务器 120 承担主要计算工作, 终端承担次要计算工作; 或者, 第一服务器 120 承担次要计算工作, 终端承担主要计算工作; 或者, 第一服务器 120 和终端之间采用分布式计算架构进行协同计算。

第二服务器 130 可以是独立的物理服务器, 也可以是多个物理服务器构成的服务器集群或者分布式系统, 还可以是提供云服务、云数据库、云计算、云函数、云存储、网络服务、云通信、中间件服务、域名服务、安全服务、内容分发网络 (Content Delivery Network, CDN)、以及大数据和人工智能平台等基础云计算服务的云服务器。本申请实施例中, 第二服务器 130 用于支持三维虚拟环境的应用程序提供虚拟角色的动作决策服务。在第二服务器 130 提供的动作决策服务开启时, 在 ETCD (et-see-dee, 分布式键值对存储) 中注册自身地址。在具有虚拟角色的动作决策需求的情况下, 第一服务器 120 向服务调度请求连接推理服务, 并且获取服务调度返回的动作决策服务的地址。从而第一服务器 120 通过得到的地址向第二服务器 130 提供的动作决策服务发送动作决策请求, 在推理服务确定角色动作后, 向第一服务器 120 返回动作指令。

在一些实施例中, 通过动作决策模型执行虚拟角色的动作决策方法, 在训练动作决策模型的过程中, 第二服务器 130 基于第一服务器 120 发送的动作决策请求中的角色状态信息以及多模态的环境感知信息, 通过强化学习的方式对动作决策模型进行训练。

本申请实施例可以应用在虚拟环境中第一虚拟角色的导航场景、角色动作完成设定任务的场景或者第一虚拟角色与其他虚拟角色的对局场景中, 本实施例对此不进行限定。下面对本申请实施例在虚拟场景中的应用过程进行示意性的说明。

本申请实施例提供的虚拟角色的动作决策方法应用于游戏场景中, 当游戏场景中要进行第一虚拟角色的动作决策的情况下, 通过终端 110 或者第一服务器 120 获取第一虚拟角色的角色状态信息以及第一虚拟角色对当前所处虚拟环境的多模态的环境感知信息, 并通过第二

服务器 130 基于角色状态信息以及环境感知信息确定出第一虚拟角色所要执行的角色动作。若由动作决策模型执行虚拟角色的动作决策方法，则在动作决策模型的训练过程中，获取第一虚拟角色的样本角色状态信息以及样本环境感知信息，并将样本角色状态信息以及样本环境感知信息输入动作决策模块，以对动作决策模型进行训练。

需要说明的是，本申请实施例还可以应用于多种人工智能场景下。上述实施环境中仅作为示意性的例子，并不对本申请实施例的应用场景构成限定。为了方便表述，下述各个实施例以虚拟场景中虚拟角色的动作决策方法由计算机设备执行为例进行说明。

请参考图 2，其示出了本申请一个示例性实施例提供的虚拟角色的动作决策方法的流程图。本实施例以该方法用于计算机为例进行说明，该方法包括如下步骤：

步骤 201，获取第一虚拟角色的角色状态信息，以及第一虚拟角色对当前所处虚拟环境的环境感知信息。

示意性的，第一虚拟角色是虚拟环境中的任意一个虚拟角色。可选地，第一虚拟角色是玩家控制的主控虚拟角色，如虚拟英雄、虚拟坐骑等；或者，第一虚拟角色是系统自动控制的自动虚拟角色，如在玩家挂机状态下自动控制执行的虚拟角色等。

其中，环境感知信息包括多个模态的感知信息，多个模态分别代表一个信息维度。也因此，环境感知信息还可以称为多模态环境感知信息，多模态环境感知信息中不同模态的感知信息具有不同的信息维度。

在游戏场景中，计算机设备提取游戏场景中的有用信息，该有用信息至少包括第一虚拟角色对应的角色状态信息以及环境感知信息。角色状态信息是指第一虚拟角色当前的属性信息、第一虚拟角色与其他虚拟角色交互产生的交互信息以及第一虚拟角色与环境交互的交互信息。能够反应第一虚拟角色的当前状态。

角色状态信息为一维向量，即为向量特征，在游戏场景中角色状态信息可以从游戏引擎接口直接获取，此外一部分角色状态信息也可以通过射线探测的方式获取。

在游戏场景中，游戏运行时每秒所运行的帧数称为游戏帧，在每个游戏帧运行过程中，计算机设备均会获取当前角色状态信息以及当前环境感知信息，或者，计算机设备基于一定的周期间隔获取当前角色状态信息以及当前环境感知信息。游戏帧数越大，则界面显示越平滑，游戏帧数可以为 36，则计算机设备每秒获取 36 帧角色状态信息以及当前环境感知信息。而在帧数较大的情况下，例如帧数为 72，对于计算机设备性能要求较高，则此时可以设定周期为 3，即每三个游戏帧获取一帧信息，每秒获取 24 帧角色状态信息以及当前环境感知信息。

示意性的，请参考表 1，其示出了本申请一个示例性实施例提供的角色状态信息的分类。

表 1

<<属性信息>>	<<交互信息>>
<对手虚拟角色共有>	<第一虚拟角色与对手虚拟角色之间>
血量	是否收到伤害
位置	是否视野内可见
朝向	是否躲避攻击
当前虚拟道具数	欧氏距离
阵营	寻路距离
是否进攻	基于第一虚拟角色朝向×距离相对坐标
是否静步	攻击距离
是否移动	<第一虚拟角色与掩体之间>
是否冲刺	是否在掩体中
是否切换道具	掩体是否安全
是否蹲伏	寻路距离
是否跳跃	欧氏距离

是否侧身	掩体相对坐标
是否趴下	<第一虚拟角色与障碍物之间>
是否架枪	是否存在射线遮挡
是否出局	相对距离

在环境感知信息中，不同模态的感知信息的获取方式可能不同，并且，在动作决策模型中对不同模态的感知信息的获取方式也不同。环境感知信息用于获取到更加全面立体的虚拟环境特征。可选的，计算机设备可以在每一帧游戏画面显示时获取一次虚拟角色的角色状态信息以及环境感知信息。或者，计算机设备也可以在每N帧游戏画面显示结束后获取一次角色状态信息以及环境感知信息。获取特征的频率较快能够在进行动作预测时获得更丰富的信息，使得角色动作的确定更加准确；获取特征的频率较慢，能够避免第一虚拟角色执行第一角色动作未结束的情况下接收到新的角色动作，执行第一角色动作被打断，可能控制虚拟角色执行动作出现错误。例如，计算机设备可以获取每一帧的角色状态信息以及环境感知信息，若通过动作决策模型指定动作决策过程，则可以将第4帧获取到的角色状态信息以及环境感知信息输入动作决策模型。

步骤202，融合角色状态信息和环境感知信息分别对应的编码结果，得到融合特征。

示意性的，特征提取角色状态信息对应的编码结果，以及特征提取环境感知信息对应的编码结果，融合两个编码结果可以得到融合特征。其中，编码结果是通过不同编码方式对角色状态信息以及多个感知信息分别进行特征编码得到的结果。

可选地，通过编码网络（或称编码器）执行特征提取过程，将角色状态信息和环境感知信息输入至相同的编码网络中，但是采用不同的编码方式对角色状态信息和环境感知信息分别进行分析，得到与角色状态信息和环境感知信息分别对应的编码结果。编码方式包括卷积编码、循环编码、自编码器编码、注意力编码等多种方式中的至少一种，不同的编码方式包括上述编码方式的单独应用或者组合应用。

可选地，通过编码网络执行特征提取过程，将角色状态信息和环境感知信息分别输入至采用不同编码方式的不同编码网络中，得到与角色状态信息和环境感知信息分别对应的编码结果。可选地，考虑到环境感知信息中包括多个模态的感知信息，多个模态分别代表一个信息维度，信息维度用于表示衡量信息的维度。示意性的，不同的信息来源对应不同的信息维度；或者，不同的信息获取方式对应不同的信息维度等。因此在对环境感知信息进行编码处理时，对多个模态的感知信息分别进行编码处理。例如：在通过编码网络执行特征提取的过程中，将角色状态信息以及多个模态的感知信息分别输入至采用不同编码方式的不同编码网络中，输出得到与角色状态信息对应的编码结果以及与多个模态的感知信息分别对应的编码结果，多个模态的感知信息的编码结果可以统称为环境感知信息的编码结果等。

在一些实施例中，通过动作决策模型执行动作决策过程，动作决策模型中包含编码网络以及决策网络。编码网络用于对输入的信息进行编码，而决策网络用于确定角色动作。由于不同模态的感知信息具有不同的信息维度，因此，编码网络中应当包含不同维度的卷积核，从而通过不同维度的卷积核对不同模态的感知信息分别进行编码。

机器学习算法都是在矩阵上执行线性代数计算，所以参加计算的特征必须是数值型的，而输入的角色状态信息以及多模态的环境感知信息中存在非数值型的信息，因此对于非数值型的信息需要进行编码处理。例如上述表中的角色状态，可以采用标签编码或者独热编码的方式将其转换为数值型的特征。

在分别对角色状态信息以及环境感知信息进行编码后，得到两种编码结果，在对两种编码结果进行特征拼接后，再基于卷积神经网络进行特征提取，得到融合特征。其中卷积神经网络可以为长短期记忆网络。通过特征提取能够减少特征数据的维度，并且能够从特征中提取出更适合动作决策的信息。

步骤203，基于融合特征对第一虚拟角色进行得到角色动作。

可选地，在基于融合特征确定角色动作时，通过从候选动作中选择当前场景下适宜执行的最佳动作的方式实现该过程。示意性的，预先给定多个候选动作，基于融合特征从多个候选动作中选择至少一个候选动作作为角色动作。例如：提取多个候选动作分别对应的候选动作特征，比较融合特征与多个候选动作特征之间的特征相似度（如确定特征之间的余弦值），并将特征相似度最大的至少一个候选动作作为角色动作。

在一些实施例中，若通过动作决策模型执行动作决策过程，可以在动作决策模型的训练过程中，根据融合特征对动作决策模型中用于进行动作决策的决策网络不断调整模型参数，从而学习到在各类场景下的最佳决策，即学习到在各类场景下最佳的动作。

可选地，决策网络是动作决策模型中的网络层，用于基于策略确定角色动作，策略定义了智能体对于给定状态所作出的行为，即定义了第一虚拟角色对于给定的角色状态信息以及环境感知信息所作出的行为。对动作决策模型的训练过程可以理解为：基于确定的预估角色动作所带来的奖励调整决策网络中策略的过程。

示意性的，在通过决策网络进行动作决策的过程中，首先基于融合特征确定各个动作类型中不同子动作之间的概率分布，随后，再基于概率分布确定各个动作类型中的对象子动作。

在一种可能的实施方式中，在游戏场景下，开发人员会为部分子动作设置冷却时间，在冷却时间未结束的情况下，该子动作的冷却状态也作为第一虚拟角色的角色状态信息输入动作决策模型，从而降低决策网络确定的概率分布中该子动作的执行概率，避免出现决策模型确定出的对象子动作无法执行的情况。

步骤 204，控制第一虚拟角色执行角色动作。

示意性的，基于融合特征确定的角色动作是控制第一虚拟角色在虚拟环境中执行的动作。可选地，角色动作包括行走动作、奔跑动作、跳跃动作、匍匐前进动作、屈膝动作、弯腰动作、跨栏动作、摇头动作等多种动作中的至少一种。也即：角色动作包括第一虚拟角色身体各个部位的整体运动，也包括至少一个部位的部分运动，此处对于角色动作的动作形式不加以限定。

在一些实施例中，以通过动作决策模型执行动作决策过程并确定角色动作为例，动作决策模型部署于第二服务器中，在通过动作决策模型确定角色动作的情况下，将动作指令发送至为支持虚拟环境的应用程序提供后台服务的第一服务器，再由第一服务器将该动作执行指令发送至客户端以实现控制第一虚拟角色执行角色动作的目的。

在一种可能的实施方式中，计算机设备对不同动作类型中的各个子动作名称分别进行编码，得到多个候选动作标签，计算机设备基于动作决策模型中的决策网络得到各动作类型下各子动作的概率分布，基于动作执行概率从子动作标签中确定对象子动作对应标签，在确定对象子动作标签的情况下，计算机设备对对象子动作标签进行解码，得到对象子动作，并控制第一虚拟角色执行由多个对象子动作构成的角色动作。

综上所述，在虚拟环境中需要对第一虚拟角色进行动作决策时，获取第一虚拟角色的角色状态信息以及多模态的环境感知信息，从而对角色状态信息以及多模态的环境感知信息进行特征编码融合，最终基于融合特征进行动作决策。在进行动作决策时获取第一虚拟角色感知到的不同信息维度的环境感知信息，有助于获取到第一虚拟角色周围更加全面、细致的三维虚拟环境，从而能够进行更加合理的动作决策。采用不同的编码方式对角色状态信息以及多个不同模态的感知信息进行特征编码，能够使得动作决策模型对当前场景下的各类信息拟合更加精准。此外，相较于相关技术中采用图像识别的方式获取环境特征的方式，通过对量化的多模态的环境感知信息进行特征提取进而融合角色状态信息的编码结果，有助于在量化环境感知信息的同时更丰富地考虑第一虚拟角色本身的角色状态信息，在节省计算资源的同时提高决策的角色动作的准确性，使得第一虚拟角色执行角色动作的效果更加逼真。

在本申请实施例中，环境感知信息包括第一模态感知信息、第二模态感知信息以及第三模态感知信息，对于不同模态的感知信息，其获取方式也不尽相同。下面将对不同模态的感

知信息的获取方式进行说明。

一、获取第一模态感知信息。

第一模态感知信息为一维向量信息，第一模态感知信息用于表征第一虚拟角色周围同一水平高度上障碍物的深度情况。在一种可能的实施方式中，基于第一虚拟角色在虚拟环境中的角色位置，通过射线探测获取第一模态感知信息。

示意性的，角色位置用于表示第一虚拟角色在虚拟环境中所处位置的位置信息。可选地，角色位置通过第一虚拟角色在虚拟环境对应的世界坐标系中的位置坐标予以表示；或者，角色位置通过第一虚拟角色在其他虚拟角色对应的局部坐标系中的位置坐标予以表示等。

可选地，以第一虚拟角色所处的角色位置为起始点，向多个方向发射环境感知射线，环境感知射线是用来检测和测量虚拟环境中物体和场景的射线，环境感知射线被用来模拟光线、传感器、雷达等设备的工作原理，从而实现虚拟环境中的物体检测、碰撞检测、视线检测等功能，环境感知射线的发射通过起始点和发射方向共同确定。如起始点为角色位置，发射方向为虚拟环境的各个方向。

可选地，环境感知射线在碰撞到障碍物表面的情况下，会沿环境感知射线的发现方向反射。进而计算机设备可以获取环境感知射线的反射情况，从而根据环境感知射线的反射情况生成第一模态感知信息。其中，不同环境感知射线位于同一水平高度。

进一步的，为更加准确的获取第一虚拟角色周围虚拟环境的第一模态的感知信息，计算机设备以第一虚拟角色所处位置的至少两种高度为起始点，向多个方向发射环境感知射线。并且根据环境感知射线的反射情况生成第一模态感知信息。其中，相同高度发射的不同环境感知射线的发射方向位于同一水平高度。例如，以第一虚拟角色所处位置的脚部、腰部以及头部作为起始点，向多个方向发射环境感知射线，从而获取三种不同高度的第一模态感知信息。其中，多个不同高度分别发射的环境感知射线为环形射线，环形射线是以第一虚拟角色自身为起点环绕自身周边向外打出的若干条射线，综合多个不同高度得到的多个环形射线可以统称为多层环形射线（如三个高度分别发射的环形射线统称为三层环形射线），借助多层环形射线可以避免单一环形射线的局限性问题，有助于更加丰富立体地表征周边地形信息。

示意性的，图3示出了本申请一个示例性实施例提供的射线探测方案的示意图，计算机设备分别以第一虚拟角色的头部、腰部以及腿部为起始点，向多个方向发射环境感知射线，在环境感知射线碰撞到虚拟环境中障碍物的情况下发生反射，从而获取三种不同高度的环境感知射线的反射情况，并生成第一模态感知信息。

在一种可能的实施方式中，以不同高度为起始点发射的环境感知射线的数量相同，即每一层环形射线的数量相同。不同高度的环形射线的反射情况能够表征不同高度下虚拟环境中障碍物与第一虚拟角色的距离。每一层中的环境射线数量越多，则对于第一虚拟角色周围障碍物的感知越精细，对于计算机设备的性能需求越高。

本申请实施例中，采用至少两种高度的环境感知射线进行射线探测，以获取第一模态感知信息，有利于获取周围环境在不同高度的环境特征，提高环境特征的全面性。同时，也能够使动作决策模型在进行动作决策的过程中充分考虑当前虚拟环境的高度特征，有助于使得第一虚拟角色执行角色动作后表现出的行为更加的拟人化，即更加符合人类在物理环境中的行为逻辑，例如，可以使第一虚拟角色展现出合理上下楼梯等拟人化行为，提高角色动画的生动性。

二、获取第二模态感知信息。

第二模态感知信息为第一虚拟角色朝向方向的二维深度图，第二模态感知信息用于表征第一虚拟角色朝向方向上障碍物的深度情况。

在一种可能的实施方式中，计算机设备基于第一虚拟角色在虚拟环境中的角色位置以及第一虚拟角色的朝向，通过射线探测获取第二模态感知信息。

示意性的，朝向用于表示第一虚拟角色在虚拟环境中面对的方向。例如：第一虚拟角色

面朝东方则代表朝向表示为 0° ，第一虚拟角色面朝南方则代表朝向表示为 90° 等。

可选地，以第一虚拟角色为起始点，向第一虚拟角色的朝向方向发射环境感知射线。在环境感知射线碰撞到障碍物表面的情况下，沿环境感知射线的法线方向反射，从而使得计算机设备能够根据环境感知射线的反射情况，生成第二模态感知信息。

示意性的，以第一虚拟角色为起始点，向第一虚拟角色的朝向方向发射单一的环境感知射线；或者，以第一虚拟角色为起始点，向第一虚拟角色的朝向方向发射多条环境感知射线。

示意性的，请参考图 4，其示出了本申请一个示例性实施例提供的第二模态感知信息的示意图，虚拟环境的示意图与二维深度图相对应。第二模态感知信息为当前虚拟角色朝向方向上的二维深度图，深度图中的像素颜色越深，则表明该像素对应的障碍物与第一虚拟角色的距离越近；深度图的像素颜色越浅，则表明该像素对应的障碍物与第一虚拟角色的距离越远。

二维深度图所表征虚拟环境的精细程度与分辨率呈正相关关系，二维深度图的采样分辨率越大，则二维深度图表征的第一虚拟角色前方障碍物的深度信息越精细。

本申请实施例中，在环境感知射线具有反射特性的情况下，通过向朝向方向发射环境感知射线的方式获取第一虚拟角色朝向方向上的深度图作为第二模态感知信息，有利于增强第一虚拟角色在朝向方向上的环境感知能力，从而有助于借助第二模态感知信息显著提升场景理解、物体检测、视觉效果等方面的优势，实现更准确的距离感知、环境映射等环境感知效果，充分提升第一虚拟角色的前方视野范围内的环境特征的准确性，以便提供更为立体且具有方向针对性的环境特征。

三、获取第三模态感知信息。

第三模态感知信息为三维占栅网格图，第三模态感知信息用于表征预设范围内障碍物的空间分布情况。占栅网格图是用于描述整张地图障碍物分布概率的分布网格图，通过预先给定的占栅网格将虚拟环境空间划分为多个栅格，占栅网格由多个立方体组成，多个立方体与多个栅格一一对应，栅格可以理解为被划分之后的部分虚拟环境；检测每个栅格中被障碍物占据的比例，所有栅格构成一个网，其本质是一张占栅网格图，图中每个像素表征实际环境中存在障碍物的概率分布。

在大规模的三维虚拟环境中，占栅网格具有对虚拟环境的宏观感知能力。可选的，计算机设备基于第一虚拟角色在虚拟环境中所处位置，通过射线探测获取在第一虚拟角色的预设范围内虚拟环境的第三模态感知信息。其中，预设范围是基于第一虚拟角色的角色位置而预先设定的范围。可选地，预设范围为第一虚拟角色周侧 1 米的圆形区域；或者，预设范围为第一虚拟角色周侧 5 米的矩形区域等。

可选地，首先，采用占栅网格划分第一虚拟角色预设范围内的虚拟环境。其中，占栅网格由多个体积相同的立方体组成。随后，从虚拟环境中至少两个方向的起始点向占栅网格内发射环境感知射线，从而使计算机设备基于环境感知射线的反射情况，确定在每一个栅格中障碍物的占据比例，进而生成占栅网格图。

通过占栅网格图可以进一步细化第一虚拟角色周围的宏观虚拟环境特征。对于相同范围的虚拟环境，划分得到的栅格越多，则占栅网格图所表征的虚拟环境信息越细节。此外，向每个栅格发射的环境感知数量越多，则占栅网格图所表征的虚拟环境信息越细节。

示意性的，请参考图 5，其示出了本申请一个示例性实施例提供的占栅网格结构的示意图，其中，不同栅格对应的立方体的体积一致。

本申请实施例中，通过预设的占栅网格对虚拟环境进行划分，并将划分得到的多个栅格组成作为第三模态感知信息的占栅网格图，充分发挥占栅网格图简单而直观的建模优势，能够更好地表征第一虚拟角色周围宏观的环境特征，占栅网格图的数据结果通常实现为三维数组形式，因此访问和分析占栅网格图的效果也更高，不仅可以预先获取第三模态感知信息，还可以实时获取第三模态感知信息；此外通过控制占栅网格的细致程度还可以进一步细化宏

观环境特征，从而表征更加细节的障碍物分布情况。

上述介绍了获取第一模态环境信息、第二模态环境信息以及第三模态环境信息的方式，虽然不同模态的环境信息的获取依赖于第一虚拟角色在虚拟环境中的角色位置，但是由于射线探测时获取的维度差异，能够从一维、二维以及三维情况下对虚拟环境进行更全面地分析，提高了环境感知信息的获取全面性，有助于提高虚拟环境中的角色感知能力，提升虚拟环境交互效果。

在一个可选的实施例中，通过动作决策模型执行虚拟角色的动作决策方法，动作决策模型中包含编码网络以决策网络，编码网络用于对角色状态信息和环境感知信息进行特征编码及融合，而决策网络用于基于融合特征确定角色动作。并且，编码网络中包含感知信息编码器以及状态信息编码器。在决策网络中包含 n 个动作输出头， n 个动作输出头用于串行输出 n 个对象子动作，不同动作输出头对应不同动作类型，且 n 个动作输出头基于动作类型之间的依赖关系串行连接，依赖关系用于表征不同动作类型下子动作之间的依赖限制情况， n 个对象子动作构成角色动作。

在动作决策模型中， n 个动作输出头属于动作决策模型的输出层，用于对不同维度下的动作进行预测，从而输出不同维度的动作参数。其中，第一虚拟角色支持执行的动作维度可能包括移动状态维度、移动方向维度、朝向维度、进攻维度以及姿态维度等等。动作决策模型中， n 个动作输出头分别对应 n 个动作维度，则 n 个动作输出头输出 n 个不同维度的动作参数，例如，在移动方向维度对应的动作输出头输出的动作参数为一个移动方向，如 90° 方向；朝向维度对应的动作输出头输出的动作参数为能表征方向的转向角度，如 -5° ，表征向逆时针方向旋转 5° ；进攻维度对应的动作输出头输出的动作参数为进攻或者不进攻，等等。本实施例对于动作维度划分以及不同动作维度对应的动作输出头输出的动作参数不进行限定。

下面将通过一个示例性实施例对通过动作决策模型生成角色动作的过程进行说明。

请参考图 6，其示出了本申请一个示例性实施例提供的生成角色动作的过程的流程图。

步骤 601，将角色状态信息以及环境感知信息输入动作决策模型，通过编码网络编码角色状态信息以及环境感知信息，得到角色状态信息对应的状态信息编码结果和环境感知信息对应的感知信息编码结果。

可选的，通过状态信息编码器编码角色状态信息，得到状态信息编码结果；通过感知信息编码器编码环境感知信息，得到感知信息编码结果。其中，两个编码过程可以先后执行（如先通过状态信息编码器编码，再通过感知信息编码器编码；或者，先通过感知信息编码器编码，再通过状态信息编码器编码），也可以同时执行，此处不加以限定。

可选地，在感知信息编码器中包含不同维度的卷积核，不同维度的卷积核用于对不同模态的感知信息进行特征编码。可选的，通过感知信息编码器中一维卷积核对第一模态感知信息进行编码，得到第一感知信息编码结果。为使动作决策模型感知到第一虚拟角色周围立体的虚拟环境信息，采用循环卷积的方式基于每一条环境感知射线所表征的障碍物与第一虚拟角色之间的距离，获取第一虚拟角色周围虚拟环境的环境感知信息。

示意性的，图 7 示出了本申请一个示例性实施例提供的对环境感知信息进行一维卷积的过程示意图。其中共包含三种不同高度的环境感知射线，每一层环形射线包括 144 条环境感知射线，每一层的环境感知射线中存在一条环境感知射线作为首部，一条环境感知射线作为尾部，由于是环形射线具有首尾相连的特性。则采用一维卷积核对三层环形射线中对应的环境感知射线进行循环卷积计算，从而实现对第一模态感知信息的编码。

此外，在训练训练动作模型的过程中，通过一维循环卷积对三种不同高度的环境感知射线进行特征编码，能够避免遗漏具有重要价值的信息，且能够加速训练过程，提高训练效果。

可选的，通过感知信息编码器中的二维卷积核对第二模态感知信息进行编码，得到第二感知信息编码结果。二维卷积核是一个二维矩阵，在通过二维卷积核对第二模态感知信息进

行编码的过程中, 通过将该二维矩阵和输入的二维深度图进行逐元素乘法以及求和计算的方式, 完成对第二模态感知信息的特征编码。

示意性的, 对逐元素乘法以及求和计算的过程予以说明。假设有一个二维矩阵 A 和一个相同尺寸的二维深度图, 对两者的元素在相同位置上进行如下操作:

(1) 逐元素乘法 (Element-wise Multiplication):

在逐元素乘法中, 对应元素位置的元素相乘, 得到一个新的矩阵 C, 该矩阵 C 中的每个元素都是矩阵 A 和深度图对应位置元素的乘积。

(2) 求和计算 (Summation):

将矩阵 C 中所有元素相加, 得到一个标量结果 S, 表示所有乘积的总和, 即逐元素乘法以及求和计算的计算结果。

可选的, 通过感知信息编码器中三维卷积核对第三模态感知信息进行编码, 得到第三感知信息编码结果, 第一感知信息编码结果、第二感知信息编码结果以及第三感知信息编码结果构成感知信息编码结果。通过三维卷积核对占栅网格图进行特征提取, 可以为动作决策模型中的决策网络提供宏观立体的场景感知信息。例如, 第一虚拟角色在游戏场景中存在预设任务, 为完成该预设任务, 需要第一虚拟角色在大范围的地图中进行移动, 在这种情况下, 为训练用于该场景中的动作决策模型, 通过三维卷积核对第三模态感知信息进行编码, 有利于向动作决策模型提供宏观场景的感知信息。

上述内容介绍了通过不同维度的卷积核编码不同模态感知信息并得到对应感知信息编码结果的过程。设定不同维度的卷积核有利于区别化分析不同模态的感知信息, 也有利于针对性地对相同维度下的模态的感知信息进行独特分析, 从而可以更好地保留和利用每种模态的特征; 综合不同卷积核编码多个模态的感知信息, 能够有效地综合感知信息, 以使得环境感知信息表达更强的整体感知能力和环境理解能力, 避免因为单一模态下因为噪音、遮挡等恶劣条件造成的不可靠问题, 有助于在复杂的虚拟环境中仍然获取较为准确且适应性较强的环境感知信息。

步骤 602, 拼接状态信息编码结果以及感知信息编码结果, 得到拼接特征编码。

在获取到状态信息编码结果、第一环境感知信息编码结果、第二环境感知信息编码结果以及第三环境感知信息编码结果之后, 通过全连接层以及激活函数对状态信息编码结果以及感知信息编码结果进行特征拼接, 得到拼接特征编码。

示意性的, 依照状态信息编码结果、第一环境感知信息编码结果、第二环境感知信息编码结果以及第三环境感知信息编码结果的拼接顺序, 执行特征拼接过程, 得到拼接特征编码; 或者, 依照第一环境感知信息编码结果、第二环境感知信息编码结果以及第三环境感知信息编码结果、状态信息编码结果的拼接顺序, 执行特征拼接过程等, 此处对于拼接顺序不加以限定。全连接层用于将前层卷积计算得到的特征空间映射到样本标记空间 (即将特征表示整合为一个值), 能够减少特征位置对于分类结果的影响, 提高动作决策模型的鲁棒性。激活函数用于加入非线性因素, 从而提高动作决策模型的表达能力。

步骤 603, 通过编码网络特征提取拼接特征编码, 得到融合特征。

在一种可能的实施方式中, 编码网络中还包括 LSTM (Long Short Term Memory, 长短期记忆) 网络。通过将融合特征输入 LSTM 网络, 从而进行特征提取, 以获取到融合特征。在进行特征提取的过程中, 长短期记忆网络会保留具有更重要价值的特征, 并将价值较低的特征遗忘掉, 从而实现特征提取。

可选的, 该 LSTM 网络可以通过其他卷积神经网络进行替换, 本实施例对此不进行限定。

上述步骤 601 至步骤 603 中介绍了通过动作决策模型中的编码网络获取融合特征的内容。通过编码网络中的状态信息编码器和感知信息编码器分别编码角色状态信息和环境感知信息, 有助于利用不同编码器的针对性分析特性, 在提高角色状态信息精准、细化角色状态的分析粒度的同时, 更针对性地捕捉虚拟环境中的各种感知信息, 实现较好地资源管理和任务

分配，采用更高效地方式获取更准确的融合特征。

步骤 604，将融合特征输入决策网络中的第一动作输出头，通过决策网络中第一动作输出头从第一动作类型的子动作中确定第一对象子动作。

将融合特征输入决策网络，即为将融合特征输入决策网络的第一动作输出头，通过决策网络中第一动作输出头从第一动作类型中的子动作中确定第一对象子动作。

可选的，将第一虚拟角色的可执行动作进行正交分解，得到 n 个动作类型，不同动作类型中包含至少两个可执行动作，不同动作类型下的子动作可独立控制，即可以控制第一虚拟角色仅执行一个动作类型中的子动作。其中，正交分解用于将第一虚拟角色的可执行动作划分为多个离散化动作类型，正交分解后得到的 n 个动作类型互为正交动作类型，不同正交动作类型中的子动作之间不存在交集，即不存在一个子动作属于两个动作类型的情况，不同正交动作类型支持同时执行，且同一正交动作类型中的不同子动作不支持同时执行。

在游戏场景中，存在控制第一虚拟角色的最基本，最细粒的原子动作，这类动作不可进一步进行拆分细化。在一个动作类型中的子动作即为原子动作，在同一动作类型中的不同子动作之间不能够同时执行。例如，第一虚拟角色可执行动作中包括匍匐前进这一个动作，则可将这个动作划为两个原子动作分别为匍匐以及前进，其中匍匐对应姿态动作类型，而前进对应移动方向动作类型。再例如，第一虚拟角色可执行动作中包括原地探头射击这一动作，则可将这个动作划分为停留在原地、探头以及射击这三个原子动作，其中停留在原地对应移动状态动作类型，探头对应探头动作类型，射击对应进攻动作类型。

示意性的，请参考图 8，其示出了本申请一个示例性实施例提供的提供虚拟环境的应用程序的用户界面示意图。其中包含移动状态控件 801、进攻控件 802、姿态调整控件 803、水平转向控件、竖直转向控件 804、左右探头控件 805 以及移动状态控件 806，不同控件可触发控制虚拟角色执行不同动作。

对上述图中示出的各个控件对应的动作进行正交分解，能够得到多种动作类型，在一些实施例中，不同动作类型之间可能存在依赖关系，例如，在需要进行攻击时才会进行站立或侧身，则角色动作的姿态动作决策需要依赖于进攻动作类型的动作决策。并且，依赖限制情况指示不同动作类型中的子动作之间可能存在依赖限制关系，例如，在第一虚拟角色进行冲刺动作时不支持同时执行进攻动作，则冲刺动作与进攻动作之间存在依赖限制关系。因此，为使动作决策模型在进行动作决策时能够考虑到不同动作之间的依赖关系，以及不同动作类型下不同子动作之间的依赖限制情况，则将决策网络中的 n 个动作输出头基于动作类型之间的依赖关系串行连接。

可选的，在确定不同动作输出头的连接顺序时，根据不同动作类型对于第一虚拟角色所要完成任务的关联程度进行排序。例如，在第一虚拟角色的任务为击败敌人，则可以确定进攻动作类型的关联程度较高，将其对应的动作输出头排列在靠前的位置；而侧身动作类型的关联程度显然较低，则将侧身动作类型对应的动作输出头排列在最后的位置。

示意性的，表 2 示出了本申请一个示例性实施例提供的正交分解后动作输出头与动作类型之间的对应关系。

表 2

动作输出头名称	动作维度（包含子动作的数目）	动作类型
Yaw_to_focus	7	粗粒度水平朝向
bFire	2	是否进攻
Yaw_to_aim	21	细粒度水平朝向
Pitch_to_aim	15	细粒度竖直朝向
Move_type	4	移动方式
Yaw_to_move	8	移动方向
Posture_type	4	站姿

Lean_type	3	是否在侧身
-----------	---	-------

在一些实施例中，具有第一虚拟角色瞄准其他虚拟角色的需求，其他虚拟角色可能作为第一虚拟角色的对手，即为第一虚拟角色具有瞄准对手的需求的情况下，若采取在转向范围内均匀分布的转向设计方案则在转角粒度过大或过小的情况下，都难以使第一虚拟角色快速精准的瞄准敌人。因此，可以采用粗粒度转向角度以及细粒度转向角度结合的方式实现精确瞄准。粗粒度转向角度的子动作对应角度间隔为第一间隔，细粒度转向角度的子动作对应角度间隔为第二间隔，第一间隔与粗粒度水平朝向的动作维度的乘积为第一虚拟角色的当前视野范围，当前视野范围根据第一虚拟角色的移动方向而改变。第二间隔与动作维数的乘积等于第一动作间隔。

请参考图 9，其示出了本申请一个示例性实施例提供的移动方向以及描述动作的子动作组成示意图，其中，移动方向动作类型被划分为八种不同角度，以第一虚拟角色当前朝向的中心区域为基准，在 360° 范围内分布。而转向角度包括水平转向角度以及竖直转向角度，其中水平转向角度包括粗粒度水平转向角度以及细粒度的水平转向角度。粗粒度的水平转向角度是以第一虚拟角色当前朝向为基准，当前朝向设置为 0°，转向范围为 120 度，并且粗粒度转向角度的每个子动作之间间隔为 20°，即粗粒度转向角度中包含子动作对应的转向角度分别为 0°，20°，40°，60°，-20°，-40°，-60°，则在确定粗粒度转向角度的对象子动作对应 40° 的情况下，执行该转向动作后，第一虚拟角色的粗粒度水平朝向为 40°。细粒度转向角度以第一虚拟角色预执行粗粒度的子动作后的预估朝向为基准，即为 0°，在设定范围内均匀分布，设定范围可以为 20，并且细粒度转向角度的每个子动作之间间隔可以为 3°，则细粒度转向角度中包含子动作对应的转向角度分别为 3°，6°，9°，...，21°，则确定细粒度转向角度的对象子动作对应向右转动 6°。而对于竖直方向的朝向范围较小，则可以都采用细粒度的转向角度进行子动作的划分，以虚拟角色当前朝向为基准，即为 0°，设定范围为-30°-45°，子动作之间的间隔可设置为 3°。

可选的，顺时针旋转角度为正值，逆时针旋转角度为负值，向上旋转角度为正值，向下旋转角度为负值，或者与其相反，本实施例对此不进行限定。

步骤 605，将已确定的第一对象子动作至第 i-1 对象子动作对应的嵌入编码向量，以及融合特征输入第 i 动作输出头，通过第 i 动作输出头从第 i 动作类型的子动作中确定第 i 对象子动作。

其中，n 为大于 2 的正整数，i 为大于 1 的正整数。

n 个动作输出头通过自回归嵌入层连接，由于不同动作类型之间存在依赖关系，则在前序输出头确定对象子动作后，应当基于自回归嵌入层将已确定的至少一个对象子动作以及融合向量输入到后序的动作输出头中，以使得后序动作输出头根据依赖关系确定对象子动作。

自回归是指在统计上处理时间顺序的方法，用同一变量之前各个时刻预测当前时刻该变量的表现，自回归具有时间序列的自相关性。嵌入 (embedding) 的本质是进行数据压缩，用较低维度的特征去表述具有冗余信息的较高维度的特征。

请参考图 10，其示出了本申请一个示例性实施例提供的 n 个动作输出头的自回归结构的示意图。在一个动作输出头中包含全连接层、Logits 层、采样层以及嵌入层。Logits 层即为确定的动作类型中各子动作之间的执行概率分布，从而确定出对象子动作，在确定对象子动作后，通过嵌入层将以确定的子动作传输至下一串行连接的动作输出头，由该动作输出头中的融合层对前序已经确定的各个对象子动作进行融合，从而根据已确定的对象子动作确定新的对象子动作。此外，编码网络也会将融合特征输入至每个动作输出头。

本申请实施例中，基于不同维度的卷积核对不同模态（即不同信息维度）的环境感知信息进行编码，能够使得对各个信息维度的环境信息能以合适的编码方式进行处理，提高环境感知信息进行编码后的特征表达，从而使动作决策模型获取到更加准确的环境特征。

在本申请实施例中，还介绍了通过动作决策模型中的决策网络进行动作决策的内容。借

助决策网络中的动作输出头可以从动作类型的子动作中确定第一对象子动作,通过综合前 $i-1$ 个对象子动作的嵌入编码向量以及融合特征至第 i 动作输出头的方式,有助于更综合地确定第 i 对象子动作,这种结构有助于系统在复杂环境中做出高效的决策和动作计划。它利用融合的多模态信息指导每一步动作的选择,同时通过嵌入编码向量传递先前决策的结果,确保后续动作的连贯性和合理性。

由于不同动作类型下的子动作存在依赖限制关系,存在依赖限制关系的至少两个子动作之间不支持同时执行,则在存在已确定的前序的对象子动作的情况下,可能在后续动作类型中存在部分子动作与已确定的对象子动作之间具有依赖限制关系。并且,由于进行正交分解后的动作空间维度依旧很高,会导致丰富的拟人化原子动作空间组合爆炸,因此,可以根据不同子动作之间组合的合理性(即依赖限制关系)进行子动作的确定,提升动作组合的拟人性。此外,在对动作决策模型进行训练的过程中,也可以采用动作遮罩的方式减少决策网络进行无效的决策策略探索,加速训练效率的同时提升动作决策模型进行动作决策的拟人性。

动作遮罩可以用于强化学习模型(对应本申请实施例中的动作决策模型),可以在大规模决策空间中,屏蔽掉不合理或者无效的子动作集合,从而减少无意义的动作探索,能够使决策网络更快收敛。本质上,动作遮罩是用于屏蔽掉不同动作类型中不合理或者与已确定对象子动作之间存在依赖限制关系的子动作。通过动作遮罩,使得需要遮罩的子动作在动作输出头得到的概率分布中的概率降低,从而在依照概率进行采样确定对象子动作时,难以将遮罩后的子动作确定为对象子动作。通过动作遮罩的方式,能够在进行动作决策过程中直接排除掉不必要探索的子动作分支,从而减小动作空间,在提升动作决策模型探索效率的同时保证动作决策模型的强度。

其中,存在依赖限制关系的至少两个子动作属于不同的动作类型,且存在依赖限制关系的至少两个子动作不能同时执行。可选的,存在第一子动作与第二子动作之间存在依赖限制关系,则这两个子动作不能同时执行,且在已经确定第一子动作为角色动作的情况下,对第二子动作进行动作遮罩。可选的,存在第二子动作结合第三子动作后,与第四子动作之间存在依赖限制关系,则第二子动作与第三子动作的结合,和第四子动作不能同时执行,且在已经确定第二子动作以及第三子动作均为对象子动作的情况下,对第四子动作进行动作遮罩。

可选的,动作决策模型中包含动作遮罩网络,动作遮罩网络与决策网络中 n 个动作输出头相连接,用于对各个输出头进行动作决策过程中各个动作类型中的子动作进行动作遮罩。

在一些实施例中,在确定角色动作之前,计算机设备基于已确定的第一对象子动作至第 $i-1$ 对象子动作,以及不同动作类型中不同子动作之间的依赖限制情况,对第 i 动作输出头对应的第 i 动作类型中的子动作进行动作遮罩。再通过第 i 动作输出头,从第 i 动作类型中未被遮罩的子动作中确定第 i 对象子动作。

可选的,在进行动作遮罩时,在网络原始 logits 动作输出层,通过对不合理子动作的 logits 加上一个绝对值很大的负数,保证该动作对应的 logits 值比所有合理的子动作的 logits 值小,进而重塑了输出动作层的 logits 分布,将不合理动作被采样的概率映射接近于零。

上述内容介绍了决策网络在决策过程中通过动作遮罩和依赖限制来优化子动作的选择过程。通过应用动作遮罩能够根据不同动作类型中子动作之间的依赖关系和限制,排除不合理或无效的子动作选择,有助于确保所选的子动作在逻辑上是连贯和有效的,也可以避免冲突和不一致的动作选择,从而保证整体动作的合理性;有效的动作选择也意味着可以更快速、更有效地执行合适的动作,节省资源并优化整体执行过程。

在进行动作遮罩之前,首先基于已确定的对象子动作,在待决策动作类型中,确定与已确定对象子动作具有依赖限制关系的待遮罩子动作。具体可包括以下三种情况:

一、存在一个已确定的对象子动作与待决策动作类型中待遮罩子动作存在依赖限制关系。

在一种可能的实施方式中,在存在已确定的第 j 对象子动作,且依赖限制情况指示第 j 对象子动作与第 i 动作类型中的至少一个第一子动作存依赖限制关系的情况下,对第 i 动作类

型中的第一子动作进行动作遮罩, j 小于 i (即第 j 动作头串联在第 i 动作头之前)。

例如, 存在已确定的冲刺动作, 且依赖限制情况指示冲刺动作与下蹲动作存在依赖限制关系, 则在确定第一虚拟角色姿态这一动作类型的情况下, 对姿态动作类型中的下蹲动作进行动作遮罩, j 为大于 2 的正整数。

可选的, 计算机设备可以对前序动作输出头所确定的对象子动作进行独热处理, 随后根据用于表征后序动作类型中各个子动作与前序动作类型中的子动作的依赖限制关系的限制矩阵, 与前序对象子动作对应的独热矩阵进行矩阵乘法, 从而获取后序动作类型对应的动作遮罩矩阵。示意性的, 计算机设备基于动作遮罩矩阵将前序动作输出头确定的动作传输至后序动作输出头。

示意性的, 存在动作类型 $a1$ 为发起进攻 (包括进攻和不进攻)、动作类型 $a2$ 为虚拟角色姿态 (包括下蹲、跳跃、匍匐以及站立) 以及动作类型 $a7$ 为虚拟角色的移动状态 (包括静步、快走、冲刺以及原地不动), 动作类型 $a1$ 对应动作输出头 $H1$ 、动作类型 $a2$ 对应动作输出头 $H2$, 动作类型 $a7$ 对应动作输出头 $H7$ 。其中, 进攻与冲刺存在依赖限制关系, 且下蹲也与冲刺存在依赖限制关系, 由于子动作的值在神经网络中以张量的形式流动, 因此不能够直接采样对象子动作来避免梯度断开。

因此, 首先将样本确定为独热 (one-hot) 矩阵 $\tau_{n \times m}$, 假设存在 $n=3$ 个样本, $a1$ 的动作维度 (即包含的子动作数量) $m=2$, 获取到的三个样本的动作输出头 $H1$ 的输出对象子动作为 {进攻, 不进攻, 进攻}, 则这三个样本对应的 $n \times m$ 的 one-hot 矩阵 $\tau_{n \times m}$ 为:

$$\tau_{n \times m} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}_{3 \times 2}$$

$\tau_{n \times m}$ 是样本中动作输出头 $H1$ 输出的独热矩阵, 行数为样本个数 n , 列数代表动作类型 $a2$ 中所包含的子动作数量 m 。

随后, 构造辅助映射矩阵 M , 即为构造能够表征动作类型 $a1$ 中子动作与动作类型 $a7$ 的子动作之间的依赖限制关系的映射矩阵。由于只是基于单个动作输出头 $H1$ 输出的对象子动作采样从而构造对动作类型 $a7$ 的对映射关系, 则动作输出头 $H1$ 所有不同动作组合成的独热矩阵起始是一个 $m \times m$ 的单位矩阵, 第一行映射对象子动作进攻, 第二行映射对象子动作不进攻。

$$E = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}_{2 \times 2}$$

由于动作输出头 $H7$ 对应维度 $P=4$, 则为了体现出进攻的同时不冲刺, 则可以得到辅助映射矩阵 $M_{m \times p}$ 为:

$$M = E \times \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \end{bmatrix}_{2 \times 4}$$

$M_{m \times p}$ 矩阵的本质为由 $H1$ 到 $H7$ 的映射, 行数代表前序动作头 $H1$ 的维度 m , 列数代表后序被影响的动作头 $H7$ 的维度 P , 每一行都与单位矩阵 $E_{m \times m}$ 中的每一行对应关联, 每行的具体值是指前序动作输出头 $H1$ 中每个子动作对后序动作输出头 $H7$ 的影响。

最后, 再进行矩阵相乘, 将样本中 $\tau_{n \times m}$ 中发起进攻时不能冲刺的影响映射到动作输出头 $H7$, 得到动作输出头 $H7$ 对应的动作遮罩矩阵 X_7 为:

$$X_7 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}_{3 \times 2} \times \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix}_{2 \times 4} = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \end{bmatrix}_{3 \times 4}$$

动作遮罩矩阵 X_7 用于表征所有样本中前序动作输出头 H1 输出的对象子动作对后层动作输出头 H7 对应的动作类型中子动作的影响, 可以看出, X_7 尺寸只和样本个数和动作输出头 H7 维度有关, 行数代表采样了 3 个样本, 列数代表动作类型 a7 中不同的子动作, 上述过程已经将所有样本中确定的对象子动作为进攻时对应的冲刺子动作屏蔽掉, 再对动作输出头 H7 确定的动作类型 a7 中子动作的概率分布进行概率分布转换, 即可保证动作输出头 H7 不会采样冲刺动作。

同样的, 动作类型 a2 也使用上述处理得到的动作类型 a2 中对象子动作对动作类型 a3 的动作遮罩矩阵 Y_7 , 由于动作类型 a1 和动作类型 a2 均独立影响动作类型 a7, 则动作输出头 H7 对前序动作输出的对象子动作对动作类型 a7 中子动作的影响进行叠加, 从而得到的动作输出头 H7 对应的动作遮罩矩阵 L_7 为:

$$L_7 = X_7 \odot Y_7$$

L_7 矩阵是 X_7 和 Y_7 的哈达玛积, L_7 所有值为 0 的元素映射的动作类型 a7 中子动作会被进行遮罩。

基于上述方式, 可以将前序确定的对象子动作对于后序动作输出头对应的动作类型中子动作的影响映射到动作遮罩矩阵, 从而将该影响在 n 个动作输出头中依次向后传递, 能够在动作决策模型的训练过程中有效解决在高维度的动作空间下, 决策模型进行大量无效探索的问题。

二、存在至少两个已确定的对象子动作进行组合后, 与待决策动作类型中的待遮罩子动作存在依赖限制关系。

在另一种可能的实施方式中, 可能存在已确定的对象子动作, 这两个对象子动作对于后序动作类型中的子动作不产生独立影响, 在至少两个已确定的对象子动作进行结合后, 会与待决策的动作类型中的子动作存在依赖限制关系。例如, 存在两个动作类型为移动方向以及水平转向角度, 这两个动作类型中子动作均不会单独对待确定的动作类型中的冲刺动作产生影响, 然而在确定出的对象子动作中水平转向角度与移动方向之间差距较大的情况下 (极端情况为 180°), 例如出现第一虚拟角色注视方向为 0° 但移动方向为 180° 方向, 由于在常规的设定中不支持第一虚拟角色倒退冲刺, 则此时第一虚拟角色不支持同时执行冲刺动作。

在这种情况下, 应当将至少两个动作输出头输出的对象子动作进行结合, 从而确定至少两个对象子动作结合对待决策的动作类型中子动作的影响, 并将该影响映射为待决策动作类型对应的动作遮罩矩阵。

可选地, 先将至少两个动作输出头输出的对象子动作各自进行独热矩阵编码, 再将至少两个独热矩阵进行拼接, 形成合并独热矩阵, 再对合并独热矩阵经过逻辑转换, 使其映射到一个单位矩阵上。随后, 基于上述方式一中所示出的方式, 确定辅助映射矩阵, 并基于辅助映射矩阵实现每行 (每种动作) 映射至少两个对象子动作进行结合后对待决策动作类型的影响, 从而实现对待决策动作类型中子动作的动作遮罩。

示意性的, 存在动作类型 a3 为移动方向 (包括在 360° 内均匀分布的八种移动方向)、动作类型 a4 为水平朝向 (包括在 360° 内均匀分布的八种朝向角度) 以及动作类型 a7 为虚拟角色的移动状态 (包括静步、快走、冲刺以及原地不动), 动作类型 a3 对应动作输出头 H3、动作类型 a4 对应动作输出头 H4, 动作类型 a7 对应动作输出头 H7。

其中, 动作类型 a3 对应动作维度为 s, 动作类型 a4 对应动作维度为 h, 动作输出头 H3 的不同采样子动作组合成独热矩阵为 $s \times s$ 的单位矩阵 $E_{s \times s}$, 动作输出头 H4 所有不同采样动作组合成的 one-hot 矩阵是一个 $h \times h$ 的单位矩阵 $E_{h \times h}$ 。对单位矩阵 $E_{s \times s}$ 以及单位矩阵 $E_{h \times h}$ 进行特征拼接得到的合并独热矩阵 $T_{(s \times h) \times (s + h)}$ 并不是一个方阵, 行数为 $s \times h$ 代表所有动作组合情况, 列数是 $s + h$ 代表结合的维度, 即动作类型 a3 以及动作类型 a4 的子动作数量之和。随后, $T_{(s \times h) \times (s + h)}$ 将转换为一个维度是 $s \times h$ 的单位矩阵 $E_{(s \times h) \times (s \times h)}$, 则该单位矩阵的每一行

代表一种动作结合的情况，随后可以根据依赖限制关系确定映射辅助矩阵 M。

可选的，假设所有样本中动作输出头 H3 和动作输出头 H4 输出对象子动作对应的合并独热矩阵为 $\tau_{n \times (s+h)}$ ，则通过以下计算可以确定动作输出头 H7 对应的动作遮罩矩阵：

首先计算样本中每行动作组合转化为近似独热矩阵的映射为：

$$U_{n \times (s \times h)} = \tau_{n \times (s+h)} \times X_{(s+h) \times (s \times h)}$$

其中， $X_{(s+h) \times (s \times h)}$ 为 $T_{(s \times h) \times (s+h)}$ 的广义逆矩阵， $U_{n \times (s \times h)}$ 为样本中每行动作组合转化为近似独热矩阵的映射，由于 $X_{(s+h) \times (s \times h)}$ 不是 $T_{(s \times h) \times (s+h)}$ 精确的逆矩阵，则需要 $U_{n \times (s \times h)}$ 转换为精确地独热矩阵：

$$Z_{n \times (s \times h)} = one - hot(\arg\max(U_{n \times (s \times h)}))$$

其中， $Z_{n \times (s \times h)}$ 的每一行代表一个样本的动作组合，每行存在 $s \times h$ 个矩阵值，只存在一个元素值为 1，其余元素值均为 0，元素值为 1 表征动作组合将会与映射辅助矩阵 M 对应行关联映射，从而影响动作输出头 H7 对应的动作遮罩矩阵对应行的矩阵值。 $F_{n \times p}$ 为动作输出头 H7 中的动作遮罩矩阵，每行表征动作类型 a3 以及动作类型 a4 会对动作类型 a7 中子动作产生的影响。

通过上述计算，能够在不同子动作之间进行组合对待决策动作类型中子动作存在依赖限制关系的情况下，实现动作遮罩，从而解决强耦合高维度复杂动作空间下决策网络探索效率低的问题。

三、角色状态信息以及环境感知信息指示当前虚拟环境的情况，与待决策动作类型中的待遮罩子动作存在依赖限制关系，即在当前虚拟环境下不适合执行待遮罩子动作。

在一种可能的实施方式中，第一虚拟角色可能具有攻击其他虚拟角色，或者躲避其他虚拟角色等任务，因此在进行动作决策时，可能会基于第一虚拟角色对其他虚拟角色的可视情况对待决策动作类型中的子动作进行动作遮罩。

可选地，以第一虚拟角色的眼部位置为起始点，向第一虚拟角色朝向方向上的其他虚拟角色的各个部位发射环境感知射线。根据环境感知射线的反射情况，确定虚拟环境中其他虚拟角色的各个部位的可视情况，可视情况用于表征其他虚拟角色各个部位是否被掩体遮挡。在环境感知射线被障碍物遮挡，则其他虚拟角色的该部位不可视，在环境感知射线未被障碍物遮挡，且碰撞到其他虚拟角色的情况下，其他虚拟角色该部位可视。

请参考图 11，其示出了本申请一个示例性实施例提供的可视情况感知的示意图。从第一虚拟角色 1101 的眼部为起始点，向对手虚拟角色 1102 的各个部位发射环境感知射线，其中存在部分感知射线被障碍物 1103 遮挡，则这部分射线对应的对手虚拟角色身体部位不可视，而存在感知射线 1104 未被障碍物遮挡，与对手虚拟角色的头部发生碰撞，则对手虚拟角色的头部可视。

上述内容描述了虚拟环境中基于环境感知射线和可视情况进行动作优化的过程。通过考虑虚拟环境中的视觉障碍和可见性，有助于能够更智能地决定虚拟角色的动作行为，使其更符合现实世界中的行为逻辑，提高第一虚拟角色在虚拟环境中交互的真实性和逼真感，使玩家具有更加沉浸的体验，第一虚拟角色还能够根据周围环境的实时变化和障碍物的移动或消失，灵活调整其行动，增强虚拟环境中第一虚拟角色的自主性。

在确定其他虚拟角色各个部位的可视情况后，计算机设备基于其他虚拟角色的各个部位的可视情况，对角色动作类型中的子动作进行动作遮罩。

其中，对象子动作与其他虚拟角色的各个部位的可视情况具有依赖限制关系。例如，角色动作类型为射击动作的情况下，若其他虚拟角色的各个部位的可视情况指示该其他虚拟角色不可视，则将角色动作中的“射击”子动作进行动作遮罩，能够避免第一虚拟角色在未瞄准敌人的情况下进行频繁射击。

在对角色动作类型中子动作进行遮罩后，计算机设备通过角色动作输出头，从角色动作

类型中未被遮罩的子动作中确定子动作。

在另一种可能的实施方式中，对于 n 个动作输出头，其串行顺序基于对应的动作类型之间的依赖关系确定。则对于相同动作类型，但两个动作输出头对应的子动作细粒度不同的情况下，可以先根据粗粒度确定出大致的动作范围，再基于已确定的粗粒度对象子动作确定细粒度对象子动作。

在一种可能的实施方式中， n 个动作输出头中存在至少两个动作输出头对应动作类型相同，存在第 i 动作输出头对应粗粒度的子动作，存在第 j 动作输出头对应细粒度的子动作， i 小于 j 。则在确定对象子动作的过程中，计算机设备首先通过决策网络中第 i 动作输出头从粗粒度的子动作中确定第 i 对象子动作，随后基于融合向量以及第 i 对象子动作对应的嵌入编码向量，通过第 j 动作输出头从细粒度的子动作中确定第 j 对象子动作，第 j 对象子动作所指示的动作范围小于第 i 角色动作所指示的动作范围。

例如，针对于表 2 中的粗粒度水平朝向以及细粒度水平朝向，在进行动作决策时，应当先确定粗粒度水平朝向再确定细粒度水平朝向。

则在第一虚拟角色具有瞄准其他虚拟角色的需求的情况下可以先确定可以瞄准到对手虚拟角色的转向角度范围，再对转向动作类型中无法使第一虚拟角色瞄准对手虚拟角色的转向角度进行动作遮罩。

请参考图 12，其示出了本申请一个示例性实施例提供的确定竖直方向有效瞄准范围的示意图。首先根据对手虚拟角色所处位置坐标以及第一虚拟角色所处位置坐标确定第一虚拟角色与对手虚拟角色之间的距离 d ，并且获取对手虚拟角色的高度 h ， $h=h_z-h_f$ 。则第一虚拟角色所持虚拟道具的准心在竖直方向上落在第一虚拟角色身高范围内的情况下，即可瞄准对手虚拟角色，则根据三角函数可以得到：

$$\theta_{\max} = \arctan[(h_z - h_f)/d]$$

$$\theta_{\min} = \arctan[(h_f - h_z)/d]$$

$$\tau_{pitch} = \theta_{\max} - \theta_{\min}$$

其中，以水平方向为当前朝向方向即为 0° ， $\theta_{\max}$ 为向上转动的有效瞄准范围， $\theta_{\min}$ 为向下转动的有效瞄准范围，则二者之差 τ_{pitch} 即为竖直转向动作的有效瞄准范围。

而对于水平转向动作类型可以先大致确定出能够瞄准到对手虚拟角色的粗粒度转向动作，再通过调整细粒度转向动作以实现瞄准对手虚拟角色的目的。

示意性的， n 个动作输出头中存在第一转向动作输出头对应粗粒度转向动作，以及存在第二转向动作输出头对应细粒度转向动作，第一转向动作输出头用于确定第一转向角度，第二转向动作输出头用于确定第二转向角度。

示意性的，第一转向角度是基于多个粗粒度转向角度确定的粗粒度形式的转向角度；第二转向角度是基于多个细粒度转向角度确定的细粒度形式的转向角度。

首先，计算机设备基于第一虚拟角色的朝向、第一虚拟角色所处位置坐标以及其他虚拟角色的位置坐标，确定理想转向角度，以及确定第一虚拟角色与其他虚拟角色之间的瞄准距离。计算机设备基于第一虚拟角色所处位置坐标以及其他虚拟角色的位置坐标，首先计算出瞄准距离随后基于三角函数其他虚拟角色相较于第一虚拟角色的方位，从而根据第一虚拟角色当前朝向以及方位确定出理想转向角度。

随后，基于瞄准距离以及其他虚拟角色的宽度，确定有效瞄准范围。

示意性的，请参考图 13，其示出了本申请一个示例性实施例提供的水平方向有效瞄准范围的示意图。第一虚拟角色当前朝向为 0° ，第一虚拟角色与其他虚拟角色之间的距离为 d ，对手虚拟角色的宽度为 W ，则确定有效瞄准范围为：

$$\theta = \arctan[(W/2)/d]$$

$$\tau_{pitch} = 2\theta$$

其中, θ 为第一虚拟角色向顺时针和逆时针旋转角度中一个有效旋转范围, 将顺时针转动有效范围与逆时针转动有效范围进行结合, 从而得到有效瞄准范围 τ_{pitch} 。

在确定理想转向角度后, 先基于第一转向动作输出头确定的第一转向角度, 确定第一转向角度与理想转向角度之间的差值, 该差值可能为正值也可能为负值。正值表示进行第一转向角度之后, 第一虚拟角色的预估朝向在其他虚拟角色的顺时针方位, 负值表示进行第一转向角度之后, 第一虚拟角色的预估朝向在其他虚拟角色的逆时针方位。

可选的, 计算机设备基于理想转向角度确定有效粗粒度转向角度, 并对虚拟角色的粗粒度转向动作中不属于有效粗粒度转向角度的转向角度进行动作遮罩。再从未被遮罩的子动作中确定出第一转向角度。随后, 计算机设备基于角度差值以及有效瞄准范围确定有效细粒度转向角度。确定出的有效细粒度转向角度为第一虚拟角色在当前朝向按照细粒度转向角度执行转向动作后朝向在有效瞄准范围内。再对粗粒度转向动作中的不属于有效细粒度转向角度的转向角度进行动作遮罩, 并通过第二转向动作输出头, 从细粒度转向动作中未被遮罩的有效细粒度转向角度中确定第二转向角度。最终使得第一虚拟角色执行第一转向动作以及第二转向动作后, 能够瞄准其他虚拟角色。

上述内容介绍了虚拟环境中第一虚拟角色转向过程中的粗粒度和细粒度控制优化。过粗粒度和细粒度的转向控制, 能够使得第一虚拟角色的转向行为更加自然和智能, 更接近真实世界中的行为表现, 优化的转向控制能够提升第一虚拟角色在游戏中的战术效果和战斗策略, 增强游戏体验的沉浸感, 通过动作遮罩技术也能够减少不必要的计算和转向操作, 提升动作执行效率和性能, 精确控制转向角度可以使得第一虚拟角色更精准地执行各种动作和交互, 从而增强动作操作精度和可控性。

上述内容介绍了决策网络在多层次动作选择中的作用, 包括粗粒度和细粒度的子动作选择过程: 通过多层次的动作选择, 有助于在粗粒度和细粒度上指导每一步动作的执行, 从而提高了整体动作的执行精度和效率, 通过逐步缩小动作范围, 可以根据虚拟环境的变化和任务需求调整和优化动作的选择和执行方式, 增强了对对象子动作确定的适应性和灵活性, 每一步的动作选择都基于前一步的结果和详细的动作描述, 减少了由于信息不完整或误解而导致的决策误差, 提高了对象子动作确定的可靠性和稳定性。

在一种可能的实施方式中, 为避免角色动作频繁做出某个动作, 不符合人类行为逻辑, 因此, 可以为部分子动作设置冷却时间, 在该子动作冷却未完成的情况下, 对该子动作进行动作遮罩。从而避免出现第一虚拟角色反复执行同一动作造成抽搐现象。

本申请实施例中, 基于三种不同的依赖限制情况以及已确定的对象子动作对待决策的动作类型中的子动作进行动作遮罩, 从而提升高维度的复杂动作空间下的动作决策效率在一定程度上节约计算资源。

在一些实施例中, 角色状态信息中包含多种能够描述当前局势状态的标量信息, 可能为第一虚拟角色的状态信息以及第一虚拟角色与环境进行交互的信息。实质上, 环境感知信息是第一虚拟角色对周围虚拟环境感知得到的, 刻画了第一虚拟角色相对于虚拟环境的相对特征, 因此, 只需要在足够丰富的虚拟环境中对动作决策模型进行训练, 即可使得动作决策模型获取到的相对特征足够丰富多样, 从而保证动作决策模型拥有对环境感知的泛化能力。

然而, 角色状态信息中可能会包含一些与特定地图场景具有强关联的特异性特征, 因此计算机设备会基于该特征进行进一步的泛化处理。

可选的, 计算机设备对角色状态信息中包含的绝对状态信息进行泛化处理, 得到相对状态信息, 相对状态信息是第一虚拟角色相对于其他虚拟角色或者障碍物的状态信息。

可选地, 可以采用相对信息取代绝对信息的方式实现泛化处理, 例如将第一虚拟角色、其他虚拟角色以及掩体点等绝对位置信息或绝对位置朝向, 替换为虚拟角色与虚拟角色之间, 虚拟角色与掩体点/障碍物之间的相对位置即位置朝向。

本申请实施例中, 通过对动作决策模型的输入信息进行泛化处理, 并结合多种多样的训

练场景，能够训练得到具有良好泛化性的动作决策模型，使动作决策模型易于部署于各类地图场景；其中，相对状态信息更有助于第一虚拟角色在动态环境中进行决策和行动，通过泛化处理消除了绝对坐标系的依赖性，使得虚拟角色能够更灵活地响应周围环境的变化，提高了第一虚拟角色的感知和决策能力，还增强了动态环境中的应变能力和智能表现，有效地促进虚拟角色在复杂情景下的行为优化和实时响应。

请参考图 14，其示出了本申请一个示例性实施例提供的动作决策模型的示意图。其中包括编码网络、决策网络、动作遮罩网络以及价值网络。计算机设备将获取到的角色状态信息（包括向量交互特征），以及环境感知信息输入动作决策模型的编码网络中进行编码、拼接，得到拼接特征。随后通过 LSTM 网络对拼接特征进行特征提取得到融合特征。将融合特征输入到决策网络中，决策网络包含全连接层以及激活函数 ReLU，决策网络基于接收到的融合特征确定角色动作，在进行动作决策的过程中，基于动作遮罩网络对子动作进行遮罩，最终通过动作采样确定对象子动作。

此外，在对动作决策模型进行训练的过程中，动作决策模型中还包括价值网络，通过价值网络为确定的对象子动作进行将至评估，并根据评估得到的价值估计调整动作决策模型。此外，还可以将部分对局信息输入价值网络以使得价值网络能够根据当前对局情况判断预估对象子动作的价值。

在应用动作决策模型之前，应当在训练对局中对模型进行训练，下面将通过一个示例性实施例对动作决策模型的训练过程进行说明。

请参考图 15，其示出了本申请一个示例性实施例提供的动作决策模型训练过程的流程图，该过程包括以下步骤：

步骤 1501，获取第一虚拟角色的样本角色状态信息以及第一虚拟角色对当前所处虚拟环境的样本环境感知信息。

其中，样本环境感知信息中不同模态的样本环境感知信息具有不同的信息维度。

本步骤的具体实施方式可参照上述步骤 201 中获取角色状态信息以及环境感知信息的过程，本实施例在此不做赘述。

步骤 1502，通过动作决策模型的编码网络对样本角色状态信息和样本环境感知信息进行特征编码及融合，得到样本融合特征。

其中，编码网络通过不同编码方式对样本角色状态信息以及不同模态的样本环境感知信息进行特征编码。

本步骤的具体实施方式可参照上述步骤 202 中基于角色状态信息以及环境感知信息得到融合编码的过程，本实施例在此不做赘述。

步骤 1503，基于样本融合特征，通过强化学习的方式训练动作决策模型。

步骤 1501 至步骤 1503 介绍了基于样本角色状态信息和样本环境感知信息训练决策模型的内容。训练模型的过程与应用模型进行动作决策的过程相对应，借助训练过程有助于提高决策模型的鲁棒性，提升决策模型进行动作决策时的准确性。

基于强化学习训练动作决策模型的过程包括如下内容。

步骤 1503a，将样本融合特征输入动作决策模型中的决策网络，得到预估角色动作。

实际上，对动作决策模型进行训练的过程即为对动作模型中决策网络进行强化训练的过程。

此过程的具体实施方式可参照上述实施例中确定角色动作的内容，本实施在此不做赘述。

步骤 1503b，控制第一虚拟角色执行预估角色动作，得到第一虚拟角色的预估动作执行结果。

该预估动作执行结果包括第一虚拟角色执行预估角色动作之后第一虚拟角色的角色状态信息的变化以及第一虚拟角色感知到的环境感知信息的变化。

步骤 1503c，基于第一虚拟角色的预估动作执行结果，确定预估动作执行奖励。

在进行训练的过程中，会为第一虚拟角色设定一定的任务，因此，为避免第一虚拟角色为完成任务而采取极致的操作，导致第一虚拟角色拟人化程度较低，因此，在基于任务进度预估角色动作的执行奖励外，还会基于预估角色动作的拟人化程度判定预估角色动作的执行奖励。

首先，计算机设备会执行各项奖惩项目的权重系数比例，分为主导奖励（即针对于预设任务进度的奖励）以及辅助拟人化奖励。在动作决策模型的训练初期，将主导奖励的权重系数调高，辅助拟人化奖励的权重系数降低，以使得动作决策模型优先以完成任务为目的进行学习。在进行多轮训练后，动作决策模型的强度不断提升，从而需要根据动作决策模型的表现和对应拟人化指标变化趋势动作调整辅助拟人化奖励的权重系数。其中，拟人化指标是指第一虚拟角色执行特定动作的概率与真实玩家执行特征动作的概率的匹配程度。例如，在第一虚拟角色在对局中无法合理使用掩体进行拉扯战斗的情况下，则可以确定主导奖励的权重过大，从而降低主导奖励的权重，以保证动作决策模型在拥有一定强度的基础上能够确定出拟人化的角色动作。

本申请实施例中，预估动作执行奖励主要包括属性奖励、对局胜利奖励、对局失败奖励以及任务奖励中的至少一项。

属性奖励主要包括第一虚拟角色的生命值属性奖励，在第一虚拟角色的属性值降低的情况下，将属性奖励确定为预估动作执行奖励。例如，在第一虚拟角色执行预估角色动作之后，第一虚拟角色的生命值属性降低，则确定该角色动作获得反向的属性奖励。或者，在第一虚拟角色的属性值提升的情况下，将属性奖励确定为预估动作执行奖励。例如，在第一虚拟角色执行预估角色动作之后，第一虚拟角色的生命值属性提升，则确定该角色动作获得正向的属性奖励。

对局失败奖励是指在第一虚拟角色取得对局失败的情况下，会获取极大的反向奖励（即惩罚），在第一虚拟角色取得对局失败的情况下，将对局失败奖励确定为预估动作执行奖励。

对局成功奖励与对局失败奖励相反，在第一虚拟角色取得对局成功的情况下，会获得极大的正向奖励，在第一虚拟角色的属性值降低的情况下，将属性奖励确定为预估动作执行奖励。

在一些实施例中，第一虚拟角色在对局中存在预设任务，例如守护宝箱、躲避追击或者抵达指定地点等等，因此，在第一虚拟角色在训练对局中任务成功的情况下，将任务奖励确定为预估动作奖励。

此外，为使动作决策模型能够进行拟人化的动作决策，在预估动作奖励中还包括拟人化奖励。计算机设备基于第一虚拟角色在训练对局中执行至少两次预估角色动作的预估动作执行结果，确定至少两次预估角色动作的拟人化属性值。该拟人属性值可能为执行特定动作的次数、连续确定的至少两个预估角色动作的走位位置等等。在拟人化属性值低于拟人属性阈值的情况下，将拟人化奖励确定为预估动作执行奖励。

示意性的，请参考表 3，其示出了本申请一个示例性实施例提供的预估动作执行奖励。

表 3

类别	子分类	奖惩设置
主导奖励	任务奖励	属性变化
		对局胜负
		任务是否成功
辅助拟人化奖励	姿态拟人化奖励	合理侧身
		合理蹲起
		合理静步
		大幅度转向惩罚
		走位分散

战斗拟人化奖励	躲避三秒再进行进攻占比
	躲避 2.5 秒再进行进攻占比
	掩体利用
	危险区域停留
	射击后撤
	暗处伏击
	远程攻击
	对手靠近时进攻
	危险时刻探头
	攻击占比

在上表中，存在部分预估动作奖励需要根据连续多次预估角色动作进行判断，例如，合理侧身为拟人化奖励，真实玩家控制的虚拟角色通常不会在短时间内进行多次侧身，或者在周围没有掩体的情况下进行侧身，因此，在确定出的多次角色动作中，侧身动作的拟人化属性值（即侧身动作的次数）超过拟人属性阈值的情况下，则将其确定为不合理侧身，基于动作决策模型反向奖励。

在表 3 中包括走位分散奖励、掩体利用奖励、暗处伏击奖励以及合理动作奖励（合理侧身、合理蹲起、合理静步等）。

一、走位分散奖励是指第一虚拟角色执行预估动作的情况下，连续多次走位的紧密程度降低的奖励。即第一虚拟角色不会连续多次在原地徘徊，在大规模地图中能够更加分散的走位，探索地图的不同位置。

可选的，预估动作执行结果包括第一虚拟角色执行前 $n+1$ 次动作后的位置点。计算机设备基于第一虚拟角色执行第 $n+1$ 次预估角色动作后的位置点，确定执行第 $n+1$ 次动作后的位置点与前 n 次执行动作后的位置点的第一紧密度，以及确定执行第 n 次动作后的位置点与前 $n-1$ 次执行动作后位置点的第二紧密度，在第一紧密度小于第二紧密度的情况下，将走位分散奖励确定为预估动作执行奖励。在对动作决策模型进行训练时，可能会导致第一虚拟角色陷入局部区域会进行绕圈的现象，因此，可以基于紧密中心性算法，引导第一虚拟角色执行动作决策模型决策出的预估角色动作后，能够走位足够分散，走位路线尽可能遍及地图全局，以训练第一虚拟角色的走位能力。

可选地，在执行第 $n+1$ 次动作后的位置点与前 n 次执行动作后的位置点的第一紧密度小于定执行第 n 次动作后的位置点与前 $n-1$ 次执行动作后位置点的第二紧密度的情况下，确定获取到正向走位分散奖励。或者，在执行第 $n+1$ 次动作后的位置点与前 n 次执行动作后的位置点的紧密度大于定执行第 n 次动作后的位置点与前 $n-1$ 次执行动作后位置点的紧密度的情况下，确定获取到反向的走位分散奖励。从而引导第一虚拟角色尽量远离最近去过的历史位置，避免出现虚拟角色原地徘徊的情况。

上述过程介绍了将走位分散奖励作为预估动作执行奖励的内容。通过奖励第一虚拟角色在不同时间段内的位置变化，鼓励了更加多样化的战术策略和走位方式，增加了游戏或仿真中的战术深度和挑战性：走位分散奖励也使得第一虚拟角色在战斗或竞技环境中更难被对手预测，从而提高了第一虚拟角色的生存能力和逃生能力；此外，通过考虑位置点的紧密度评估，有助于更好地优化动作决策模型，使其更加智能地选择和执行动作，以最大化奖励并改善整体表现。

二、掩体利用奖励是指第一虚拟角色执行预估动作的情况下，所处位置与掩体相邻的奖励。在存在其他虚拟角色的情况下，期望第一虚拟角色能够尽量处于掩体旁边，以防止被其他虚拟角色攻击到。

可选的，预估动作执行结果包括第一虚拟角色执行预估角色动作后的样本环境感知信息，样本环境感知信息中包含二维深度图，二维深度图用于表征第一虚拟角色朝向方向上的障碍

物对第一虚拟角色的遮掩情况。

计算机设备对二维深度图进行区域划分，得到至少两个深度区域。确定第一深度区域的第一遮掩率，以及与第一深度区域相邻的至少两个第二深度区域的第二遮掩率，第一深度区域位于第一虚拟角色的视野中心。基于第一遮掩率与第二遮掩率的最小差值，确定掩体利用奖励，掩体利用奖励的奖励程度与最小差值呈正相关关系，将掩体利用奖励确定为预估动作执行奖励。

上述内容介绍了对二维深度图进行区域划分，并基于这些区域的遮掩率确定掩体利用奖励的内容。通过评估和奖励掩体利用能后鼓励第一虚拟角色在战斗中选择更优的掩体位置，以优化其防御或进攻能力，考虑到深度区域的遮掩率差异，虚拟角色被鼓励选择能够提供更好视野和防护的位置，从而增强了其战术意识和视觉策略能力，不仅增加了游戏的战术深度，还提升了玩家的战斗体验，使游戏更加真实和具有挑战性。

本申请实施例中通过重塑改造深度图矩阵实现对掩体利用奖励的确定。在游戏场景中掩体是在第一虚拟角色战斗过程中为第一虚拟角色提供遮掩的区域，在虚拟环境中存在多个掩体，不同掩体之间的位置、朝向以及形状中至少有一种存在区别。而在二维深度图中，能够表征第一虚拟角色朝向方上的障碍物深度情况，因此，可以通过二维深图的像素值分布情况确定掩体的像素分布区域。

请参考图 16, 其示出了本申请一个示例性实施例提供的对深度图进行区域划分的示意图。按照深度图数值方向进行划分，得到 1-19 共 19 个区域，将深度图竖直划分成若干列，计算每一列的明亮率，明亮率越大表明前方越亮，越没有可能会有障碍物遮挡，则遮掩率越低。在应用中，期望第一虚拟角色能够选择存在掩体又能够攻击到其他虚拟角色的区域，因此，期望深度图中间列的明亮率最小，由于中间列反应第一虚拟角色正前方的视野信息，明亮率越小则遮掩率越大，正前方越有可能是掩体区域。此外，期望第一虚拟角色能够处于掩体边缘，即能够通过侧身等动作对其他虚拟角色发起进攻，因此，要保中间列的明亮率和其左右两列中较大明亮率的差值绝对值应尽可能大，即使得中间列的明亮率与其左右两列的明亮率（遮掩率）差值的最小值尽可能大。对应图 16 中，期望第 10 列的遮掩率尽可能大，且第 9 列和第 11 列中存在一个区域的遮掩率尽可能小。

可选的，深度区域对应的明亮率 v_j 可以通过以下公式进行计算：

$$v_j = \frac{\sum_{i=1}^{i=n} x_{i,j}}{\sum_{i=1}^{i=n} 1}$$

其中 v_j 为每个深度区域中实际像素与每列理论最大像素和的比值，而深度区域的遮掩率为 $1-v_j$ 。则掩体利用奖励可以通过下式进行计算：

$$r = \alpha \cdot (1 - v_{10}) + \beta \cdot |1 - v_{10} - \min(1 - v_9, 1 - v_{11})|$$

其中， r 为掩体利用奖励，而 α 与 β 为奖励超参数。

需要说明的是，对二维深度图进行区域划分的方式还可以基于水平方向进行划分，或者，划分的区域数量可以基于虚拟环境进行调整，在虚拟环境较为复杂的情况下，可以采用更多的深度区域，对二维深度图进行更精细的划分，从而获取更精细的掩体效果。在地图场景较大的情况下，控件结构较为简单，则可以粗粒度进行深度区域划分，本实施例对此不进行限定。

三、暗处伏击奖励即为可视性奖励，可视性奖励是指虚拟角色执行预估动作的情况下，所处位置为对其他虚拟角色的有效攻击位置的奖励。在存在其他虚拟角色的情况下，期望其他虚拟角色能够处于第一虚拟角色的可视位置，但第一虚拟角色不处于其他虚拟角色的可视位置，有利于发起进攻。

可选的，预估动作执行结果包括第一虚拟角色与其他虚拟角色之间的相互可视关系，相

互可视关系用于表征第一虚拟角色以及其他虚拟角色相对于彼此的可视情况。

计算机设备在相互可视关系指示第一虚拟角色不在其他虚拟角色的可视范围，且其他虚拟角色在第一虚拟角色的可视范围的情况下，确定预估动作执行奖励为正向的可视性奖励。

在相互可视关系指示第一虚拟角色在其他虚拟角色的可视范围，且其他虚拟角色不在第一虚拟角色的可视范围的情况下，确定预估动作执行奖励为反向的可视性奖励。

即第一虚拟角色相较于其他虚拟角色处于有利进攻位置的情况下确定获取正向的可视性奖励。反之，第一虚拟角色相较于其他虚拟角色处于不利的进攻位置的情况下，确定获取反向的可视性奖励。

可选的，相互可视关系可以通过射线探测的方式获取。以第一虚拟角色眼部位置为起始点，向第一虚拟角色朝向方向上的其他虚拟角色的各个部位发射环境感知射线。根据环境感知射线的反射情况，确定虚拟环境中其他虚拟角色的各个部位的可视情况，可视情况用于表征其他虚拟角色各个部位是否被掩体遮挡。在可视情况指示其他虚拟角色各个部位均被障碍物遮挡的情况下，确定其他虚拟角色不在第一虚拟角色的可视范围。同理，以其他虚拟角色眼部位置为起始点，向其他虚拟角色朝向方向上的第一虚拟角色的各个部位发射环境感知射线，根据环境感知射线的反射情况，确定第一虚拟角色的各个部位的可视情况，在可视情况指示第一虚拟角色各个部位均被障碍物遮挡的情况下，确定第一虚拟角色不在其他虚拟角色的可视范围。

上述内容介绍了根据虚拟角色之间的相互可视关系确定预估动作执行奖励的方向性的过程。正向的可视性奖励鼓励虚拟角色在不被其他角色察觉的情况下采取行，这可以用于增强潜行、隐蔽和执行战术攻击等策略，从而增加游戏或仿真环境中的战斗深度和战术选择；反向的可视性奖励鼓励虚拟角色在被其他角色发现时采取行动。这可以促进快速反应、避免暴露和实施防守策略，有助于提升虚拟角色的生存能力和游戏体验；有效地丰富了虚拟角色在游戏或仿真环境中的行为选择，提高了整体游戏体验和参与度。

四、合理动作奖励是指第一虚拟角色执行预估动作的情况下，符合真实玩家操作逻辑的奖励。在进行模型训练的过程中，期望第一虚拟角色的行为能够尽可能的拟人化，执行更符合人类逻辑的行为。

在确定预估角色动作之前，计算机设备基于第一虚拟角色的样本角色状态信息以及样本环境感知信息，确定第一虚拟角色的理想动作。例如，环境感知信息表征第一虚拟角色前方存在较低的掩体，则可以为第一虚拟角色的理想动作可以为下蹲动作，再例如，环境感知信息表征第一虚拟角色右手边存在其他虚拟角色，则可以确定理想动作应当为向右转向。

随后，在动作决策模型对应的预估角色动作与理想动作一致的情况下，将合理动作奖励确定为预估动作执行奖励。

上述内容介绍了基于虚拟角色状态信息和环境感知信息确定理想动作并据此决定合理动作奖励的使用情况。通过结合角色状态和环境信息计算理想动作，能够帮助虚拟角色做出更明智和基于情境的行动选择，提高了整体游戏或仿真系统的表现和智能水平，也有助于玩家更直观地与虚拟角色进行互动，因为虚拟角色表现出更接近真实行为的智能和逼真性，慈爱通过奖励合理动作，也有有效地促进资源的有效利用和任务的高效执行，从而优化了动作执行效率和性能。

在一些实施例，以第一虚拟角色为起始点向多个方向发射环境感知射线，不同环境感知射线位于同一水平高度，环境感知射线碰撞到障碍物表面的情况下，沿环境感知射线的法线方向反射；在环境感知射线的反射情况指示存在第一环境感知射线的投射距离小于距离阈值的情况下，将与第一环境感知射线相邻，且投射距离大于距离阈值的环境感知射线的发射方向确定为理想转向方向，投射距离是指环境感知射线从起始点到反射点之间的距离；在转向予动作指示的转向方向与理想转向方向一致的情况下，将合理动作奖励确定为预估动作执行奖励。

其中, 通过使用环境感知射线计算理想转向方向, 虚拟角色可以更有效地规避障碍物和选择最佳路径, 从而优化其移动和导航能力, 通过模拟现实世界中物体反射光线的行为, 使得虚拟环境更加逼真和具有挑战性, 通过奖励合理动作, 不仅提高了虚拟角色的操作效率, 还优化了在处理复杂场景和动态环境中的响应速度和智能水平, 有助于提升人机交互效率。

上述内容介绍了预估动作执行奖励的实现方式。通过上述四种奖励中的至少一种, 不同的奖励机制不仅能够优化第一虚拟角色的行为决策, 还能提升在游戏或仿真环境中的表现和玩家体验。通过合理动作、掩体利用、可视性和走位分散奖励的设计和实现, 可以更有效地模拟真实世界的战术和策略选择, 提高第一虚拟角色的智能性和逼真度。

示意性的, 请参考图 17, 其示出了本申请一个示例性实施例提供的角色动作中存在转向子动作。则计算机设备以第一虚拟角色为起始点向多个方向发射环境感知射线, 不同环境感知射线位于同一水平高度。计算机设备根据环境感知射线的反射情况, 确定周围的障碍物分布。图 17 中, 在存在一条最短的环境感知射线的长度小于距离阈值的情况下, 则说明第一虚拟角色距离障碍物之间的距离较近, 可以确定此时应当进行转向。环境感知射线以最短射线作为基准向左右扩散, 计算机设备判定射线长度更早超过距离阈值的射线方向(第一虚拟角色的逆时针方向)为第一虚拟角色的理想转向方向。

而当动作决策模型确定出的转向方向与理想转向方向一致的情况下, 将合理动作奖励确定为预估动作执行奖励。在当动作决策模型确定出的转向方向与理想转向方向不一致的情况下则第一虚拟角色可能出现卡墙的行为, 给予一定的反向奖励。

步骤 1503d, 基于预估动作执行奖励更新动作决策模型的模型参数。

基于预估动作执行奖励更新动作决策模型的过程即为, 更改决策模型的模型参数以获取更大的预估动作执行奖励的过程。从而使得动作决策模型能够在进行预设任务的前提下尽可能的实现动作拟人化。

本申请实施例中, 在对动作决策模型进行训练的过程中, 增加辅助拟人化动作奖励, 能够满足模型强度以及拟人性的双重需求, 在准确预测和控制第一虚拟角色的角色行为的同时, 结合动作遮罩逻辑使得第一虚拟角色的拟人性得到极大提升, 由于实际执行结果的奖励还可以即时反馈到动作决策模型中, 从而通过奖励的反馈优化模型的决策策略, 使其逐步改进和优化, 在更好地适应各种场景和挑战的同时更加符合真实玩家的行为逻辑。

本申请实施例中, 第一虚拟角色所处对局中存在至少一个其他虚拟角色, 且其他虚拟角色执行第二动作决策模型输出的动作。在一些场景中, 其他虚拟角色的任务与第一虚拟角色的任务相对应, 例如, 第一虚拟角色的任务为守护宝箱, 而其他虚拟角色的任务可能为夺取宝箱。

其他虚拟角色作为对手虚拟角色, 在对第一虚拟角色对应的第一动作决策模型训练后, 同样需要对第二动作决策模型进行迭代, 以获取不同强度的第二动作决策模型。

可选地, 在第一虚拟角色对应的第一动作决策模型完成一轮训练的情况下, 基于训练后的第一动作决策模型对第二动作决策模型进行训练, 并将完成一轮训练后的第二动作决策模型存储至对手模型池。对手模型池中包含完成至少一轮训练后的第二动作决策模型。

在对第一决策模型进行下一轮的训练时, 从对手模型池中按照一定规则采样处过往的第二决策模型作为对手, 进行下一轮训练。计算机设备从对手模型池中选择训练过的第二动作决策模型, 并构建全新对局, 全新对局中包含执行第一动作决策模型输出的预估角色动作的第一虚拟角色, 以及执行第二动作决策模型输出动作的其他虚拟角色。再在全新对局中对第一动作决策模型进行训练。

示意性的, 具体过程如下:

- (1) 基于客户端内置行为树控制对手虚拟角色, 对第一动作决策模型进行第 0 轮训练;
- (2) 将第 0 轮训练的模型存入评估模型池进行评估;
- (3) 根据完成第 0 轮训练的第一动作决策模型, 对第二动作决策模型进行第 0 轮训练, 并将其添加至对手模型池;
- (4) 从对手模型

池中选择一个过往训练过的第二决策模型作为对手，对第一动作决策模型进行第 K 轮训练；
(5) 将第 K 轮训练的第一动作决策模型存入评估模型池进行评估；(6) 根据第 K 轮训练的第一动作决策模型对第二动作决策模型进行第 K 轮训练，并将其存入对手模型池；(7) 重复 (4) 至 (6) 中的步骤，直至第一动作决策模型的评估结果满足性能评估结果的情况下，确定第一动作决策模型完成训练。

上述内容中描述了虚拟角色的动作决策模型训练和对局构建的过程，通过使用第一模型训练经验加速第二模型的学习过程，可以更快地优化多个决策模型；在对手模型池中存储多个训练后的模型，有助于构建更具挑战性和变化性的对局，提高虚拟角色的对抗学习能力和适应性；通过在实际对局中训练和调整第一动作决策模型，可以不断优化角色的行为策略和智能水平，以应对复杂和动态的游戏或仿真环境。

在一种可能的实施方式中，为使得第一动作决策模型在具有一定强度的同时具有拟人性，可以为对手虚拟角色增加以下几种陪练行为：

- 1、创建全程冲刺追击的对手虚拟角色，有利于训练第一虚拟角色进行全地图探索的能力。
- 2、将对手虚拟角色的生成点设定为随机生成点，在对手虚拟角色与第一虚拟角色的生成距离较近的情况下，为对手虚拟角色设置冷静期，能够避免在训练过程中第一虚拟角色在开局被瞬间击败。
- 3、在对手虚拟角色与第一虚拟角色之间的距离小于安全距离的情况下，判定第一虚拟角色取得对局失败，能够训练第一虚拟角色与对手虚拟角色保持一定的距离，同时能够避免第一虚拟角色被地形卡住。
- 4、使对手虚拟角色小概率的移动至第一虚拟角色正前方，从而模拟真实玩家维度第一虚拟角色的场景，以提升第一虚拟角色的强度。

本申请实施例中，通过对陪练的对手虚拟角色对应的第二动作决策模型以及第一虚拟角色对应的第一决策模型，从而提升第一决策模型的强度，使第一决策模型能够具有稳定的动作决策能力。

本申请实施例中，判断动作决策模型是完成训练，需要结合对局指标以及拟人化指标共同判断。对局指标用于评判动作决策模型的强度，拟人化指标用于评判动作决策模型的拟人化程度。

请参考表 4，其示出了本申请一个示例性实施例提供的对局指标以及拟人化指标。

表 4

对局指标		拟人化指标 (统计 500 局，计算平均每局的拟人性指标)	
平局率	双方平局方式结束对局	侧身率	合理侧身攻击对手虚拟角色
胜率	第一虚拟角色打败对手虚拟角色	蹲伏率	合理利用蹲起动作来隐藏自己不被对手发现的概率
败率	第一虚拟角色被对手虚拟角色打败	躲避率	平均一局游戏中不被发现的占比
遇敌进攻率	遇敌进攻帧数/遇敌总帧数	静步率	第一虚拟角色利用静步方式隐藏声音而不被发现的占比
证据开火率	遇敌打中敌人帧数/遇敌进攻总帧数	掩体利用率	第一虚拟角色利用掩体隐藏自身占比
存活时间	第一虚拟角色在一局游戏中存活的时长	走位分散率	是指第一虚拟角色走位灵活，不会长时间停留在同一个位置
卡墙次数	视野内 30° 内且与障碍物距离小于 0.5m 的帧数	危险区域停留率	第一虚拟角色在特定的战斗场景下处于没有掩体的危险

			空旷区域频率
		大幅转向率	第一虚拟角色的移动朝向不应当大幅度骤变
		进攻回退率	第一虚拟角色在射击对手虚拟角色时能主动一边射击一边远离对手虚拟角色
		暗中射击率	在第一虚拟角色在对手虚拟角色视野范围之外射击
		缠斗率	用于刻画利用掩体进行拉扯战斗的拟人行为

其中，在确定拟人性指标时，根据大量的对局数据确定平均的拟人性指标。

计算机设备在完成对动作决策模型的第*i*轮训练的情况下，获取第*i*轮训练对局的对局数据。基于至少两轮训练对局的对局数据，确定对局指标以及拟人化指标，拟人化指标是指特定动作在至少训练对局的对局数据中的实际执行占比。

在对局指标达到训练完成标准，且拟人化指标与真实玩家在对局中执行特定动作的动作执行占比相匹配的情况下，确定动作决策模型完成训练。

本申请实施例中，基于对局指标以及拟人化指标共同作为动作决策模型的评估标准，从而使训练完成的动作决策模型能够兼顾强度以及动作模型的拟人化程度；不仅提高了动作决策模型的性能和实用性，还能够确保其在多样化的对局情景中表现出色，符合真实玩家的期望和需求。

请参考图 18，其示出了本申请一个示例性实施例提供的虚拟角色的动作决策装置的结构示意图，该装置包括以下结构。

获取模块 1801，用于获取第一虚拟角色的角色状态信息，以及获取所述第一虚拟角色对当前所处虚拟环境的环境感知信息，其中，所述环境感知信息包括多个模态的感知信息，所述多个模态分别代表一个信息维度；

编码模块 1802，用于融合所述角色状态信息和所述环境感知信息分别对应的编码结果，得到融合特征，其中，所述编码结果是通过不同编码方式对所述角色状态信息以及多个感知信息分别进行特征编码得到的结果；

决策模块 1803，用于基于所述融合特征对所述第一虚拟角色进行动作决策，得到角色动作；

控制模块 1804，用于控制所述第一虚拟角色执行所述角色动作。

可选的，所述多模态环境感知信息包括第一模态感知信息、第二模态感知信息以及第三模态感知信息；

所述获取模块 1801，用于：基于所述第一虚拟角色在所述虚拟环境中的角色位置，通过射线探测获取所述第一模态感知信息，所述第一模态感知信息为一维向量，所述第一模态感知信息用于表征所述第一虚拟角色周围同一水平高度上障碍物的深度情况；基于所述第一虚拟角色在所述虚拟环境中的所述角色位置以及所述第一虚拟角色的朝向，通过射线探测获取所述第二模态感知信息，所述第二模态感知信息为所述第一虚拟角色朝向方向上的二维深度图，所述第二模态感知信息用于表征所述第一虚拟角色朝向方向上障碍物的深度情况；基于所述第一虚拟角色在所述虚拟环境中的所述角色位置，通过射线探测获取在所述第一虚拟角色的预设范围内所述虚拟环境的所述第三模态感知信息，所述第三模态感知信息为三维占栅网格图，所述第三模态感知信息用于表征所述预设范围内障碍物的空间分布情况。

可选的，所述获取模块 1801，用于：以所述第一虚拟角色的所述角色位置的至少两种高度为起始点，向多个方向发射环境感知射线；其中，在所述环境感知射线碰撞到障碍物表面的情况下，沿所述环境感知射线的法线方向反射，相同高度发射的不同所述环境感知射线的

发射方向位于同一水平高度；根据所述环境感知射线的反射情况，生成所述第一模态感知信息。

可选的，所述获取模块 1801，用于：以所述第一虚拟角色为起始点，向所述第一虚拟角色的朝向方向发射环境感知射线，其中，在所述环境感知射线碰撞到障碍物表面的情况下，沿所述环境感知射线的法线方向反射；根据所述环境感知射线的反射情况，生成所述第二模态感知信息。

可选的，所述获取模块 1801，用于：采用占栅网格划分所述第一虚拟角色预设范围内的所述虚拟环境，所述占栅网格由体积相同的多个立方体组成；从所述虚拟环境中至少两个方向的起始点向所述占栅网格内发射环境感知射线，基于所述环境感知射线的反射情况，生成所述第三模态感知信息。可选的，通过动作决策模型对所述第一虚拟角色进行动作决策，所述动作决策模型包括编码网络以及决策网络；

所述编码模块 1802，用于：将所述角色状态信息以及所述环境感知信息输入所述动作决策模型，通过所述编码网络编码所述角色状态信息以及所述环境感知信息，得到所述角色状态信息对应的状态信息编码结果和所述环境感知信息对应的感知信息编码结果；拼接所述状态信息编码结果以及所述感知信息编码结果，得到拼接特征编码；通过所述编码网络特征提取所述拼接特征编码，得到所述融合特征。

可选的，所述编码网络包括状态信息编码器以及感知信息编码器，所述感知信息编码器中包含多个维度的卷积核，所述多个维度的卷积核分别用于对多个模态的感知信息进行特征编码；

所述编码模块 1802，用于：通过所述状态信息编码器编码所述角色状态信息，得到所述角色状态信息对应的所述状态信息编码结果；通过所述感知信息编码器中的一维卷积核编码第一模态感知信息，得到第一感知信息编码结果；通过所述感知信息编码器中的二维卷积核编码第二模态感知信息，得到第二感知信息编码结果；通过所述感知信息编码器中的三维卷积核编码第三模态感知信息，得到第三感知信息编码结果，所述第一感知信息编码结果、所述第二感知信息编码结果以及所述第三感知信息编码结果构成所述感知信息编码结果。

可选的，所述角色状态信息包含绝对状态信息，所述绝对状态信息用于表征所述第一虚拟角色在所述虚拟环境中与其他虚拟角色或障碍物无关的状态信息；

所述装置还包括，泛化模块，用于：对所述角色状态信息中包含的所述绝对状态信息进行泛化处理，得到相对状态信息，所述相对状态信息是所述第一虚拟角色相对于其他虚拟角色或者障碍物的状态信息。

可选的，所述决策网络中包含 n 个动作输出头， n 个所述动作输出头用于串行输出 n 个对象子动作，不同动作输出头对应不同动作类型，且 n 个所述动作输出头基于所述动作类型之间的依赖关系串行连接，所述依赖关系用于表征不同动作类型下子动作之间的依赖限制情况， n 个所述对象子动作构成所述角色动作， n 为大于 2 的正整数；

所述决策模块 1803，包括：将所述融合特征输入决策网络中的第一动作输出头，通过所述决策网络中所述第一动作输出头从第一动作类型的子动作中确定第一对象子动作；将已确定的第一对象子动作至第 $i-1$ 对象子动作对应的嵌入编码向量，以及所述融合特征输入第 i 动作输出头，通过所述第 i 动作输出头从第 i 动作类型的子动作中确定第 i 对象子动作， i 为大于 1 的正整数。可选的， n 个所述动作输出头中存在至少两个所述动作输出头对应动作类型相同，存在所述第 i 动作输出头对应粗粒度的子动作，存在第 j 动作输出头对应细粒度的子动作， i 小于 j ， j 为大于 2 的正整数。

所述决策模块 1803，包括：通过所述决策网络中所述第 i 动作输出头从粗粒度的所述子动作中确定所述第 i 对象子动作；基于所述融合向量以及所述第 i 对象子动作对应的所述嵌入编码向量，通过所述第 j 动作输出头从细粒度的所述子动作中确定第 j 对象子动作，所述第 j 对象子动作所指示的动作范围小于所述第 i 角色动作所指示的动作范围。

可选的，所述装置还包括：

动作遮罩模块，用于：基于已确定的所述第一对象子动作至所述第*i*-1对象子动作，以及不同动作类型中不同子动作之间的依赖限制情况，对所述第*i*动作输出头对应的第*i*动作类型中的子动作进行动作遮罩，所述依赖限制情况指示的具有依赖限制关系的至少两个子动作无法同时执行，所述动作遮罩用于屏蔽不合理的子动作或者无效的子动作；

所述决策模块 1803，包括：通过所述第*i*动作输出头，从所述第*i*动作类型中未被遮罩的子动作中确定所述第*i*对象子动作。

可选的，所述装置还包括：

确定模块，用于以所述第一虚拟角色的眼部位置为起始点，向所述第一虚拟角色朝向方向上的其他虚拟角色的各个部位发射环境感知射线：根据所述环境感知射线的反射情况，确定所述虚拟环境中其他虚拟角色的各个部位的可视情况，所述可视情况用于表征其他虚拟角色各个部位是否被掩体遮挡；

所述遮罩模块，用于基于其他虚拟角色的各个部位的所述可视情况，对角色动作类型中的子动作进行动作遮罩，其中，对象子动作与其他虚拟角色的各个部位的可视情况具有所述依赖限制关系；

所述决策模块 1803，包括通过角色动作输出头，从所述角色动作类型中未被遮罩的子动作中确定对象子动作。

可选的，*n*个所述动作输出头中存在所述第一转向动作输出头对应粗粒度转向动作，以及存在第二转向动作输出头对应细粒度转向动作，所述第一转向动作输出头用于确定第一转向角度，所述第二转向动作输出头用于确定第二转向角度；

所述确定模块，还用于：基于所述第一虚拟角色的朝向、所述第一虚拟角色所处位置坐标以及其他虚拟角色的位置坐标，确定理想转向角度，以及确定所述第一虚拟角色与其他虚拟角色之间的瞄准距离；基于所述瞄准距离以及其他虚拟角色的宽度，确定有效瞄准范围；

所述动作遮罩模块，用于：基于所述第一转向动作输出头确定的第一转向角度，确定所述第一转向角度与所述理想转向角度的角度差值；基于所述角度差值以及所述有效瞄准范围确定有效细粒度转向角度；对所述粗粒度转向动作中的不属于所述有效细粒度转向角度的转向角度进行动作遮罩；

所述决策模块 1803，用于通过所述第二转向动作输出头，从所述细粒度转向动作中未被遮罩的所述有效细粒度转向角度中确定所述第二转向角度。

可选的，所述装置还包括：

训练模块，所述训练模块用于获取所述第一虚拟角色的样本角色状态信息以及所述第一虚拟角色对当前所处虚拟环境的样本环境感知信息，其中所述样本环境感知信息中包括多个模态的样本感知信息，所述多个模态分别代表一个信息维度；

所述训练模块，还用于通过动作决策模型的编码网络融合所述样本角色状态信息和所述样本环境感知信息分别对应的样本编码结果，得到样本融合特征，其中，所述编码网络通过不同编码方式对所述样本角色状态信息以及不同模态的所述样本环境感知信息进行特征编码；

所述训练模块，还用于基于所述样本融合特征，通过强化学习的方式训练所述动作决策模型。

可选的，所述训练模块，用于：将所述样本融合特征输入所述动作决策模型中的所述决策网络，得到预估角色动作；控制所述第一虚拟角色执行所述预估角色动作，得到所述第一虚拟角色的预估动作执行结果；基于所述第一虚拟角色的所述预估动作执行结果，确定预估动作执行奖励；基于所述预估动作执行奖励更新所述动作决策模型的模型参数。

可选的，所述预估动作执行奖励包括合理动作奖励、掩体利用奖励、可视性奖励以及走位分散奖励中的至少一项；

所述合理动作奖励是指所述第一虚拟角色执行所述预估角色动作的情况下，符合真实玩家操作逻辑的奖励；所述掩体利用奖励是指所述第一虚拟角色执行所述预估角色动作的情况下，所处位置与掩体相邻的奖励；所述可视性奖励是指所述虚拟角色执行所述预估角色动作的情况下，所处位置为对其他虚拟角色的有效攻击位置的奖励；所述走位分散奖励是指所述第一虚拟角色执行所述预估角色动作的情况下，连续多次走位的紧密程度降低的奖励。

可选的，所述预估动作执行结果包括所述第一虚拟角色执行前 $n+1$ 次动作后的位置点；

所述训练模块，用于基于所述第一虚拟角色执行第 $n+1$ 次所述预估角色动作后的位置点，确定执行第 $n+1$ 次动作后的位置点与前 n 次执行动作后的位置点的第一紧密度，以及确定执行第 n 次动作后的位置点与前 $n-1$ 次执行动作后位置点的第二紧密度；

在所述第一紧密度小于所述第二紧密度的情况下，将所述走位分散奖励确定为所述预估动作执行奖励。

可选的，所述预估动作执行结果包括所述第一虚拟角色执行所述预估角色动作后的所述样本环境感知信息，所述样本环境感知信息中包含二维深度图，所述二维深度图用于表征所述第一虚拟角色朝向方向上的障碍物对第一虚拟角色的遮掩情况；

所述训练模块，用于：对所述二维深度图进行区域划分，得到至少两个深度区域；确定第一深度区域的第一遮掩率，以及与所述第一深度区域相邻的至少两个第二深度区域的第二遮掩率，所述第一深度区域位于所述第一虚拟角色的视野中心；基于所述第一遮掩率与所述第二遮掩率的最小差值，确定所述掩体利用奖励，所述掩体利用奖励的奖励程度与所述最小差值呈正相关关系；将所述掩体利用奖励确定为所述预估动作执行奖励。

所述预估动作执行结果包括所述第一虚拟角色与其他虚拟角色之间的相互可视关系，所述相互可视关系用于表征所述第一虚拟角色以及其他虚拟角色相对于彼此的可视情况；

所述训练模块，用于：在所述相互可视关系指示所述第一虚拟角色不在其他虚拟角色的可视范围，且其他虚拟角色在所述第一虚拟角色的可视范围的情况下，确定所述预估动作执行奖励为正向的所述可视性奖励；在所述相互可视关系指示所述第一虚拟角色在其他虚拟角色的可视范围，且其他虚拟角色不在所述第一虚拟角色的可视范围的情况下，确定所述预估动作执行奖励为反向的所述可视性奖励。

所述训练模块，还用于基于所述第一虚拟角色的所述样本角色状态信息以及所述样本环境感知信息，确定所述第一虚拟角色的理想动作；在所述预估角色动作与所述理想动作一致的情况下，将所述合理动作奖励确定为所述预估动作执行奖励。

可选的，所述角色动作中存在转向子动作；

所述训练模块，用于：以所述第一虚拟角色为起始点向多个方向发射环境感知射线，不同所述环境感知射线位于同一水平高度，所述环境感知射线碰撞到障碍物表面的情况下，沿所述环境感知射线的法线方向反射；在所述环境感知射线的反射情况指示存在第一环境感知射线的投射距离小于距离阈值的情况下，将与所述第一环境感知射线相邻，且投射距离大于所述距离阈值的所述环境感知射线的发射方向确定为理想转向方向，所述投射距离是指所述环境感知射线从起始点到反射点之间的距离；在所述转向子动作指示的转向方向与所述理想转向方向一致的情况下，将所述合理动作奖励确定为所述预估动作执行奖励。

可选的，所述第一虚拟角色所处对局中存在至少一个其他虚拟角色，且其他虚拟角色执行所述第二动作决策模型输出的动作；

所述训练模块，用于：在所述第一虚拟角色对应的第一动作决策模型完成一轮训练的情况下，基于训练后的第一动作决策模型对所述第二动作决策模型进行训练，并将完成一轮训练后的所述第二动作决策模型存储至对手模型池；从所述对手模型池中选择训练过的所述第二动作决策模型，并构建全新对局，所述全新对局中包含执行所述第一动作决策模型输出的预估角色动作的所述第一虚拟角色，以及执行所述第二动作决策模型输出动作的其他虚拟角色；在所述全新对局中对所述第一动作决策模型进行训练。

可选的,所述训练模型,用于:在完成对所述动作决策模型的第*i*轮训练的情况下,获取所述第*i*轮训练对局的对局数据;基于至少两轮训练对局的所述对局数据,确定对局指标以及拟人化指标,所述拟人化指标是指特定动作在至少所述训练对局的所述对局数据中的实际执行占比;在所述对局指标达到训练完成标准,且所述拟人化指标与真实玩家在对局中执行所述特定动作的动作执行占比相匹配的情况下,确定所述动作决策模型完成训练。

综上所述,在虚拟环境中需要对第一虚拟角色进行动作决策时,获取第一虚拟角色的角色状态信息以及多模态的环境感知信息,从而对角色状态信息以及多模态的环境感知信息进行特征编码融合,最终基于融合特征进行动作决策。在进行动作决策时获取第一虚拟角色感知到的不同信息维度的环境感知信息,有助于获取到第一虚拟角色周围更加全面、细致的三维虚拟环境,从而能够进行更加合理的动作决策。采用不同的编码方式对角色状态信息以及多个不同模态的感知信息进行特征编码,能够使得动作决策模型对当前场景下的各类信息拟合更加精准。此外,相较于相关技术中采用图像识别的方式获取环境特征的方式,通过对量化的多模态的环境感知信息进行特征提取进而融合角色状态信息的编码结果,有助于在量化环境感知信息的同时更丰富地考虑第一虚拟角色本身的角色状态信息,在节省计算资源的同时提高决策的角色动作的准确性,使得第一虚拟角色执行角色动作的效果更加逼真。

请参考图 19,其示出了本申请一个示例性实施例提供的计算机设备的结构示意图,计算机设备可实现为上述实施例中的终端或服务器。具体来讲:所述计算机设备 1900 包括中央处理单元(Central Processing Unit, CPU) 1901、包括随机存取存储器 1902 和只读存储器 1903 的系统存储器 1904,以及连接系统存储器 1904 和中央处理单元 1901 的系统总线 1905。所述计算机设备 1900 还包括帮助计算机内的各个器件之间传输信息的基本输入/输出系统(Input/Output, I/O 系统) 1906,和用于存储操作系统 1913、应用程序 1914 和其他程序模块 1915 的大容量存储设备 1907。

在一些实施例中,所述基本输入/输出系统 1906 包括有用于显示信息的显示器 1908 和用于用户输入信息的诸如鼠标、键盘之类的输入设备 1909。其中所述显示器 1908 和输入设备 1909 都通过连接到系统总线 1905 的输入输出控制器 1910 连接到中央处理单元 1901。所述基本输入/输出系统 1906 还可以包括输入输出控制器 1910 以用于接收和处理来自键盘、鼠标、或电子触控笔等多个其他设备的输入。类似地,输入输出控制器 1910 还提供输出到显示屏、打印机或其他类型的输出设备。所述大容量存储设备 1907 通过连接到系统总线 1905 的大容量存储控制器(未示出)连接到中央处理单元 1901。所述大容量存储设备 1907 及其相关联的计算机可读介质为计算机设备 1900 提供非易失性存储。也就是说,所述大容量存储设备 1907 可以包括诸如硬盘或者驱动器之类的计算机可读介质(未示出)。

存储器存储有一个或多个程序,一个或多个程序被配置成由一个或多个中央处理单元 1901 执行,一个或多个程序包含用于实现上述方法的指令,中央处理单元 1901 执行该一个或多个程序实现上述各个方法实施例提供的方法。根据本申请的各种实施例,所述计算机设备 1900 还可以通过诸如因特网等网络连接到网络上的远程计算机运行。也即计算机设备 1900 可以通过连接在所述系统总线 1905 上的网络接口单元 1911 接到网络 1912,或者说,也可以使用网络接口单元 1911 来连接到其他类型的网络或远程计算机系统(未示出)。所述存储器还包括一个或者一个以上的程序,所述一个或者一个以上程序存储于存储器中,所述一个或者一个以上程序包含用于进行本申请实施例提供的方法中由计算机设备所执行的步骤。

本申请实施例还提供一种计算机可读存储介质,该可读存储介质中存储有至少一条指令、至少一段程序、代码集或指令集,至少一条指令、至少一段程序、代码集或指令集由处理器加载并执行以实现上述任一实施例所述的虚拟角色的动作决策方法。

本申请实施例提供了一种计算机程序产品或计算机程序,该计算机程序产品或计算机程序包括计算机指令,该计算机指令存储在计算机可读存储介质中。计算机设备的处理器从计算机可读存储介质读取该计算机指令,处理器执行该计算机指令,使得该计算机设备执行上

述方面提供的虚拟角色的动作决策方法。

本领域普通技术人员可以理解上述实施例的各种方法中的全部或部分步骤是可以通程序来指令相关的硬件来完成，该程序可以存储于一计算机可读存储介质中，该计算机可读存储介质可以是上述实施例中的存储器中所包含的计算机可读存储介质；也可以是单独存在，未装配入终端中的计算机可读存储介质。该计算机可读存储介质中存储有至少一条指令、至少一段程序、代码集或指令集，所述至少一条指令、所述至少一段程序、所述代码集或指令集由处理器加载并执行以实现上述任一方法实施例所述的虚拟角色的动作决策方法。

可选地，该计算机可读存储介质可以包括：ROM、RAM、固态硬盘（Solid State Drives, SSD）或光盘等。本领域普通技术人员可以理解实现上述实施例的全部或部分步骤可以通过硬件来完成，也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，上述提到的存储介质可以是只读存储器，磁盘或光盘等。

需要说明的是，本申请所涉及的信息（包括但不限于用户设备信息、用户个人信息等）、数据（包括但不限于用于分析的数据、存储的数据、展示的数据等）以及信号，均为经用户授权或者经过各方充分授权的，且相关数据的收集、使用和处理需要遵守相关地区的相关法律法规和标准。且本申请在收集用户的相关数据之前以及在收集用户的相关数据的过程中，都可以显示提示界面、弹窗或输出语音提示信息，该提示界面、弹窗或语音提示信息用于提示用户当前正在搜集其相关数据，使得本申请仅仅在获取到用户对该提示界面或者弹窗发出的确认操作后，才开始执行获取用户相关数据的相关步骤，否则结束获取用户相关数据的相关步骤，即不获取用户的相关数据。

应当理解的是，在本文中提及的“多个”是指两个或两个以上。本文中提及的“第一”、“第二”等用于区别类似对象，而并不用于限定特定的顺序或先后次序。另外，本文中描述的步骤编号，仅示例性示出了步骤间的一种可能的执行先后顺序，在一些其它实施例中，上述步骤也可以不按照编号顺序来执行，如两个不同编号的步骤同时执行，或者两个不同编号的步骤按照与图示相反的顺序执行，本申请实施例对此不作限定。

以上所述仅为本申请的可选的实施例，并不用以限制本申请，凡在本申请的精神和原则之内，所作的任何修改、等同替换、改进等，均应包含在本申请的保护范围之内。

权利要求书

1. 一种虚拟角色的动作决策方法，由计算机设备执行，所述方法包括：

获取第一虚拟角色的角色状态信息，以及获取所述第一虚拟角色对当前所处虚拟环境的环境感知信息，其中，所述环境感知信息包括多个模态的感知信息，所述多个模态分别代表一个信息维度；

融合所述角色状态信息和所述环境感知信息分别对应的编码结果，得到融合特征，其中，所述编码结果是通过不同编码方式对所述角色状态信息以及多个感知信息分别进行特征编码得到的结果；

基于所述融合特征对所述第一虚拟角色进行动作决策，得到角色动作；

控制所述第一虚拟角色执行所述角色动作。

2. 根据权利要求1所述的方法，其中，所述环境感知信息包括第一模态感知信息、第二模态感知信息以及第三模态感知信息；

所述获取第一虚拟角色对当前所处虚拟环境的环境感知信息，包括：

基于所述第一虚拟角色在所述虚拟环境中的角色位置，通过射线探测获取所述第一模态感知信息，所述第一模态感知信息为一维向量，所述第一模态感知信息用于表征所述第一虚拟角色周围同一水平高度上障碍物的深度情况；

基于所述第一虚拟角色在所述虚拟环境中的所述角色位置以及所述第一虚拟角色的朝向，通过射线探测获取所述第二模态感知信息，所述第二模态感知信息为所述第一虚拟角色朝向方向上的二维深度图，所述第二模态感知信息用于表征所述第一虚拟角色朝向方向上障碍物的深度情况；

基于所述第一虚拟角色在所述虚拟环境中的所述角色位置，通过射线探测获取在所述第一虚拟角色的预设范围内所述虚拟环境的所述第三模态感知信息，所述第三模态感知信息为三维占栅网格图，所述第三模态感知信息用于表征所述预设范围内障碍物的空间分布情况。

3. 根据权利要求1或2所述的方法，其中，所述基于所述第一虚拟角色在所述虚拟环境中的角色位置，通过射线探测获取所述第一模态感知信息，包括：

以所述第一虚拟角色的所述角色位置的至少两种高度为起始点，向多个方向发射环境感知射线：其中，在所述环境感知射线碰撞到障碍物表面的情况下，沿所述环境感知射线的法线方向反射，相同高度发射的不同所述环境感知射线的发射方向位于同一水平高度；

根据所述环境感知射线的反射情况，生成所述第一模态感知信息。

4. 根据权利要求1至3任一所述的方法，其中，所述基于所述第一虚拟角色在所述虚拟环境中的所述角色位置以及所述第一虚拟角色的朝向，通过射线探测获取所述第二模态感知信息，包括：

以所述第一虚拟角色为起始点，向所述第一虚拟角色的朝向方向发射环境感知射线，其中，在所述环境感知射线碰撞到障碍物表面的情况下，沿所述环境感知射线的法线方向反射；

根据所述环境感知射线的反射情况，生成所述第二模态感知信息。

5. 根据权利要求1至4任一所述的方法，其中，所述基于所述第一虚拟角色在所述虚拟环境中的所述角色位置，通过射线探测获取在所述第一虚拟角色的预设范围内所述虚拟环境的所述第三模态感知信息，包括：

采用占栅网格划分所述第一虚拟角色预设范围内的所述虚拟环境，所述占栅网格由体积相同的多个立方体组成；

从所述虚拟环境中至少两个方向的起始点向所述占栅网格内发射环境感知射线，基于所述环境感知射线的反射情况，生成所述第三模态感知信息。

6. 根据权利要求1至5任一所述的方法，其中，通过动作决策模型对所述第一虚拟角色进行动作决策，所述动作决策模型包括编码网络以及决策网络；

所述融合所述角色状态信息和所述环境感知信息分别对应的编码结果,得到融合特征,包括:

将所述角色状态信息以及所述环境感知信息输入所述动作决策模型,通过所述编码网络编码所述角色状态信息以及所述环境感知信息,得到所述角色状态信息对应的状态信息编码结果和所述环境感知信息对应的感知信息编码结果;

拼接所述状态信息编码结果以及所述感知信息编码结果,得到拼接特征编码;

通过所述编码网络特征提取所述拼接特征编码,得到所述融合特征。

7. 根据权利要求1至6任一所述的方法,其中,所述编码网络包括状态信息编码器以及感知信息编码器,所述感知信息编码器中包含多个维度的卷积核,所述多个维度的卷积核分别用于对多个模态的感知信息进行特征编码;

所述通过所述编码网络编码所述角色状态信息以及所述环境感知信息,得到所述角色状态信息对应的状态信息编码结果和所述环境感知信息对应的感知信息编码结果,包括:

通过所述状态信息编码器编码所述角色状态信息,得到所述角色状态信息对应的所述状态信息编码结果;

通过所述感知信息编码器中的一维卷积核编码第一模态感知信息,得到第一感知信息编码结果;

通过所述感知信息编码器中的二维卷积核编码第二模态感知信息,得到第二感知信息编码结果;

通过所述感知信息编码器中的三维卷积核编码第三模态感知信息,得到第三感知信息编码结果,所述第一感知信息编码结果、所述第二感知信息编码结果以及所述第三感知信息编码结果构成所述感知信息编码结果。

8. 根据权利要求1至7任一所述的方法,其中,所述角色状态信息包含绝对状态信息,所述绝对状态信息用于表征所述第一虚拟角色在所述虚拟环境中与其他虚拟角色或障碍物无关的状态信息;

所述获取第一虚拟角色的角色状态信息之后,所述方法还包括:

对所述角色状态信息中包含的所述绝对状态信息进行泛化处理,得到相对状态信息,所述相对状态信息是所述第一虚拟角色相对于其他虚拟角色或者障碍物的状态信息。

9. 根据权利要求1至8任一所述的方法,其中,所述决策网络中包含 n 个动作输出头, n 个所述动作输出头用于串行输出 n 个对象子动作,不同动作输出头对应不同动作类型,且 n 个所述动作输出头基于所述动作类型之间的依赖关系串行连接,所述依赖关系用于表征不同动作类型下子动作之间的依赖限制情况, n 个所述对象子动作构成所述角色动作, n 为大于2的正整数;

所述基于所述融合特征对所述第一虚拟角色进行动作决策,得到角色动作,包括:

将所述融合特征输入决策网络中的第一动作输出头,通过所述决策网络中所述第一动作输出头从第一动作类型的子动作中确定第一对象子动作;

将已确定的第一对象子动作至第 $i-1$ 对象子动作对应的嵌入编码向量,以及所述融合特征输入第 i 动作输出头,通过所述第 i 动作输出头从第 i 动作类型的子动作中确定第 i 对象子动作, i 为大于1的正整数。

10. 根据权利要求1至9任一所述的方法,其中, n 个所述动作输出头中存在至少两个所述动作输出头对应动作类型相同,存在所述第 i 动作输出头对应粗粒度的子动作,存在第 j 动作输出头对应细粒度的子动作, i 小于 j , j 为大于2的正整数;

所述通过所述第 i 动作输出头从第 i 动作类型的子动作中确定第 i 对象子动作,包括:

通过所述决策网络中所述第 i 动作输出头从粗粒度的所述子动作中确定所述第 i 对象子动作;

所述方法还包括:

基于所述融合向量以及所述第*i*对象子动作对应的所述嵌入编码向量,通过所述第*j*动作输出头从细粒度的所述子动作中确定第*j*对象子动作,所述第*j*对象子动作所指示的动作范围小于所述第*i*角色动作所指示的动作范围。

11. 根据权利要求1至10任一所述的方法,其中,所述通过所述第*i*动作输出头从第*i*动作类型的子动作中确定第*i*对象子动作之前,所述方法还包括:

基于已确定的所述第一对象子动作至所述第*i*-1对象子动作,以及不同动作类型中不同子动作之间的依赖限制情况,对所述第*i*动作输出头对应的第*i*动作类型中的子动作进行动作遮罩,所述依赖限制情况指示的具有依赖限制关系的至少两个子动作无法同时执行,所述动作遮罩用于屏蔽不合理的子动作或者无效的子动作;

所述通过所述第*i*动作输出头从第*i*动作类型的子动作中确定第*i*对象子动作,包括:

通过所述第*i*动作输出头,从所述第*i*动作类型中未被遮罩的子动作中确定所述第*i*对象子动作。

12. 根据权利要求1至11任一所述的方法,其中,所述方法还包括:

以所述第一虚拟角色的眼部位置为起始点,向所述第一虚拟角色朝向方向上的其他虚拟角色的各个部位发射环境感知射线;

根据所述环境感知射线的反射情况,确定所述虚拟环境中其他虚拟角色的各个部位的可视情况,所述可视情况用于表征其他虚拟角色各个部位是否被掩体遮挡;

所述对所述第*i*动作输出头对应的第*i*动作类型中的子动作进行动作遮罩,包括:

基于其他虚拟角色的各个部位的所述可视情况,对角色动作类型中的子动作进行动作遮罩,其中,对象子动作与其他虚拟角色的各个部位的可视情况具有所述依赖限制关系;

所述通过所述第*i*动作输出头从第*i*动作类型的子动作中确定第*i*对象子动作,包括:

通过角色动作输出头,从所述角色动作类型中未被遮罩的子动作中确定对象子动作。

13. 根据权利要求1至12任一所述的方法,其中,*n*个所述动作输出头中存在所述第一转向动作输出头对应粗粒度转向动作,以及存在第二转向动作输出头对应细粒度转向动作,所述第一转向动作输出头用于确定第一转向角度,所述第二转向动作输出头用于确定第二转向角度;

在所述第一虚拟角色具有瞄准其他虚拟角色的情况下,所述方法还包括:

基于所述第一虚拟角色的朝向、所述第一虚拟角色所处位置坐标以及其他虚拟角色的位置坐标,确定理想转向角度,以及确定所述第一虚拟角色与其他虚拟角色之间的瞄准距离;

基于所述瞄准距离以及其他虚拟角色的宽度,确定有效瞄准范围;

所述对所述第*i*动作输出头对应的第*i*动作类型中的子动作进行动作遮罩,包括:

基于所述第一转向动作输出头确定的第一转向角度,确定所述第一转向角度与所述理想转向角度的角度差值;

基于所述角度差值以及所述有效瞄准范围确定有效细粒度转向角度;

对所述粗粒度转向动作中不属于所述有效细粒度转向角度的转向角度进行动作遮罩;

所述通过所述第*i*动作输出头从第*i*动作类型的子动作中确定第*i*对象子动作,包括:

通过所述第二转向动作输出头,从所述细粒度转向动作中未被遮罩的所述有效细粒度转向角度中确定所述第二转向角度。

14. 根据权利要求1至13任一所述的方法,其中,所述获取第一虚拟角色的角色状态信息之前,所述方法包括:

获取所述第一虚拟角色的样本角色状态信息以及所述第一虚拟角色对当前所处虚拟环境的样本环境感知信息,其中所述样本环境感知信息中包括多个模态的样本感知信息,所述多个模态分别代表一个信息维度;

通过动作决策模型的编码网络融合所述样本角色状态信息和所述样本环境感知信息分别对应的样本编码结果,得到样本融合特征,其中,所述编码网络通过不同编码方式对所述样

本角色状态信息以及不同模态的所述样本环境感知信息进行特征编码；

基于所述样本融合特征，通过强化学习的方式训练所述动作决策模型。

15. 根据权利要求1至14任一所述的方法，其中，所述基于所述样本融合特征，通过强化学习的方式训练所述动作决策模型，包括：

将所述样本融合特征输入所述动作决策模型中的所述决策网络，得到预估角色动作；

控制所述第一虚拟角色执行所述预估角色动作，得到所述第一虚拟角色的预估动作执行结果；

基于所述第一虚拟角色的所述预估动作执行结果，确定预估动作执行奖励；

基于所述预估动作执行奖励更新所述动作决策模型的模型参数。

16. 根据权利要求1至15任一所述的方法，其中，所述预估动作执行奖励包括合理动作奖励、掩体利用奖励、可视性奖励以及走位分散奖励中的至少一项；

所述合理动作奖励是指所述第一虚拟角色执行所述预估角色动作的情况下，符合真实玩家操作逻辑的奖励；

所述掩体利用奖励是指所述第一虚拟角色执行所述预估角色动作的情况下，所处位置与掩体相邻的奖励；

所述可视性奖励是指所述虚拟角色执行所述预估角色动作的情况下，所处位置为对其他虚拟角色的有效攻击位置的奖励；

所述走位分散奖励是指所述第一虚拟角色执行所述预估角色动作的情况下，连续多次走位的紧密程度降低的奖励。

17. 根据权利要求1至16任一所述的方法，其中，所述预估动作执行结果包括所述第一虚拟角色执行前 $n+1$ 次动作后的位置点；

所述基于所述第一虚拟角色的预估动作执行结果，确定预估动作执行奖励，包括：

基于所述第一虚拟角色执行第 $n+1$ 次所述预估角色动作后的位置点，确定执行第 $n+1$ 次动作后的位置点与前 n 次执行动作后的位置点的第一紧密度，以及确定执行第 n 次动作后的位置点与前 $n-1$ 次执行动作后位置点的第二紧密度；

在所述第一紧密度小于所述第二紧密度的情况下，将所述走位分散奖励确定为所述预估动作执行奖励。

18. 根据权利要求1至17任一所述的方法，其中，所述预估动作执行结果包括所述第一虚拟角色执行所述预估角色动作后的所述样本环境感知信息，所述样本环境感知信息中包含二维深度图，所述二维深度图用于表征所述第一虚拟角色朝向方向上的障碍物对第一虚拟角色的遮掩情况；

所述基于所述第一虚拟角色的所述预估动作执行结果，确定预估动作执行奖励，包括：

对所述二维深度图进行区域划分，得到至少两个深度区域；

确定第一深度区域的第一遮掩率，以及与所述第一深度区域相邻的至少两个第二深度区域的第二遮掩率，所述第一深度区域位于所述第一虚拟角色的视野中心；

基于所述第一遮掩率与所述第二遮掩率的最小差值，确定所述掩体利用奖励，所述掩体利用奖励的奖励程度与所述最小差值呈正相关关系；

将所述掩体利用奖励确定为所述预估动作执行奖励。

19. 根据权利要求1至18任一所述的方法，其中，所述预估动作执行结果包括所述第一虚拟角色与其他虚拟角色之间的相互可视关系，所述相互可视关系用于表征所述第一虚拟角色以及其他虚拟角色相对于彼此的可视情况；

所述基于所述第一虚拟角色的所述预估动作执行结果，确定预估动作执行奖励，包括：

在所述相互可视关系指示所述第一虚拟角色不在其他虚拟角色的可视范围，且其他虚拟角色在所述第一虚拟角色的可视范围的情况下，确定所述预估动作执行奖励为正向的所述可视性奖励；

在所述相互可视关系指示所述第一虚拟角色在其他虚拟角色的可视范围，且其他虚拟角色不在所述第一虚拟角色的可视范围的情况下，确定所述预估动作执行奖励为反向的所述可视性奖励。

20. 根据权利要求 1 至 19 任一所述的方法，其中，所述方法还包括：

基于所述第一虚拟角色的所述样本角色状态信息以及所述样本环境感知信息，确定所述第一虚拟角色的理想动作；

所述基于所述第一虚拟角色的所述预估动作执行结果，确定预估动作执行奖励，包括：

在所述预估角色动作与所述理想动作一致的情况下，将所述合理动作奖励确定为所述预估动作执行奖励。

21. 根据权利要求 1 至 20 任一所述的方法，其中，所述角色动作中存在转向子动作：

所述基于所述第一虚拟角色的所述样本角色状态信息以及所述样本环境感知信息，确定所述第一虚拟角色的理想动作，包括：

以所述第一虚拟角色为起始点向多个方向发射环境感知射线，不同所述环境感知射线位于同一水平高度，所述环境感知射线碰撞到障碍物表面的情况下，沿所述环境感知射线的法线方向反射；

在所述环境感知射线的反射情况指示存在第一环境感知射线的投射距离小于距离阈值的情况下，将与所述第一环境感知射线相邻，且投射距离大于所述距离阈值的所述环境感知射线的发射方向确定为理想转向方向，所述投射距离是指所述环境感知射线从起始点到反射点之间的距离；

所述在所述预估角色动作与所述理想动作一致的情况下，将所述合理动作奖励确定为所述预估动作执行奖励，包括：

在所述转向子动作指示的转向方向与所述理想转向方向一致的情况下，将所述合理动作奖励确定为所述预估动作执行奖励。

22. 根据权利要求 1 至 21 任一所述的方法，其中，所述第一虚拟角色所处对局中存在至少一个其他虚拟角色，且其他虚拟角色执行所述第二动作决策模型输出的动作；

所述方法还包括：

在所述第一虚拟角色对应的第一动作决策模型完成一轮训练的情况下，基于训练后的第一动作决策模型对所述第二动作决策模型进行训练，并将完成一轮训练后的所述第二动作决策模型存储至对手模型池；

从所述对手模型池中选择训练过的所述第二动作决策模型，并构建全新对局，所述全新对局中包含执行所述第一动作决策模型输出的预估角色动作的所述第一虚拟角色，以及执行所述第二动作决策模型输出动作的其他虚拟角色；

在所述全新对局中对所述第一动作决策模型进行训练。

23. 根据权利要求 1 至 22 任一所述的方法，其中，所述方法还包括：

在完成对所述动作决策模型的第*i*轮训练的情况下，获取所述第*i*轮训练对局的对局数据：

基于至少两轮训练对局的所述对局数据，确定对局指标以及拟人化指标，所述拟人化指标是指特定动作在至少所述训练对局的所述对局数据中的实际执行占比；

在所述对局指标达到训练完成标准，且所述拟人化指标与真实玩家在对局中执行所述特定动作的动作执行占比相匹配的情况下，确定所述动作决策模型完成训练。

24. 一种虚拟角色的动作决策装置，所述装置包括：

获取模块，用于获取第一虚拟角色的角色状态信息，以及获取所述第一虚拟角色对当前所处虚拟环境的环境感知信息，其中，所述环境感知信息包括多个模态的感知信息，所述多个模态分别代表一个信息维度；

编码模块，用于融合所述角色状态信息和所述环境感知信息分别对应的编码结果，得到融合特征，其中，所述编码结果是通过不同编码方式对所述角色状态信息以及多个感知信息

分别进行特征编码得到的结果；

决策模块，用于基于所述融合特征对所述第一虚拟角色进行动作决策，得到角色动作；

控制模块，用于控制所述第一虚拟角色执行所述角色动作。

25. 一种计算机设备，所述计算机设备包括处理器和存储器，所述存储器中存储有至少一段程序，所述至少一段程序由所述处理器加载并执行以实现如权利要求 1 至 23 任一所述的虚拟角色的动作决策方法。

26. 一种计算机可读存储介质，所述可读存储介质中存储有至少一段程序，所述至少一段程序由处理器加载并执行以实现如权利要求 1 至 23 任一所述的虚拟角色的动作决策方法。

27. 一种计算机程序产品，所述计算机程序产品包括计算机指令，所述计算机指令存储在计算机可读存储介质中；计算机设备的处理器从所述计算机可读存储介质读取所述计算机指令，所述处理器执行所述计算机指令，使得所述计算机设备执行如权利要求 1 至 23 任一所述的虚拟角色的动作决策方法。