



(12) 发明专利

(10) 授权公告号 CN 112019297 B

(45) 授权公告日 2021. 03. 30

(21) 申请号 202011176052.6

审查员 董玉慧

(22) 申请日 2020.10.29

(65) 同一申请的已公布的文献号

申请公布号 CN 112019297 A

(43) 申请公布日 2020.12.01

(73) 专利权人 上海特金无线技术有限公司

地址 201114 上海市闵行区新骏环路245号
第6层E612室

(72) 发明人 刘鑫 姜化京 姜维

(74) 专利代理机构 上海慧晗知识产权代理事务
所(普通合伙) 31343

代理人 徐海晟

(51) Int. Cl.

H04K 3/00 (2006.01)

G08G 5/00 (2006.01)

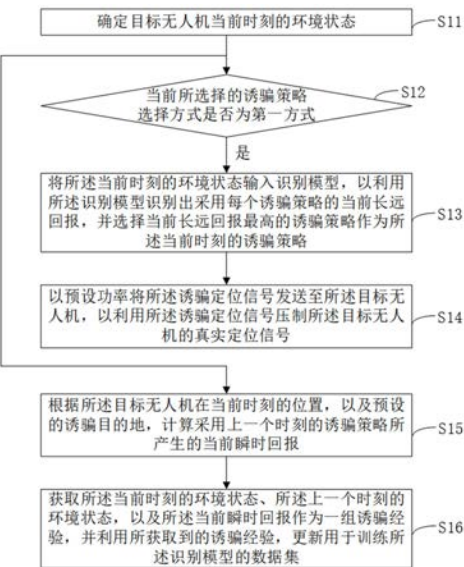
权利要求书4页 说明书12页 附图7页

(54) 发明名称

无人机定点诱骗方法、装置、电子设备与存储介质

(57) 摘要

本发明提供了一种无人机定点诱骗方法、装置、电子设备与存储介质,其中的无人机定点诱骗方法,包括:确定目标无人机当前时刻的环境状态,所述环境状态记载了对应时刻的位置以及该时刻之前多个时刻的位置;将所述当前时刻的环境状态输入识别模型,以利用所述识别模型识别出采用每个诱骗策略的当前长远回报,并选择当前长远回报最高的诱骗策略作为所述当前时刻的诱骗策略;以预设功率将所述诱骗定位信号发送至所述目标无人机,以利用所述诱骗定位信号压制所述目标无人机的真实定位信号。可见,本发明可基于环境状态与诱骗目的地确定瞬时回报,最终将目标无人机逐步诱骗到诱骗目的地或其附近区域,解决了无人机反制对周边造成不利影响的问题。



1. 一种无人机定点诱骗方法,应用于诱骗设备,其特征在于,包括:

确定目标无人机当前时刻的环境状态,所述环境状态记载了对应时刻的位置以及该时刻之前多个时刻的位置;所述目标无人机为进入监管区域的无人机;

若当前所选择的诱骗策略选择方式为第一方式,则将所述当前时刻的环境状态输入识别模型,以利用所述识别模型识别出采用每个诱骗策略的当前长远回报,并选择当前长远回报最高的诱骗策略作为所述当前时刻的诱骗策略,以实现所述第一方式;其中,不同诱骗策略对应不同的定位信号,所述当前时刻的诱骗策略所对应的定位信号为所述诱骗定位信号;

以预设功率将所述诱骗定位信号发送至所述目标无人机,以利用所述诱骗定位信号压制所述目标无人机的真实定位信号;

根据所述目标无人机在当前时刻的位置,以及预设的诱骗目的地,计算采用上一个时刻的诱骗策略所产生的当前瞬时回报;

获取所述当前时刻的环境状态、所述上一个时刻的环境状态,以及所述当前瞬时回报作为一组诱骗经验,并利用所获取到的诱骗经验,更新用于训练所述识别模型的数据集;

其中,根据所述目标无人机在当前时刻的位置,以及预设的诱骗目的地,计算采用上一个时刻的诱骗策略所产生的当前瞬时回报,包括:

根据所述诱骗目的地相对于所述目标无人机当前位置的方位,以及所述目标无人机的当前速度方向,确定当前方位收益;所述当前速度方向表征了在所述当前时刻,所述目标无人机的飞行速度的方向

根据所述目标无人机当前位置与所述诱骗目的地的当前距离,确定当前距离收益;

根据所述当前方位收益与所述当前距离收益,确定所述当前瞬时回报;

所述识别模型为卷积神经网络,所识别出的长远回报是利用Q函数的Q值来表征的;

所述Q函数的定义式为:

$$Q^*(S, a) = E \left\{ r + \gamma \max_{a'} Q^*(S, a) \mid S, a \right\}$$

其中:

γ 表示奖励折扣系数;

$Q^*(\cdot)$ 表示Q函数的Q值;

r 表示瞬时回报;

a 表示诱骗策略;

a' 表示下一个状态的诱骗策略;

S 表示环境状态。

2. 根据权利要求1所述的无人机定点诱骗方法,其特征在于,所述当前瞬时回报是通过以下公式确定的:

$$r_t = (1 - \alpha)A(\theta) + \alpha B(L);$$

$$A(\theta) = \frac{90 - \theta}{90};$$

$$\theta = |\theta_r - \theta_w|, \theta \in (0, 180);$$

$$B(L) = e^{1-L};$$

$$\alpha = \frac{1}{L};$$

其中:

r_t 表征了当前瞬时回报;

θ 表征了所述诱骗目的地相对于所述目标无人机当前位置的方向 θ_r 与所述目标无人机的当前速度方向 θ_w 之间的夹角;

$A(\theta)$ 表征了当前方位收益;

L 表征了所述目标无人机当前位置与所述诱骗目的地的当前距离;

$B(L)$ 表征了当前距离收益。

3. 根据权利要求1或2所述的无人机定点诱骗方法, 其特征在于, 确定目标无人机当前时刻的环境状态之后, 还包括:

根据选择概率信息, 选择所述第一方式或第二方式作为当前的诱骗策略选择方式, 其中, 所述第一方式的选择概率与所述第二方式的选择概率之和为1;

若选择所述第二方式, 则随机自多个诱骗策略中选择一个作为当前时刻的诱骗策略, 以实现所述第二方式。

4. 根据权利要求3所述的无人机定点诱骗方法, 其特征在于, 所述识别模型采用卷积神经网络, 所述选择概率信息中, 所述第一方式的选择概率为固定概率值与动态概率值中较大的一个概率值, 所述动态概率值是随所述卷积神经网络中当前更新步长的衰减系数确定的。

5. 根据权利要求1或2所述的无人机定点诱骗方法, 其特征在于, 所述数据集中记录了所获取到的至少部分诱骗经验;

所述识别模型是根据以下过程训练的:

在所述数据集中所记录的各诱骗经验中, 随机抽取诱骗经验;

根据所抽取的诱骗经验, 构建对应环境状态的目标Q值;

根据所述目标Q值与对应环境状态的真实Q值, 更新所述卷积神经网络。

6. 一种无人机定点诱骗装置, 应用于诱骗设备, 其特征在于, 包括:

环境状态确定模块, 用于确定目标无人机当前时刻的环境状态, 所述环境状态记载了对应时刻的位置以及该时刻之前多个时刻的位置; 所述目标无人机为进入监管区域的无人机;

诱骗策略第一选择模块, 用于若当前所选择的诱骗策略选择方式为第一方式, 则将所述当前时刻的环境状态输入识别模型, 以利用所述识别模型识别出采用每个诱骗策略的当前长远回报, 并选择当前长远回报最高的诱骗策略作为所述当前时刻的诱骗策略, 以实现所述第一方式; 其中, 不同诱骗策略对应不同的定位信号, 所述当前时刻的诱骗策略所对应的定位信号为所述诱骗定位信号;

诱骗模块, 用于以预设功率将所述诱骗定位信号发送至所述目标无人机, 以利用所述诱骗定位信号压制所述目标无人机的真实定位信号;

瞬时回报计算模块,用于根据所述目标无人机在当前时刻的位置,以及预设的诱骗目的地,计算采用上一个时刻的诱骗策略所产生的当前瞬时回报;

数据集更新模块,用于获取所述当前时刻的环境状态、所述上一个时刻的环境状态,以及所述当前瞬时回报作为一组诱骗经验,并利用所获取到的诱骗经验,更新用于训练所述识别模型的数据集;

其中,所述瞬时回报计算模块,具体用于:

根据所述诱骗目的地相对于所述目标无人机当前位置的方位,以及所述目标无人机的当前速度方向,确定当前方位收益;所述当前速度方向表征了在所述当前时刻,所述目标无人机的飞行速度的方向

根据所述目标无人机当前位置与所述诱骗目的地的当前距离,确定当前距离收益;

根据所述当前方位收益与所述当前距离收益,确定所述当前瞬时回报;

所述识别模型为卷积神经网络,所识别出的长远回报是利用Q函数的Q值来表征的;

所述Q函数的定义式为:

$$Q^*(S, a) = E \left\{ r + \gamma \max_{a'} Q^*(S, a) \mid S, a \right\}$$

其中:

γ 表示奖励折扣系数;

$Q^*(\cdot)$ 表示Q函数的Q值;

r 表示瞬时回报;

a 表示诱骗策略;

a' 表示下一个状态的诱骗策略;

S 表示环境状态。

7.根据权利要求6所述的无人机定点诱骗装置,其特征在于,所述当前瞬时回报是通过以下公式确定的:

$$r_t = (1 - \alpha)A(\theta) + \alpha B(L);$$

$$A(\theta) = \frac{90 - \theta}{90};$$

$$\theta = |\theta_r - \theta_w|, \theta \in (0, 180);$$

$$B(L) = e^{1-L};$$

$$\alpha = \frac{1}{L};$$

其中:

r_t 表征了当前瞬时回报;

θ 表征了所述诱骗目的地相对于所述目标无人机当前位置的方向 θ_r 与所述目标无人机的当前速度方向 θ_w 之间的夹角;

$A(\theta)$ 表征了当前方位收益;

L 表征了所述目标无人机当前位置与所述诱骗目的地的当前距离;

$B(L)$ 表征了当前距离收益。

8. 根据权利要求6或7所述的无人机定点诱骗装置, 其特征在于, 还包括:

方式选择模块, 用于根据选择概率信息, 选择所述第一方式或第二方式作为当前的诱骗策略选择方式, 其中, 所述第一方式的选择概率与所述第二方式的选择概率之和为1;

诱骗策略第二选择模块, 用于若选择所述第二方式, 则随机自多个诱骗策略中选择一个作为当前时刻的诱骗策略, 以实现所述第二方式。

9. 根据权利要求8所述的无人机定点诱骗装置, 其特征在于, 所述识别模型采用卷积神经网络, 所述选择概率信息中, 所述第一方式的选择概率为固定概率值与动态概率值中较大的一个概率值, 所述动态概率值是随所述卷积神经网络中当前更新步长的衰减系数确定的。

10. 根据权利要求6或7所述的无人机定点诱骗装置, 其特征在于, 所述数据集中记录了所获取到的至少部分诱骗经验;

所述识别模型是根据以下过程训练的:

在所述数据集中所记录的各诱骗经验中, 随机抽取诱骗经验;

根据所抽取的诱骗经验, 构建对应环境状态的目标Q值;

根据所述目标Q值与对应环境状态的真实Q值, 更新所述卷积神经网络。

11. 一种电子设备, 其特征在于, 包括处理器与存储器,

所述存储器, 用于存储代码和相关数据;

所述处理器, 用于执行所述存储器中的代码用以实现权利要求1至5任一项所述的方法。

12. 一种存储介质, 其上存储有计算机程序, 该程序被处理器执行时实现权利要求1至5任一项所述的方法。

无人机定点诱骗方法、装置、电子设备与存储介质

技术领域

[0001] 本发明涉及无人机领域,尤其涉及一种无人机定点诱骗方法、装置、电子设备与存储介质。

背景技术

[0002] 近几年,随着信息、控制、通讯等领域技术的不断发展,使得现代无人机具有小型化、轻量化、低成本、适应性强等优点,然而,由于无人机本身技术原因,飞行进入监管区域的无人机会带来安全威胁。

[0003] 现有的相关技术中,为了反制目标无人机,通常采用电磁信号干扰、声波干扰、激光武器、黑客入侵、物理捕获等,这些技术研制的反无人机系统大致可分为干扰阻断、毁伤抓捕和监测控制三大类。

[0004] 然而,以上方案中,会对周边造成不利影响,例如:电磁信号、声波干扰会对目标无人机之外的其他设备(例如通信设备、其他无人机)造成干扰,还会对人造成辐射。

发明内容

[0005] 本发明提供一种无人机定点诱骗方法、装置、电子设备与存储介质,以解决会对周边造成不利影响的问题。

[0006] 根据本发明的第一方面,提供了一种无人机定点诱骗方法,应用于诱骗设备,包括:

[0007] 确定目标无人机当前时刻的环境状态,所述环境状态记载了对应时刻的位置以及该时刻之前多个时刻的位置;

[0008] 若当前所选择的诱骗策略选择方式为第一方式,则将所述当前时刻的环境状态输入识别模型,以利用所述识别模型识别出采用每个诱骗策略的当前长远回报,并选择当前长远回报最高的诱骗策略作为所述当前时刻的诱骗策略;其中,不同诱骗策略对应不同的定位信号,所述当前时刻的诱骗策略所对应的定位信号为所述诱骗定位信号;

[0009] 以预设功率将所述诱骗定位信号发送至所述目标无人机,以利用所述诱骗定位信号压制所述目标无人机的真实定位信号;

[0010] 根据所述目标无人机在当前时刻的位置,以及预设的诱骗目的地,计算采用上一个时刻的诱骗策略所产生的当前瞬时回报;

[0011] 获取所述当前时刻的环境状态、所述上一个时刻的环境状态,以及所述当前瞬时回报作为一组诱骗经验,并利用所获取到的诱骗经验,更新用于训练所述识别模型的数据集。

[0012] 可选的,根据所述目标无人机在当前时刻的位置,以及预设的诱骗目的地,计算采用上一个时刻的诱骗策略所产生的当前瞬时回报,包括:

[0013] 根据所述诱骗目的地相对于所述目标无人机当前位置的方位,以及所述目标无人机的当前速度方向,确定当前方位收益;所述当前速度方向表征了在所述当前时刻,所述目

标无人机的飞行速度的方向

[0014] 根据所述目标无人机当前位置与所述诱骗目的地的当前距离,确定当前距离收益;

[0015] 根据所述当前方位收益与所述当前距离收益,确定所述当前瞬时回报。

[0016] 可选的,所述当前瞬时回报是通过以下公式确定的:

$$[0017] \quad r_t = (1 - \alpha)A(\theta) + \alpha B(L);$$

$$[0018] \quad A(\theta) = \frac{90 - \theta}{90};$$

$$[0019] \quad \theta = |\theta_r - \theta_w|, \theta \in (0, 180);$$

$$[0020] \quad B(L) = e^{1-L};$$

$$[0021] \quad \alpha = \frac{1}{L};$$

[0022] 其中:

[0023] r_t 表征了当前瞬时回报;

[0024] θ 表征了所述诱骗目的地相对于所述目标无人机当前位置的方向 θ_r 与所述目标无人机的当前速度方向 θ_w 之间的夹角;

[0025] $A(\theta)$ 表征了当前方位收益;

[0026] L 表征了所述目标无人机当前位置与所述诱骗目的地的当前距离;

[0027] $B(L)$ 表征了当前距离收益。

[0028] 可选的,确定目标无人机当前时刻的环境状态之后,还包括:

[0029] 根据选择概率信息,选择所述第一方式或第二方式作为当前的诱骗策略选择方式,其中,所述第一方式的选择概率与所述第二方式的选择概率之和为1;

[0030] 若选择所述第二方式,则随机自多个诱骗策略中选择一个作为当前时刻的诱骗策略。

[0031] 可选的,所述识别模型采用卷积神经网络,所述选择概率信息中,所述第一方式的选择概率为固定概率值与动态概率值中较大的一个概率值,所述动态概率值是随所述卷积神经网络中当前更新步长的衰减系数确定的。

[0032] 可选的,所述数据集中记录了所获取到的至少部分诱骗经验;所述识别模型为卷积神经网络,所识别出的长远回报是利用Q函数的Q值来表征的;

[0033] 所述识别模型是根据以下过程训练的:

[0034] 在所述数据集中所记录的各诱骗经验中,随机抽取诱骗经验;

[0035] 根据所抽取的诱骗经验,构建对应环境状态的目标Q值;

[0036] 根据所述目标Q值与对应环境状态的真实Q值,更新所述卷积神经网络。

[0037] 根据本发明的第二方面,提供了一种无人机定点诱骗装置,应用于诱骗设备,包括:

[0038] 环境状态确定模块,用于确定目标无人机当前时刻的环境状态,所述环境状态记载了对应时刻的位置以及该时刻之前多个时刻的位置;

[0039] 诱骗策略第一选择模块,用于若当前所选择的诱骗策略选择方式为第一方式,则将所述当前时刻的环境状态输入识别模型,以利用所述识别模型识别出采用每个诱骗策略的当前长远回报,并选择当前长远回报最高的诱骗策略作为所述当前时刻的诱骗策略;其中,不同诱骗策略对应不同的定位信号,所述当前时刻的诱骗策略所对应的定位信号为所述诱骗定位信号;

[0040] 诱骗模块,用于以预设功率将所述诱骗定位信号发送至所述目标无人机,以利用所述诱骗定位信号压制所述目标无人机的真实定位信号;

[0041] 瞬时回报计算模块,用于根据所述目标无人机在当前时刻的位置,以及预设的诱骗目的地,计算采用上一个时刻的诱骗策略所产生的当前瞬时回报;

[0042] 数据集更新模块,用于获取所述当前时刻的环境状态、所述上一个时刻的环境状态,以及所述当前瞬时回报作为一组诱骗经验,并利用所获取到的诱骗经验,更新用于训练所述识别模型的数据集。

[0043] 可选的,所述瞬时回报计算模块,具体用于:

[0044] 根据所述诱骗目的地相对于所述目标无人机当前位置的方位,以及所述目标无人机的当前速度方向,确定当前方位收益;所述当前速度方向表征了在所述当前时刻,所述目标无人机的飞行速度的方向

[0045] 根据所述目标无人机当前位置与所述诱骗目的地的当前距离,确定当前距离收益;

[0046] 根据所述当前方位收益与所述当前距离收益,确定所述当前瞬时回报。

[0047] 可选的,所述当前瞬时回报是通过以下公式确定的:

$$[0048] \quad r_t = (1 - \alpha)A(\theta) + \alpha B(L);$$

$$[0049] \quad A(\theta) = \frac{90 - \theta}{90};$$

$$[0050] \quad \theta = |\theta_r - \theta_w|, \theta \in (0, 180);$$

$$[0051] \quad B(L) = e^{1-L};$$

$$[0052] \quad \alpha = \frac{1}{L};$$

[0053] 其中:

[0054] r_t 表征了当前瞬时回报;

[0055] θ 表征了所述诱骗目的地相对于所述目标无人机当前位置的方向 θ_r 与所述目标无人机的当前速度方向 θ_w 之间的夹角;

[0056] $A(\theta)$ 表征了当前方位收益;

[0057] L 表征了所述目标无人机当前位置与所述诱骗目的地的当前距离;

[0058] $B(L)$ 表征了当前距离收益。

[0059] 可选的,所述的无人机定点诱骗装置,还包括:

[0060] 方式选择模块,用于根据选择概率信息,选择所述第一方式或第二方式作为当前的诱骗策略选择方式,其中,所述第一方式的选择概率与所述第二方式的选择概率之和为

1;

[0061] 诱骗策略第二选择模块,用于若选择所述第二方式,则随机自多个诱骗策略中选择一个作为当前时刻的诱骗策略。

[0062] 可选的,所述识别模型采用卷积神经网络,所述选择概率信息中,所述第一方式的选择概率为固定概率值与动态概率值中较大的一个概率值,所述动态概率值是随所述卷积神经网络中当前更新步长的衰减系数确定的。

[0063] 可选的,所述数据集中记录了所获取到的至少部分诱骗经验;所述识别模型为卷积神经网络,所识别出的长远回报是利用Q函数的Q值来表征的;

[0064] 所述识别模型是根据以下过程训练的:

[0065] 在所述数据集中所记录的各诱骗经验中,随机抽取诱骗经验;

[0066] 根据所抽取的诱骗经验,构建对应环境状态的目标Q值;

[0067] 根据所述目标Q值与对应环境状态的真实Q值,更新所述卷积神经网络。

[0068] 根据本发明的第三方面,提供了一种电子设备,包括处理器与存储器,

[0069] 所述存储器,用于存储代码和相关数据;

[0070] 所述处理器,用于执行所述存储器中的代码用以实现第一方面及其可选方案涉及的方法。

[0071] 根据本发明的第四方面,提供了一种存储介质,其上存储有计算机程序,该程序被处理器执行时实现第一方面及其可选方案涉及的方法。

[0072] 本发明提供的无人机定点诱骗方法、装置、电子设备与存储介质中,诱骗设备能够以预设功率将所述诱骗定位信号发送至所述目标无人机,以利用所述诱骗定位信号压制所述目标无人机的真实定位信号,进而,目标无人机可将诱骗定位信号当做真实定位信号,从而基于此来控制自身的飞行,使其被诱骗飞行。

[0073] 进一步的,本发明中,可基于环境状态与诱骗目的地确定瞬时回报,从而依此形成用于训练识别模型的数据集,进而,基于数据集的诱骗经验,可有助于将识别模型训练为能够识别出长远回报的模型,且该长远回报能反应出决策是否能够趋向于将目标无人机诱骗至诱骗目的地,最终,通过找到最优的决策(即长远回报最高的诱骗策略),将目标无人机逐步诱骗到诱骗目的地或其附近区域。

附图说明

[0074] 为了更清楚地说明本发明实施例或现有技术中的技术方案,下面将对实施例或现有技术描述中所需要使用的附图作简单地介绍,显而易见地,下面描述中的附图仅仅是本发明的一些实施例,对于本领域普通技术人员来讲,在不付出创造性劳动性的前提下,还可以根据这些附图获得其他的附图。

[0075] 图1是本发明一实施例中无人机定点诱骗方法的流程示意图一;

[0076] 图2是本发明一实施例中应用场景的示意图;

[0077] 图3是本发明一实施例中诱骗飞行过程的示意图;

[0078] 图4是本发明一实施例中步骤S15的流程示意图;

[0079] 图5是本发明一实施例中无人机定点诱骗方法的流程示意图二;

[0080] 图6是本发明一实施例中无人机定点诱骗装置的程序模块示意图一;

[0081] 图7是本发明一实施例中无人机定点诱骗装置的程序模块示意图二；

[0082] 图8是本发明一实施例中电子设备的构造示意图。

具体实施方式

[0083] 下面将结合本发明实施例中的附图，对本发明实施例中的技术方案进行清楚、完整地描述，显然，所描述的实施例仅仅是本发明一部分实施例，而不是全部的实施例。基于本发明中的实施例，本领域普通技术人员在没有做出创造性劳动前提下所获得的所有其他实施例，都属于本发明保护的范围。

[0084] 本发明的说明书和权利要求书及上述附图中的术语“第一”、“第二”、“第三”“第四”等(如果存在)是用于区别类似的对象，而不必用于描述特定的顺序或先后次序。应该理解这样使用的数据在适当情况下可以互换，以便这里描述的本发明的实施例能够以除了在这里图示或描述的那些以外的顺序实施。此外，术语“包括”和“具有”以及他们的任何变形，意图在于覆盖不排他的包含，例如，包含了一系列步骤或单元的过程、方法、系统、产品或设备不必限于清楚地列出的那些步骤或单元，而是可包括没有清楚地列出的或对于这些过程、方法、产品或设备固有的其它步骤或单元。

[0085] 下面以具体地实施例对本发明的技术方案进行详细说明。下面这几个具体的实施例可以相互结合，对于相同或相似的概念或过程可能在某些实施例不再赘述。

[0086] 请参考图1，本发明实施例提供了一种无人机定点诱骗方法，同时，请参考图2，本发明实施例提供的无人机定点诱骗方法、装置，可应用于诱骗设备22。

[0087] 诱骗设备22可与监控设备23形成数据链路，从而接收到所需的数据，例如可自监控设备23获取到监控得到的无人机24的位置信息。诱骗设备22还可与无人机24交互，从而形成信号的压制路径，该压制路径可理解为能够向无人机24发出诱骗定位信号从而压制无人机的真实定位信号。无人机24还可与卫星21之间形成信号路径，通过该信号路径，无人机24可获取到真实定位信号(例如GPS信号)。本发明实施例可实现该真实定位信号的压制，进而，通过信号的压制实现诱骗。

[0088] 在部分方案中，监控设备23与诱骗设备22可以是复用的，例如通过同一套设备实现监控与诱骗，同一诱骗设备22可以与多个监控设备23交互从而实现相应无人机的压制，同一监控设备23也可与多个诱骗设备22交互。

[0089] 请参考图1，本发明实施例中，无人机定点诱骗方法，包括：

[0090] S11:确定目标无人机当前时刻的环境状态，

[0091] S12:当前所选择的诱骗策略选择方式是否为第一方式；

[0092] 若步骤S12的结果为是，则可实施步骤S13:将所述当前时刻的环境状态输入识别模型，以利用所述识别模型识别出采用每个诱骗策略的当前长远回报，并选择当前长远回报最高的诱骗策略作为所述当前时刻的诱骗策略；

[0093] 其中，不同诱骗策略对应不同的定位信号，所述当前时刻的诱骗策略所对应的定位信号为所述诱骗定位信号；

[0094] S14:以预设功率将所述诱骗定位信号发送至所述目标无人机，以利用所述诱骗定位信号压制所述目标无人机的真实定位信号。

[0095] 步骤S11所涉及的环境状态记载了对应时刻的位置以及该时刻之前多个时刻的位

置;该位置可以是监控设备23监控从而采集到的,监控设备23可实时监测所属区域(例如图3所示的特定区域)内无人机的空间位置,其中的空间位置可表征为: $X_t = (x_t, y_t, z_t)$,并将其传送到诱骗设备22。

[0096] 基于以上 X_t ,针对于环境状态 S ,考虑到无人机速度与以往较长时间有关,可将环境状态定义为 $S_t = (X_t, X_{t-1}, \dots, X_{t-T+1})$,其中 T 表示回溯的时间长度。

[0097] 以上所涉及的诱骗定位信号,可表征为 X'_t ,对应的诱骗策略可表征为: $a_t = X'_t(x'_t, y'_t, z'_t)$ 在步骤S14的具体举例中,可以大功率发射出去压制真实的卫星信号(即真实定位信号,具体可以为真实GPS信号)。

[0098] 以上方案中,诱骗设备能够以预设功率将所述诱骗定位信号发送至所述目标无人机,以利用所述诱骗定位信号压制所述目标无人机的真实定位信号,进而,目标无人机可将诱骗定位信号当做真实定位信号,从而基于此来控制自身的飞行,使其被诱骗飞行。

[0099] 在部分方案中,其中的诱骗可以是将无人机诱骗至部分方向,而在本发明实施例中,则是基于诱骗目的地对无人机实施诱骗的,其可理解为定点诱骗的处理过程。

[0100] 进而,在未到达诱骗目的地之前(或未到达诱骗目的地一定范围内之前),可重复实施以上步骤S11至步骤S14。可见,诱骗设备或监控设备还需实施以下过程:

[0101] 计算目标无人机的位置 X_t 与指定地点 $X_d = (x_L, y_L, z_L)$ 之间的距离 $L = \sqrt{(x_t - x_L)^2 + (y_t - y_L)^2 + (z_t - z_L)^2}$ 。当 L 小于1米时,即认为到达指定地点(即诱骗目的地)。

[0102] 为了实现定点的诱骗,本发明实施例对识别模型的训练素材、识别目标等进行了具体的配置,从而使其长远回报能反映出决策是否能够趋向于将目标无人机诱骗至诱骗目的地。

[0103] 故而,在本发明实施例中,请参考图1,步骤S11之后,还可包括:

[0104] S15:根据所述目标无人机在当前时刻的位置,以及预设的诱骗目的地,计算采用上一个时刻的诱骗策略所产生的当前瞬时回报;

[0105] S16:获取所述当前时刻的环境状态、所述上一个时刻的环境状态,以及所述当前瞬时回报作为一组诱骗经验,并利用所获取到的诱骗经验,更新用于训练所述识别模型的数据集。

[0106] 针对于其中的当前瞬时回报,因其是基于诱骗目的地为依据计算的,其能够表征出:针对于将目标无人机诱骗至诱骗目的地这一目标的回报,例如,瞬时回报越高,表示对应的策略越有助于达成这一目标。可见,只要在计算瞬时回报时考虑了诱骗目的地,就能在一定程度上体现出以上内容,在此基础上,所训练出的模型所识别的长远回报能够在一定程度上反映出决策是否能够趋向于将目标无人机诱骗至诱骗目的地。

[0107] 可见,以上方案中,可基于环境状态与诱骗目的地确定瞬时回报,从而依此形成用于训练识别模型的数据集,进而,基于数据集的诱骗经验,可有助于将识别模型训练为能够识别出长远回报的模型,且该长远回报能反应出决策是否能够趋向于将目标无人机诱骗至诱骗目的地,最终,通过找到最优的决策(即长远回报最高的诱骗策略),将目标无人机逐步诱骗到诱骗目的地或其附近区域。

[0108] 具体实施过程中,针对于环境状态由 S_t ,在相应的诱骗定位信号发出从而实现压制之后,当环境状态由 S_t 转为 $S_{t+1} = (X_{t+1}, X_t, \dots, X_{t-T+2})$,则可将每一时刻的诱骗经验 $e_t = (S_t, a_t, r_t, S_{t+1})$ 保存到数据集 D 中。

[0109] 其中一种实施方式中,请参考图4,计算瞬时回报的过程,即步骤S15,可以包括:

[0110] S151:根据所述诱骗目的地相对于所述目标无人机当前位置的方位,以及所述目标无人机的当前速度方向,确定当前方位收益;所述当前速度方向表征了在所述当前时刻,所述目标无人机的飞行速度的方向;

[0111] S152:根据所述目标无人机当前位置与所述诱骗目的地的当前距离,确定当前距离收益;

[0112] S153:根据所述当前方位收益与所述当前距离收益,确定所述当前瞬时回报。

[0113] 在步骤S151的具体举例中,无人机飞向指定地点(即诱骗目的地)过程中无人机当前飞行速度 \vec{V}_r 的方向(即当前速度方向)可表征为: θ_r ;无人机到指定地点的方向(即所述诱骗目的地相对于所述目标无人机当前位置的方位)可表征为: θ_w , θ_r 与 θ_w 之间的夹角为 θ 。

[0114] 即: $\theta = |\theta_r - \theta_w|$, $\theta \in (0,180)$,该夹角 θ 会对最终结果造成影响,因此瞬时回报 r_t 中包括当前方位收益 $A(\theta)$,当 θ 越小时收益越大。

[0115] 基于此,可定义当前方位收益 $A(\theta) = \frac{90-\theta}{90}$ 。

[0116] 在步骤S152的具体举例中,无人机到指定地点(即诱骗目的地)之间的距离 L 变化也会对最终结果有影响,因此瞬时回报 r_t 中也包括当前距离收益 $B(L)$,当距离 L 越小,对应的当前距离收益 $B(L)$ 越大。

[0117] 基于此,可定义当前距离收益 $B(L) = e^{1-L}$ 。

[0118] 基于以上的当前方位收益 $A(\theta)$ 与当前距离收益 $B(L)$,所述瞬时回报 r_t 是通过以下公式确定的:

[0119] $r_t = (1 - \alpha)A(\theta) + \alpha B(L)$;

[0120] $\alpha = \frac{1}{L}$ 。

[0121] 在以上所定义的公式中,可体现出:当距离 L 过大时更加关注当前方位收益 $A(\theta)$ 的变化,当距离 L 变小时,逐渐加大关注当前距离收益 $B(L)$ 的比例。可见,瞬时回报的计算可匹配于无人机与诱骗目的地的距离变化而适应性变化,其中,越接近诱骗目的地,距离本身越能体现出针对于诱骗目标的回报大小,而离诱骗目的地越远时,方向越能体现出针对于诱骗目标的回报大小,进而,以上的瞬时回报计算过程可准确体现出策略的回报(其中的回报指使得无人机的飞行越趋向于诱骗目的地的回报)。

[0122] 其中一种实施方式中,请参考图5,步骤S11之后,还可包括:

[0123] S17:根据选择概率信息,选择所述第一方式或第二方式作为当前的诱骗策略选择方式,其中,所述第一方式的选择概率与所述第二方式的选择概率之和为1;

[0124] S18:当前所选择的诱骗策略选择方式是否为第二方式;

[0125] 若步骤S18的结果为是,即若选择所述第二方式,则可实施步骤S19:随机自多个诱骗策略中选择一个作为当前时刻的诱骗策略。

[0126] 具体实施过程中,以上所涉及的识别模型可以采用卷积神经网络(Convolutional Neural Networks, CNN),以上所涉及的长远回报可利用Q函数的Q值来表征。由于状态空间巨大,本发明的具体方案可采用卷积神经网络(CNN)对Q函数进行估计,Q函数表示在环境状态S下,选择策略a可以获得长远的回报,其定义式为:

$$[0127] \quad Q^*(S, a) = E \left\{ r + \gamma \max_{a'} Q^*(S, a) \mid S, a \right\}$$

[0128] 其中:

[0129] γ 表示奖励折扣系数;

[0130] $Q^*(\cdot)$ 表示Q函数的Q值;

[0131] r表示瞬时回报;

[0132] a表示诱骗策略;

[0133] a' 表示下一个状态的诱骗策略;

[0134] S表示环境状态。

[0135] 将环境状态 S_t 输入卷积神经网络中,可得到该卷积神经网络输出在此环境状态下所有诱骗策略对应的Q值。

[0136] 在步骤S17中,可按照 $\epsilon - greedy$ 算法选取相应策略,其中以 ϵ 概率随机选择策略, $1 - \epsilon$ 概率选择所有策略中Q值最大的策略 $a = \operatorname{argmax}_a Q^*(S, a; w_i)$ 。

[0137] 其中的 w_i 可理解为卷积神经网络的权值。

[0138] 在采用卷积神经网络与Q函数的方案中,所述数据集中记录了所获取到的至少部分诱骗经验;所述识别模型为卷积神经网络,所识别出的长远回报是利用Q函数的Q值来表征的;

[0139] 所述识别模型是根据以下过程训练的:

[0140] 在所述数据集中所记录的各诱骗经验中,随机抽取诱骗经验;

[0141] 根据所抽取的诱骗经验,构建对应环境状态的目标Q值;

[0142] 根据所述目标Q值与对应环境状态的真实Q值,更新所述卷积神经网络。

[0143] 具体举例中,可以等概率地从数据集D中随机抽取诱骗经验,并根据诱骗经验中的状态以及相应的策略和回报,构建目标值:

$$[0144] \quad \eta_i = r_t + \gamma \max_{a'_{t+1}} Q^*(S_{t+1}, a'_{t+1});$$

[0145] 其中:

[0146] γ 表示奖励折扣系数;

[0147] $\max_{a'_{t+1}} Q^*(S_{t+1}, a'_{t+1})$ 表示在环境状态 S_{t+1} 下, 经验中选用已知策略可以取得最大的 Q 值。

[0148] 然后, 可计算目标值与真实值误差 $L_i(w_i) = E(\eta_i - Q^*(S_t, a'_t; w_i))^2$, 并据此计算梯度 $\nabla_{w_i} L_i(w_i) = E[L_i(w_i) \nabla_{w_i} Q^*(S_t, a'_t; w_i)]$ 更新卷积神经网络的权值 w_i 。

[0149] 在卷积神经网络更新后, 可更新 ε -greedy 算法中 $\varepsilon = \max(0.05, \varepsilon - \Delta\varepsilon)$, 其中 $\Delta\varepsilon$ 为更新步长的衰减系数。

[0150] 可见, 所述选择概率信息中, 所述第一方式的选择概率为固定概率值与动态概率值中较大的一个概率值, 所述动态概率值是随所述卷积神经网络中当前更新步长的衰减系数确定的。

[0151] 以上过程中所涉及的字母、描述、公式等均可参照于现有的以及改进的DQN算法理解。

[0152] 以上步骤S11可以是循环实施的, 进而, 步骤S11至步骤S14可以是循环实施, 直至目标无人机到达诱骗目的地。同时, 步骤S15与步骤S16可以在至少部分时间循环实施, 例如, 在数据集中诱骗经验的数量未达到上限时, 可以循环实施步骤S11、S15、S16, 将各诱骗经验写入数据集, 直至数据集的数量到达上限, 再例如, 在到达上限后, 可以不再实施步骤S15、S16, 也可以通过步骤S15、S16, 利用新的诱骗经验迭代已写入数据集的诱骗经验。同时, 利用数据集中的诱骗经验训练的过程可以是在数据集中的诱骗经验到达上限之后才开始实施的, 也可以是在数据集中写入诱骗经验后即开始实施, 利用数据集中的诱骗经验训练的过程可以是在完成所有诱骗经验的训练后停止, 也可在无人机到达诱骗目的地后才停止。

[0153] 以上方案中, 通过实时产生虚假的 GPS 信号 (即诱骗定位信号), 并在该空间内替代真实 GPS 信号 (即真实定位信号) 诱骗无人机改变其定位结果。同时, 运用深度强化学习可以在未知场景下进行学习并且充分发话收敛速度快的特性。

[0154] 一种具体举例中, 请结合图3, 目标无人机例如为一般商用无人机, 诱骗设备覆盖范围方圆500米, 监控设备实时监测所在区域内无人机空间位置, 指定地点坐标 $X_d = (1, 1, 1)$ (即诱骗目的地为 X_d)。

[0155] 在诱骗之前, 可初始化用户的数据集 $D = \emptyset$, 设置数据集上限 $N = 500$, 策略选择概率 $\varepsilon = 1$, 奖励衰减系数 $\gamma = 0.85$ 。

[0156] 当目标无人机飞入特定区域, 监控设备检测到无人机空间位置 X_1 , 将其传输给诱骗设备, 诱骗设备可通过步骤S15、S16等步骤开始进行学习, 并根据本发明实施例所涉及的方法做出最优决策 (即步骤S13中选择当前长远回报最高的诱骗策略作为当前时刻的诱骗策略), 并据此发射诱骗定位信号 X'_1 。目标无人机接收到诱骗信号 X'_1 后可调整自己的飞行速度 (例如是速度大小、方向, 还可以调整加速度等其他参数), 转向下一个空间位置 X_2 , 将当前经验 e_1 放入数据集 D 中, 循环迭代, 直至数据集 D 达到设定上限。

[0157] 在训练时,可等概率地从数据集 D 中抽取经验,并根据经验中的状态以及相应的策略和回报,构建目标值,并计算目标值和真实值之间的误差,并据此更新网络参数。随着不停地循环迭代,最终无人机落在指定地点 X_d (即诱骗目的地为 X_d)。

[0158] 通过以上手段的实验结果,采用本发明实施例的方案之后,可以通过学习无人机运动轨迹,向无人机发送诱骗定位信号,诱骗无人机飞行到指定地点。如图3所示,其示意了实验中无人机飞机轨迹路线。

[0159] 可见,通过在线学习无人机的运动轨迹,并在回报函数(计算瞬时回报的函数)中考虑无人机当前飞行速度方向与无人机到指定地点之间方向的夹角以及无人机到指定地点的距离对无人机运动轨迹的影响,根据网络输出决策,发射GPS诱骗信号(即诱骗定位信号),无人机接收到诱骗定位信号后会自动调整自己的飞行速度,最终可以飞行到指定地点,便于控制管理。

[0160] 综上所述,本发明实施例提供的无人机定点诱骗方法中,诱骗设备能够以预设功率将所述诱骗定位信号发送至所述目标无人机,以利用所述诱骗定位信号压制所述目标无人机的真实定位信号,进而,目标无人机可将诱骗定位信号当做真实定位信号,从而基于此来控制自身的飞行,使其被诱骗飞行。进一步的,本发明实施例中,可基于环境状态与诱骗目的地确定瞬时回报,从而依此形成用于训练识别模型的数据集,进而,基于数据集的诱骗经验,可有助于将识别模型训练为能够识别出长远回报的模型,且该长远回报能反应出决策是否能够趋向于将目标无人机诱骗至诱骗目的地,最终,通过找到最优的决策(即长远回报最高的诱骗策略),将目标无人机逐步诱骗到诱骗目的地或其附近区域。

[0161] 请参考图6,本发明实施例提供了一种无人机定点诱骗装置,应用于诱骗设备,包括:

[0162] 环境状态确定模块301,用于确定目标无人机当前时刻的环境状态,所述环境状态记载了对应时刻的位置以及该时刻之前多个时刻的位置;

[0163] 诱骗策略第一选择模块302,用于若当前所选择的诱骗策略选择方式为第一方式,则将所述当前时刻的环境状态输入识别模型,以利用所述识别模型识别出采用每个诱骗策略的当前长远回报,并选择当前长远回报最高的诱骗策略作为所述当前时刻的诱骗策略;其中,不同诱骗策略对应不同的定位信号,所述当前时刻的诱骗策略所对应的定位信号为所述诱骗定位信号;

[0164] 诱骗模块303,用于以预设功率将所述诱骗定位信号发送至所述目标无人机,以利用所述诱骗定位信号压制所述目标无人机的真实定位信号;

[0165] 瞬时回报计算模块304,用于根据所述目标无人机在当前时刻的位置,以及预设的诱骗目的地,计算采用上一个时刻的诱骗策略所产生的当前瞬时回报;

[0166] 数据集更新模块305,用于获取所述当前时刻的环境状态、所述上一个时刻的环境状态,以及所述当前瞬时回报作为一组诱骗经验,并利用所获取到的诱骗经验,更新用于训练所述识别模型的数据集。

[0167] 可选的,所述瞬时回报计算模块304,具体用于:

[0168] 根据所述诱骗目的地相对于所述目标无人机当前位置的方位,以及所述目标无人机的当前速度方向,确定当前方位收益;所述当前速度方向表征了在所述当前时刻,所述目标无人机的飞行速度的方向

[0169] 根据所述目标无人机当前位置与所述诱骗目的地的当前距离,确定当前距离收益;

[0170] 根据所述当前方位收益与所述当前距离收益,确定所述当前瞬时回报。

[0171] 可选的,所述当前瞬时回报是通过以下公式确定的:

$$[0172] \quad r_t = (1 - \alpha)A(\theta) + \alpha B(L);$$

$$[0173] \quad A(\theta) = \frac{90 - \theta}{90};$$

$$[0174] \quad \theta = |\theta_r - \theta_w|, \theta \in (0, 180);$$

$$[0175] \quad B(L) = e^{1-L};$$

$$[0176] \quad \alpha = \frac{1}{L};$$

[0177] 其中:

[0178] r_t 表征了当前瞬时回报;

[0179] θ 表征了所述诱骗目的地相对于所述目标无人机当前位置的方向 θ_r 与所述目标无人机的当前速度方向 θ_w 之间的夹角;

[0180] $A(\theta)$ 表征了当前方位收益;

[0181] L 表征了所述目标无人机当前位置与所述诱骗目的地的当前距离;

[0182] $B(L)$ 表征了当前距离收益。

[0183] 可选的,请参考图7,所述的无人机定点诱骗装置300,还包括:

[0184] 方式选择模块306,用于根据选择概率信息,选择所述第一方式或第二方式作为当前的诱骗策略选择方式,其中,所述第一方式的选择概率与所述第二方式的选择概率之和为1;

[0185] 诱骗策略第二选择模块307,用于若选择所述第二方式,则随机自多个诱骗策略中选择一个作为当前时刻的诱骗策略。

[0186] 可选的,所述识别模型采用卷积神经网络,所述选择概率信息中,所述第一方式的选择概率为固定概率值与动态概率值中较大的一个概率值,所述动态概率值是随所述卷积神经网络中当前更新步长的衰减系数确定的。

[0187] 可选的,所述数据集中记录了所获取到的至少部分诱骗经验;所述识别模型为卷积神经网络,所识别出的长远回报是利用Q函数的Q值来表征的;

[0188] 所述识别模型是根据以下过程训练的:

[0189] 在所述数据集中所记录的各诱骗经验中,随机抽取诱骗经验;

[0190] 根据所抽取的诱骗经验,构建对应环境状态的目标Q值;

[0191] 根据所述目标Q值与对应环境状态的真实Q值,更新所述卷积神经网络。

[0192] 综上所述,本发明实施例提供的无人机定点诱骗装置中,诱骗设备能够以预设功率将所述诱骗定位信号发送至所述目标无人机,以利用所述诱骗定位信号压制所述目标无人机的真实定位信号,进而,目标无人机可将诱骗定位信号当做真实定位信号,从而基于此来控制自身的飞行,使其被诱骗飞行。进一步的,本发明实施例中,可基于环境状态与诱骗

目的地确定瞬时回报,从而依此形成用于训练识别模型的数据集,进而,基于数据集的诱骗经验,可有助于将识别模型训练为能够识别出长远回报的模型,且该长远回报能反应出决策是否能够趋向于将目标无人机诱骗至诱骗目的地,最终,通过找到最优的决策(即长远回报最高的诱骗策略),将目标无人机逐步诱骗到诱骗目的地或其附近区域。

[0193] 请参考图8,提供了一种电子设备40,包括:

[0194] 处理器41;以及,

[0195] 存储器42,用于存储所述处理器的可执行指令;

[0196] 其中,所述处理器41配置为经由执行所述可执行指令来执行以上所涉及的方法。

[0197] 处理器41能够通过总线43与存储器42通讯。

[0198] 本发明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,该程序被处理器执行时实现以上所涉及的方法。

[0199] 本领域普通技术人员可以理解:实现上述各方法实施例的全部或部分步骤可以通过程序指令相关的硬件来完成。前述的程序可以存储于一计算机可读取存储介质中。该程序在执行时,执行包括上述各方法实施例的步骤;而前述的存储介质包括:ROM、RAM、磁碟或者光盘等各种可以存储程序代码的介质。

[0200] 最后应说明的是:以上各实施例仅用以说明本发明的技术方案,而非对其限制;尽管参照前述各实施例对本发明进行了详细的说明,本领域的普通技术人员应当理解:其依然可以对前述各实施例所记载的技术方案进行修改,或者对其中部分或者全部技术特征进行等同替换;而这些修改或者替换,并不使相应技术方案的本质脱离本发明各实施例技术方案的范围。

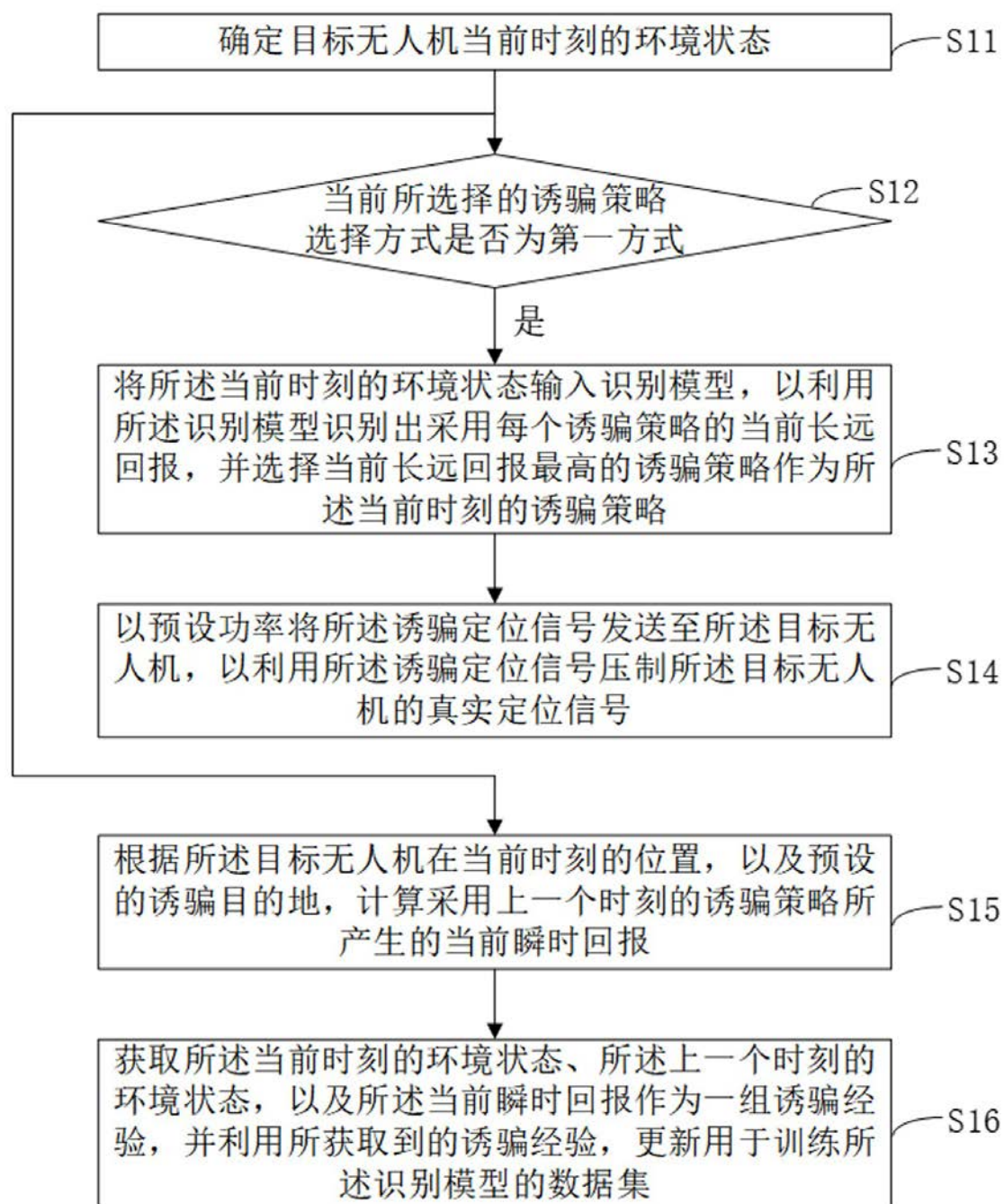


图1

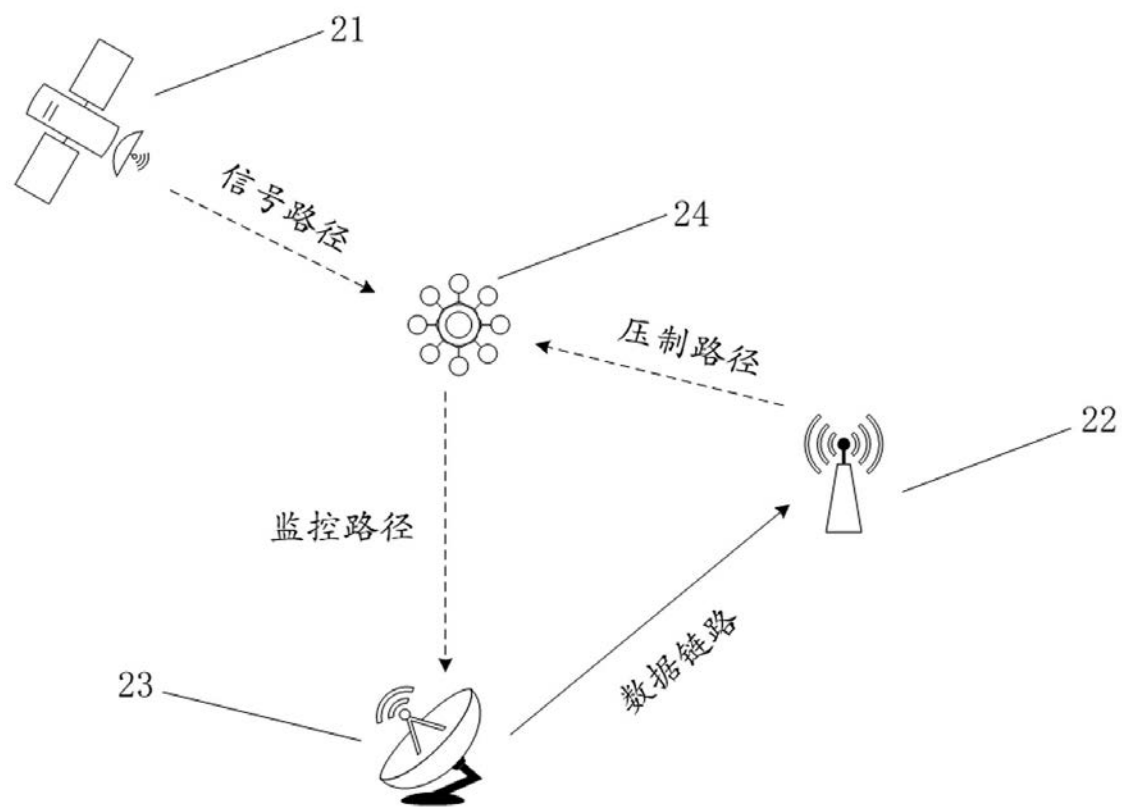


图2

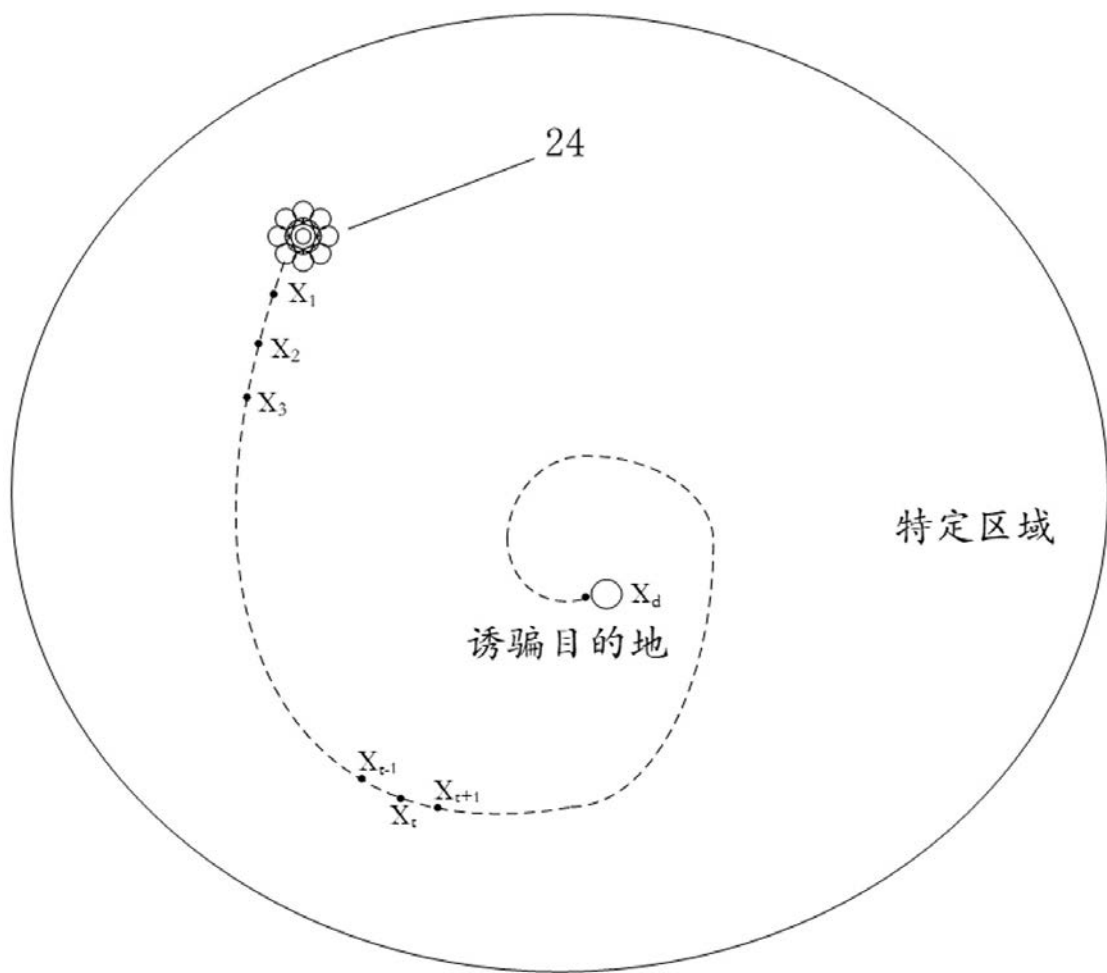


图3

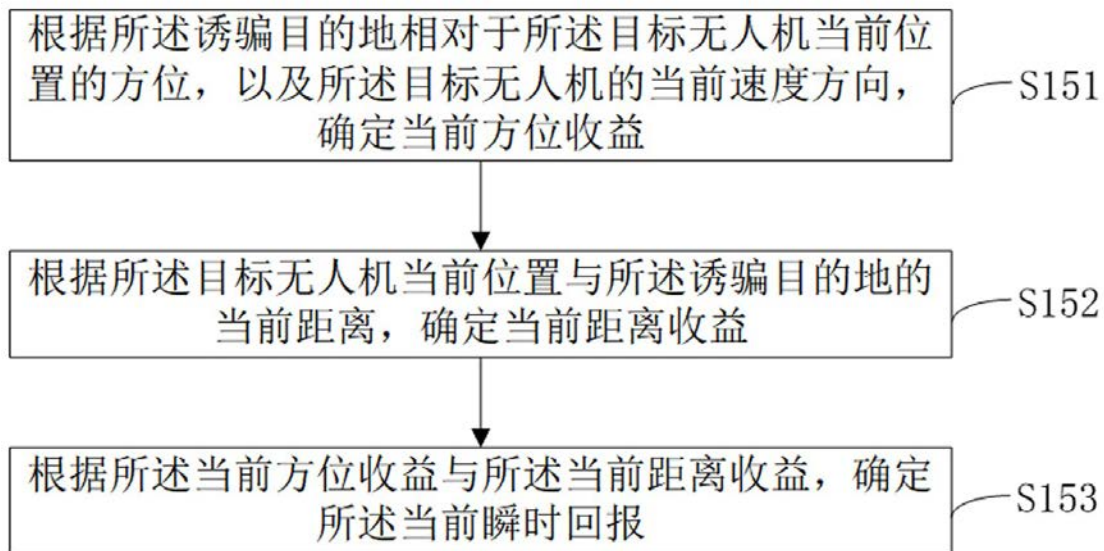


图4

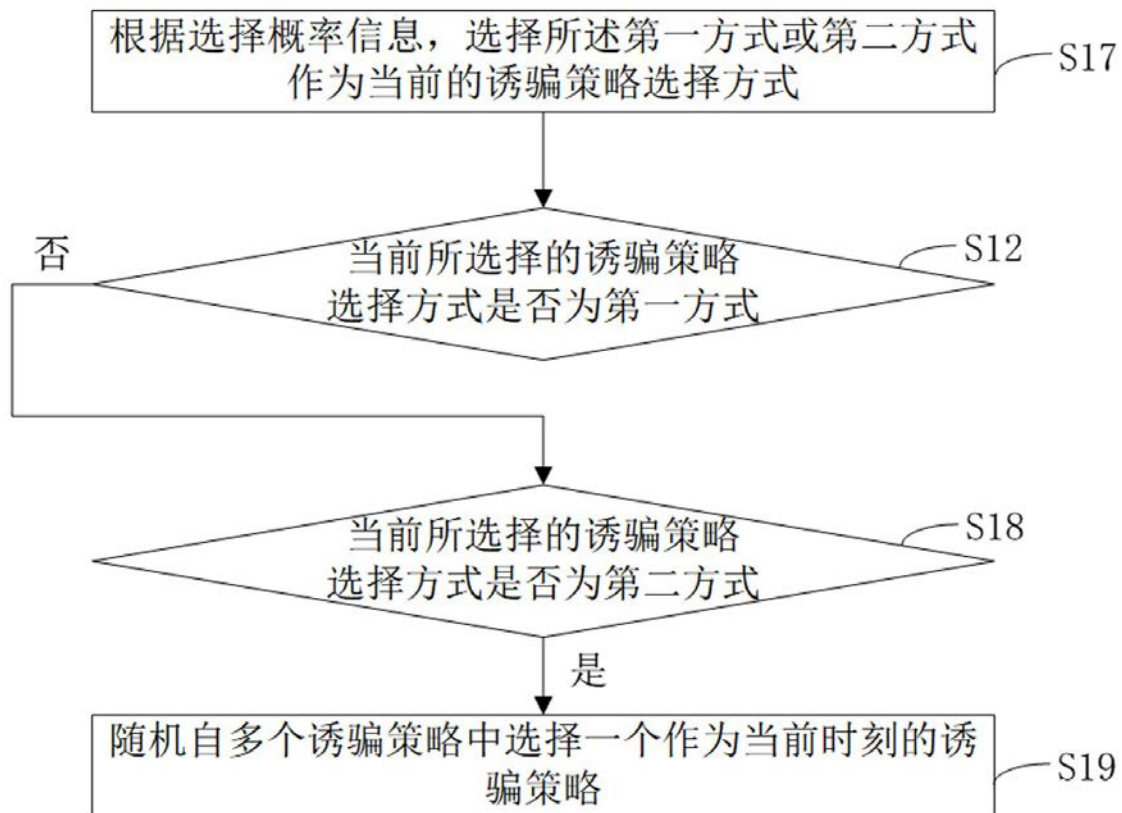


图5

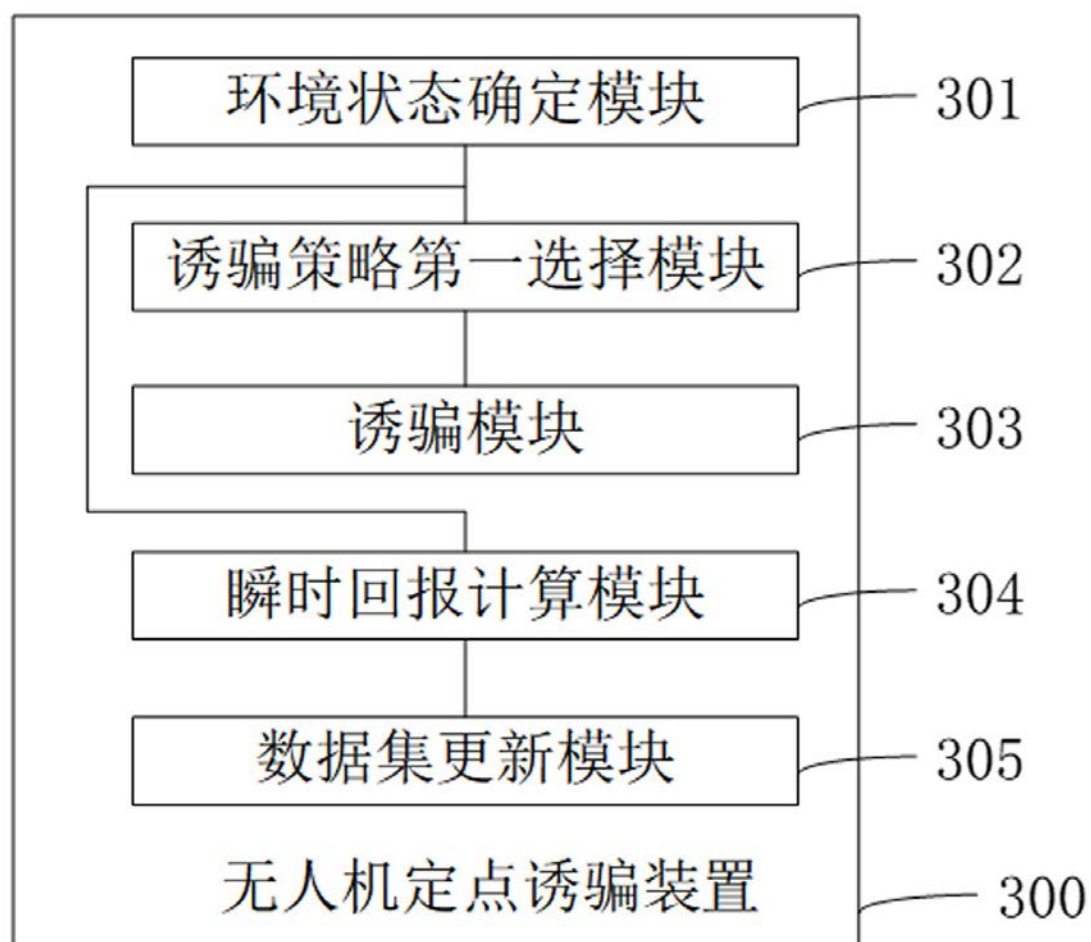


图6

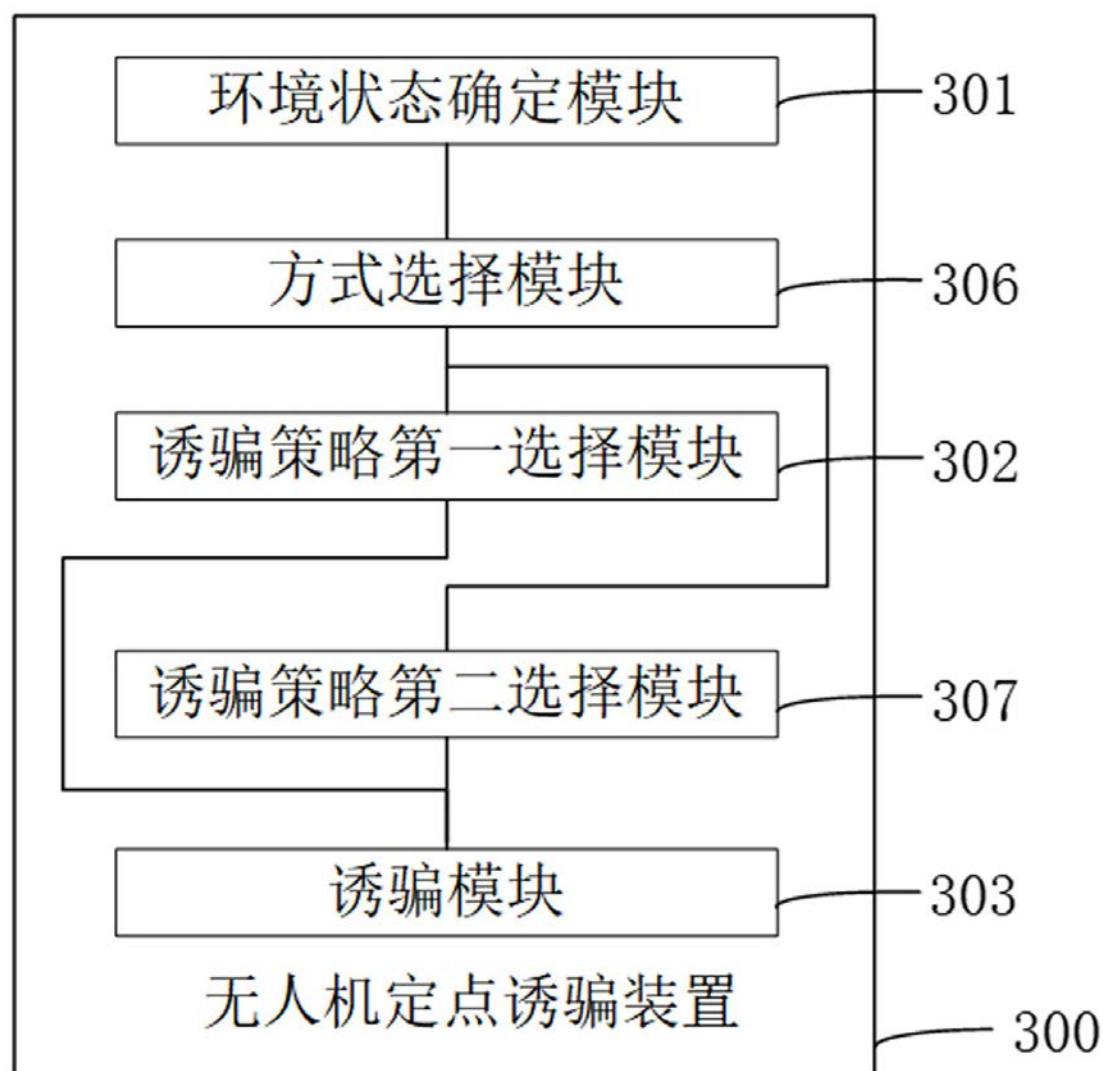


图7

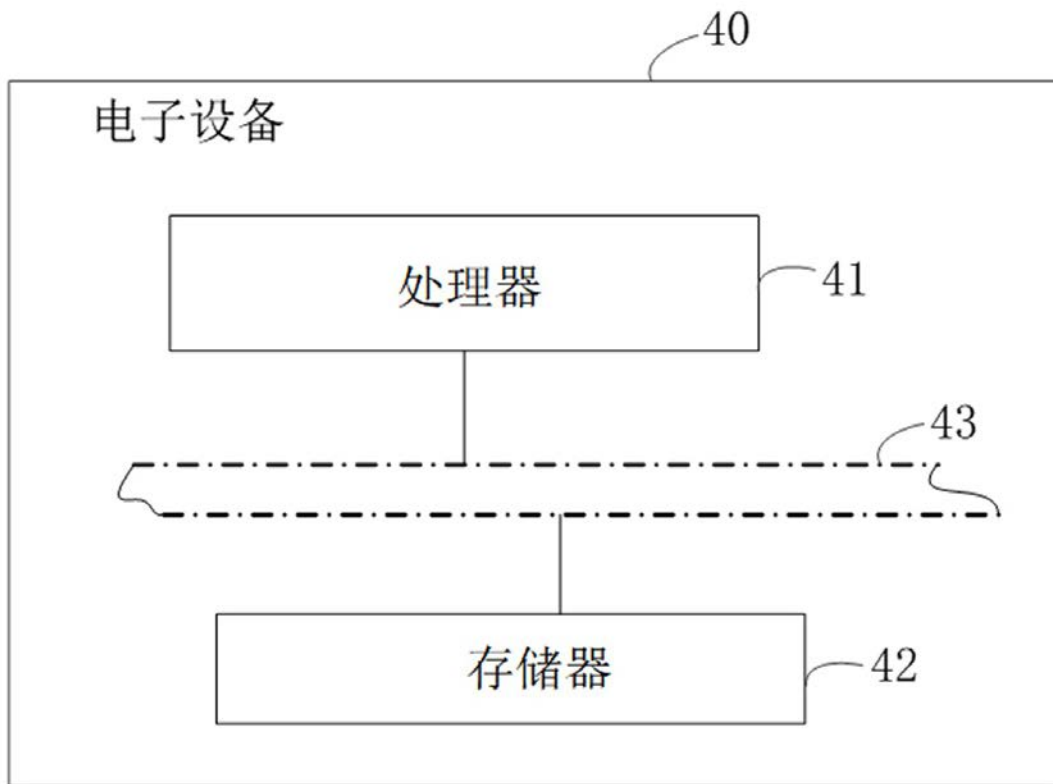


图8