智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

# 旋翼无人机在移动平台降落的控制参数自学习调节方法

张鹏鹏, 魏长赟, 张恺睿, 欧阳勇平

引用本文:
张鹏鹏,魏长,张恺睿,欧阳勇平. 旋翼无人机在移动平台降落的控制参数自学习调节方法[J]. 智能系统学报, 2022, 17(5): 931–940.
ZHANG Pengpeng,WEI Changyun,ZHANG Kairui,OUYANG Yongping. Self–learning approach to control parameter adjustment for quadcopter landing on a moving platform[J]. *CAAI Transactions on Intelligent Systems*, 2022, 17(5): 931–940.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202107040

DOI: 10.11992/tis.202107040
网络出版地址: https://kns.cnki.net/kcms/detail/23.1538.TP.20220519.1429.004.html

# 旋翼无人机在移动平台降落的控制参数自学习调节方法

张鹏鹏, 魏长赟, 张恺睿, 欧阳勇平
(河海大学 机电工程学院, 江苏 常州 213022)

**摘 要:** 无人机设备能够适应复杂地形, 但由于电池容量等原因, 无人机无法长时间执行任务。无人机与其他无人系统(无人车、无人船等)协同能够有效提升无人机的工作时间, 完成既定任务, 当无人机完成任务后, 将无人机迅速稳定地降落至移动平台上是一项必要且具有挑战性的工作。针对降落问题, 文中提出了基于矫正纠偏 COACH(corrective advice communicated humans) 方法的深度强化学习比例积分微分 (proportional-integral-derivative, PID) 方法, 为无人机降落至移动平台提供了最优路径。首先在仿真环境中使用矫正纠偏框架对强化学习模型进行训练, 然后在仿真环境和真实环境中, 使用训练后的模型输出控制参数, 最后利用输出参数获得无人机位置控制量。仿真结果和真实无人机实验表明, 基于矫正纠偏 COACH 方法的深度强化学习 PID 方法优于传统控制方法, 且能稳定完成在移动平台上的降落任务。

**关键词:** 自主降落; 强化学习; 路径规划; COACH 框架; 确定性策略梯度; 空地协同; 无人机; 最优控制
**中图分类号:** TP273+.2 **文献标志码:** A **文章编号:** 1673−4785(2022)05−0931−10

**中文引用格式: 张鹏鹏, 魏长赟, 张恺睿, 等. 旋翼无人机在移动平台降落的控制参数自学习调节方法 [J]. 智能系统学报, 2022, 17(5): 931–940.**
**英文引用格式: ZHANG Pengpeng, WEI Changyun, ZHANG Kairui, et al. Self-learning approach to control parameter adjustment for quadcopter landing on a moving platform[J]. CAAI transactions on intelligent systems, 2022, 17(5): 931–940.**

# Self-learning approach to control parameter adjustment for quadcopter landing on a moving platform

ZHANG Pengpeng, WEI Changyun, ZHANG Kairui, OUYANG Yongping
(College of Mechanical and Electrical Engineering, Hohai University, Changzhou 213022, China)

**Abstract**: Unmanned Aerial Vehicle (UAV) is a type of robot that performs well in mapping without being affected by the terrain. However, a UAV cannot perform its tasks for long due to its small battery capacity and several other reasons. The collaboration between UAVs and other unmanned ground vehicles (UGVs) is considered a crucial solution to this concern as it can save up the time taken by UAVs effectively when completing a scheduled task. When deploying a team of UAVs and UGVs, it is both important and challenging to land a UAV on a mobile platform quickly and stably. To circumvent the UAV landing issue, this study proposes a reinforcement learning PID method based on the correction COACH method, thereby providing an optimal path for the UAV to land on a mobile platform. First, the reinforcement learning agent is trained using the rectification framework in a simulated environment. Next, the trained agent is used for output control parameters in the simulated and true environments, and subsequently, the output parameters are utilized to obtain the control variables of the UAV's position. The simulation and real UAV experiment results show that the deep reinforcement learning PID method based on the correction COACH method is superior to the traditional control method and can accomplish the task of a stable landing on a mobile platform.

**Keywords**: autonomous landing; reinforcement learning; path planning; COACH frame; deterministic policy gradient; air-ground cooperation; UAV; optimal control

无人机可以应用于不同的场景, 例如日常的便民生活应用[1], 农业生产过程[2], 矿场的侦探和挖掘过程[3] 等。单一的无人机不受地形的限制, 但是由于携带电池能量的限制, 执行任务时间短, 并且难以承担较重的负载。无人车(船)移动范围受限于地形, 难以到达特定的位置。无人机–无人车(船)的组合系统可以结合两者的优点, 完成复杂

**收稿日期:** 2021−07−20. **网络出版日期:** 2022−05−20.
**基金项目:** 国家自然科学基金项目(61703138); 中央高校基本科研业务费项目(B200202224).
**通信作者:** 魏长赟. E-mail: c.wei@hhu.edu.cn.

的任务[4-5]。在执行任务结束后，无人机如何移动到指定位置是协同系统实际应用的关键问题[6]，因此本文聚焦于无人机的自主降落问题。

在文献 [7-8] 中，作者将多种传统的控制方法应用于无人机降落任务，这些方法具有稳定和低算力需求的优点，但是较难实现最优的控制效果。文献 [9-12] 将强化学习理论应用于无人机降落问题，并取得良好的效果。在文献 [13-16] 中，使用强化学习原理调整控制算法的参数，面对不同的控制情形，能够实现较优的控制效果，但是并没有进行真实无人机实验。

针对以上方法的不足，本文结合深度强化学习理论和比例积分微分 (proportional-integral-derivative, PID) 控制方法，解决无人机降落至移动平台的问题，本方法既有 PID 方法的稳定性，又能够发挥强化学习寻找最优控制策略的优点，迅速完成无人机降落到无人移动平台的任务。

## 1　无人机降落问题描述

在多机器人无人系统中，无人机在完成特定任务后需要降落至特定平台，本文基于上述任务，针对无人机降落问题，提出一种结合深度强化学习算法和比例积分微分 (proportional-integral-derivative, PID) 原理的控制方法。文中首先介绍多机器人协同系统以及实现无人机降落任务的必要性，并详细介绍传统控制方法和机器学习方法在无人机降落问题上的应用现状。

### 1.1　关键的无人机降落问题

在如今的机器人学研究中，单一的机器人难以完成复杂任务。在所有的无人设备中，无人机有着多项优点，其他设备难以替代无人机执行任务。首先无人机运动不受地形限制，可以轻易地到达特定的位置，并且无人机在空中悬停可以为地面无人设备和工作人员提供高处视角的图像信息，为发现和定位目标物品提供可能。同时，无人机由于自身结构的原因也有特定的缺点，包括由于电池能量不足造成的执行任务时间短和无法携带较重负载的问题等。对单一的无人机设备添加无人车（船）组成协同系统可以有效地解决无人机的上述缺点。实际应用中，无人机完成任务，须自动返回，以备下次任务的执行。因此，在这些协同系统中，如何将无人机降落到特定的平台上是一项必须解决的任务。在文献 [17] 中，作者提出一种无人机和无人车的协同系统，该系统作业于建筑行业，收集建筑区域内的各种关键数据。该方法结合了两种机器人的优点，弥补单一机器人的不足，高效地实现建筑行业数据收集任务，不过在文中提到的无人机降落方法依然有进步的空间，难以在复杂的环境中实行降落任务。

### 1.2　传统控制方法应用

文献 [18] 提出一种比例微分 (proportional-integral, PD) 控制器，该控制方法针对无人机自主降落问题，实现了无人机降落到固定平台的任务。在文献 [8] 中，作者将模型预测控制方法应用于无人机降落问题，该方法结构轻量，响应迅速，能够在低算力的平台中运行。同时，作者在文中使用仿真环境进行验证，且效果良好，但是并没有在真实的场景中进行降落效果的测试。PID 控制方法在控制任务中广泛使用，但是固定参数的 PID 控制方法对非线性问题适应性差，在文献 [7] 中，作者提出一种基于模糊逻辑的 PID 控制方法，结果显示，该方法优于传统的控制方法，不足之处在于该方法未考虑无人机降落到移动的平台的情形。

### 1.3　强化学习算法的应用

许多学者应用强化学习算法寻找解决问题的最优策略。强化学习算法能够实现在干扰和复杂情况下的最优控制，这是传统控制方法难以比拟的。理论上，基于马尔可夫过程的强化学习算法有潜力找到最优的控制策略，当算法训练充分后就可以实现对于无人机降落问题的最优控制。在文献 [9,11] 中，作者将确定性策略梯度方法[19] 应用于无人机降落问题中，该方法在虚拟环境中进行训练，并且可以在仿真和真实环境中，实现对无人机降落过程的控制。确定性策略梯度方法能够根据不同的状态输入输出不同的动作，进而完成当前的任务。文献 [11] 中，输入状态包含 $x$、$y$ 两个方向上的位置信息，算法根据不同的位置信息控制无人机降落，以连续的状态作为输入并输出连续动作，有潜力实现精确控制，由于文章中的方法未使用 $z$ 方向的位置信息，当面对不同高度的输入时可能有相同的输出，影响控制无人机降落的效果。文献 [9] 中，作者同样使用确定性策略梯度方法，该方法的输入包括三轴的位置信息，可根据无人机高度改变输出动作，从而实现精准控制，同上面的方法一样，此方法以连续的状态作为输入并输出连续的动作，保证无人机的精确控制。

在文献 [16,20] 中，一种结合 PID 理论和强化学习原理的方法被应用于移动机器人的路径规划问题，在仿真实验中，对比传统 PID 方法，文中提出的 Q 学习–PID 方法在路径规划实验的结果中优势明显，面对不同环境和干扰时，表现出鲁棒性强的优点。文献 [21-22] 将参数自学习调节方法应用于无人机降落至静止平台的任务，由于控制

器参数随当前状态自适应调节, 因此取得的无人机路径控制效果均优于传统 PID 方法。目前类似的方法还没有应用于无人机降落至移动平台任务, 在无人机降落问题中, 使用结合 PID 理论和强化学习原理的方法, 具有创新性和可行性。在文献[23] 中, 作者提出一种矫正纠偏(corrective advice communicated humans, COACH)框架, 使用人类建议, 用于帮助强化学习算法寻找最优的控制策略, 效果显著, 能优化获得的最终策略。

本文提出一种结合 PID 原理和强化学习理论的方法, 完成无人机降落到移动平台的任务, 应用矫正纠偏框架, 优化最终训练得到的策略。上层的控制策略选择确定性策略梯度方法, 该方法有着连续的输入和输出, 在连续空间上, 有潜力实现优秀的控制效果。下层应用 PID 方法, 用于保证无人机降落的稳定性。

## 2 强化学习算法描述

### 2.1 强化学习

人工智能领域中, 强化学习通常根据特定的状态寻找最优动作, 并将动作执行进而完成相应的任务。基于强化学习原理的方法已经在多个领域取得亮眼表现, 包括围棋[24]、电脑游戏[25-26] 等。在理论上, 基于强化学习的方法在经过一定回合的训练后, 所获得的智能体可以在不同场景实现特定的任务。一个标准的强化学习问题可以由 $\boldsymbol{S}$、$\boldsymbol{A}$、$\boldsymbol{P}$、$r$、$\gamma$ 定义。其中 $\boldsymbol{S}$ 和 $\boldsymbol{A}$ 分别代表输入状态和输出动作的集合, $\boldsymbol{s}$ 和 $\boldsymbol{a}$ 表示某一时间的状态和动作, $\boldsymbol{P}$ 表示状态转移概率, $r$ 是奖励信息, $\gamma$ 是折扣因子, 同时定义总体奖励 $R_t = \sum_{i=t}^{f} \gamma^{i-t} r(s_i, a_i)$, 其中 $f$ 是最后的回合数。

智能体在环境中进行训练, 不断优化当前的策略, 对于一个特定的策略 $\pi$, 本文以公式 $V^{\pi}(s_t) = \mathrm{E}[R_t | s = s_t, \pi]$ 定义价值函数 $V^{\pi}$。同样由公式 $Q^{\pi}(s_t, a_t) = \mathrm{E}[R_t | s = s_t, a = a_t, \pi]$ 定义动作价值函数。同时本文使用 $J(\pi)$ 定义策略 $\pi$ 的评价标准, 具体为 $J(\pi) = \mathrm{E}[R_t | \pi]$。最后使用 $\pi^*$ 来代表最优的控制策略, 即,

$$\pi^* = \arg\max Q^*(s_t, a_t)$$

智能体不断地在环境中训练, 并使用贝尔曼方程:

$$Q(s_t, a_t) = r(s_t, a_t) + \gamma \sum_{s_{t+1} \in S} P^{a_t}_{s_t, s_{t+1}} \sum_{a_{t+1} \in A} Q(s_{t+1}, a_{t+1})$$

不断更新状态价值函数, 因此智能体在强化学习算法的规则下不断训练并获得最优的控制策略。

在强化学习的发展过程中, 最具有代表性的算法[27] 是 Q 学习算法[28], 算法本身结构简单, 并为其后的算法带来启发, 这些算法包括深度 Q 学习[29]、双 Q 学习[30]、决斗 Q 学习算法[31], 但是 Q 学习算法由于其离散的输入和输出, 只能解决复杂度低的低维度问题。

### 2.2 深度强化学习

在实际任务执行时, 状态和动作的表示是连续的, 由于维度爆炸的问题, 在连续的动作状态空间中使用离散的状态和动作难以实现。因此, 本文使用神经网络非线性拟合的特点, 对动作价值函数进行估计, 为了更好的表示动作价值函数, 本方法定义损失函数 $L(\theta^Q) = \mathrm{E}[(y_t - Q(s_t, a_t|\theta^Q))^2]$ 来优化网络参数 $\theta^Q$, 其中 $y_t = r(s_t, a_t) + \gamma Q(s_{t+1}, a_{t+1}|\theta^Q)$。如果策略是确定的, 则可以将状态映射到动作上, 即 $\mu: S \rightarrow A$, 之后本文定义动作网络 $\theta^{\mu}$, 其遵循 $J(\pi)$ 进行更新, 使得 $J(\pi)$ 变大, 即根据 $\nabla_{\theta^{\mu}} J(\pi) \approx \mathrm{E}[\nabla_{\theta^{\mu}} Q(s, a|\theta^Q)|_{s=s_t, a=\mu(s_t|\theta^{\mu})}] = \mathrm{E}[\nabla_a Q(s, a|\theta^Q)|_{s=s_t, a=\mu(s_t)} \nabla_{\theta^{\mu}} \mu(s|\theta^{\mu})|_{s=s_t}]$ 进行更新。

确定性策略梯度方法[19] 是一种解决在连续状态空间和动作空间的无模型算法。此方法使用动作–评论家的结构, 有两个主要的人工神经网络, 一个用于拟合动作价值函数, 称为价值网络 $\theta^Q$, 另一个网络用于产生动作, 成为动作网络 $\theta^{\mu}$。在这两个网络进行更新时, 网络的迭代会不稳定和发散, 因此, 使用两个目标网络(目标价值网络 $\theta^{Q'}$ 和目标动作网络 $\theta^{\mu'}$)对更新的两个主网络进行软更新, 提高其稳定性。两个目标网络分别与价值网络和动作网络有着相同的结构。在训练时, 每个输出动作有随机的干扰, 用于增加算法探索的空间, 动作作用于环境后, 智能体会将观察数组 $(s_{\mathrm{step}}, a_{\mathrm{step}}, r_{\mathrm{step}}, s_{\mathrm{step}+1})$ 储存到记忆库中, 记忆库达到一定数量后, 按公式 $L = \left(\frac{1}{B}\right) \cdot \sum_i (y_i - Q(s_i, a_i|\theta^Q))^2$ 更新价值网络参数 $\theta^Q$ 使得 $L$ 减小, 其中 $B$ 是样本取样个数, $y_t = r(s_t, a_t) + \gamma Q(s_{t+1}, \mu(s_{t+1}|\theta^{\mu'})|\theta^{Q'})$, 之后按照 $\nabla_{\theta_{\mu}} J \approx \frac{1}{B} \sum_i \nabla_a Q(s, a|\theta^Q)|_{s=s_i, a=\mu(s_i)} \nabla_{\theta^{\mu}} \mu(s|\theta^{\mu})|_{s_i}$ 对动作网络 $\theta^{\mu}$ 进行更新, 使得 $J$ 增大, 接着, 本方法对两个目标网络进行软更新, 具体公式为: $\theta^{Q'} = \tau\theta^Q + (1-\tau)\theta^{Q'}$ 和 $\theta^{\mu'} = \tau\theta^{\mu} + (1-\tau)\theta^{\mu'}$。最后经过不断地训练, 会得到两个主网络, 用于实际问题的解决。

### 2.3 使用矫正纠偏框架的强化学习方法

使用深度强化学习算法需要大量的时间进行训练, 并且训练的时间随着动作空间维度的增加而显著增加。为了减少训练的时间并提升训练的效果, 一种矫正纠偏框架用于提升训练的效率, 在智能体输出动作时, 使用人类的建议(一个二值化的量)对智能体产生的动作进行增强或者减

弱。在矫正纠偏框架下的确定性策略梯度方法具体结构如图 1 所示，使用人类建议指导智能体探索，使用确定性策略梯度方法对网络进行更新，最终迅速获得最优策略。

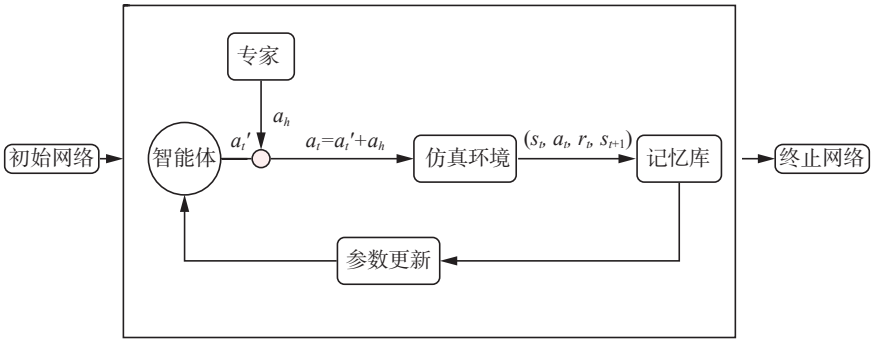



**图 1　使用矫正纠偏框架的确定性策略梯度方法**

**Fig. 1　DDPG with COACH**

在本文提出的方法中，使用人类的建议增加或者减弱智能体生成的动作，最终动作作用于环境中，并储存在$(s_t,a_t,r_t,s_{t+1})$中，且与确定性策略梯度方法一致，对网络权重进行更新。总体来看，将人类的建议用于修正智能体产生的动作，会使得智能体在相同的训练回合下，获得更佳的策略。

## 3　无人机降落控制策略

### 3.1　传统 PID 方法控制无人机降落

传统的 PID 方法结构如图 2 所示，误差信号$e(t)$是设定值和测量值的差值，有比例、积分、微分 3 个环节，分别由$k_{\mathrm{P}}$、$k_{\mathrm{I}}$、$k_{\mathrm{D}}$ 3 个参数对输入的误差信号按公式$u(t)=k_{\mathrm{P}}e(t)+k_{\mathrm{I}}\int_0^t e(\tau)\mathrm{d}\tau+k_{\mathrm{D}}\dfrac{\mathrm{d}e(t)}{\mathrm{d}t}$处理，最终获得所需的输出控制量。

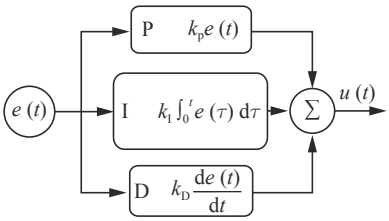



**图 2　标准 PID 控制器**

**Fig. 2　Standard structure of a PID controller**

### 3.2　深度强化学习算法控制无人机降落

本文提出的方法将强化学习算法应用在 PID 控制算法的上层，方法的结构如图 3 所示，有两个控制模块，左边框为强化学习模块，右边框为 PID 控制模块，强化学习的输入状态由 3 个方向上的位置组成，输出 $\boldsymbol{a}$ 为 PID 控制模块的参数$k_{\mathrm{P}}$、$k_{\mathrm{I}}$、$k_{\mathrm{D}}$。

强化学习模块时刻调节 PID 控制器的参数，具体的奖励函数由公式

$$r_t=\begin{cases}1, & \text{成功}\\ -1, & \text{失败}\\ d_{t-1}-d_t, & \text{其他}\end{cases}$$

定义，其中$d_t$是 $t$ 时刻无人机与目标点的欧式距离。当无人机降落到指定的地点，奖励值为 1，当降落失败（目标消失或未降落至目标点）时，奖励值为−1，其他情况下，奖励值为与上一时刻欧式距离和当前时刻欧式距离的差值。一旦记忆库存满，评价网络和动作网络便开始更新。

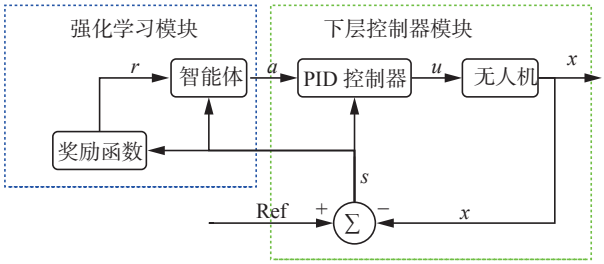



**图 3　强化学习与 PID 结合方法**

**Fig. 3　RL-based PID**

由于强化学习模块能时刻输出动作 $\boldsymbol{a}$ 对 PID 控制模块的参数进行调整，因此本方法可在多个场景控制无人机进行降落。PID 模块输出控制命令 $\boldsymbol{u}$，包含 $x$ 和 $y$ 方向上的控制位置，无人机在 $z$ 方向上的目标降落速度为 0.3m/s 保持不变，Ref 代表无人机的目标位置，$x$ 是当前机器人的位置，$\boldsymbol{s}$ 是无人机在图像中相对目标点的位置，包含 $x$、$y$ 和 $z$ 三轴的信息。由于强化学习算法的加入，本方法能够在复杂环境中更加有效地控制无人机降落。

### 3.3　应用矫正纠偏框架的深度强化学习算法控制无人机降落

本文的方法使用矫正纠偏框架优化训练过程，强化学习算法和矫正纠偏框架的结构如图 4 所示，矫正纠偏框架使用人的建议代替干扰信号，用于智能体探索环节，因为人类建议的加入，所以增强了最终获得策略的鲁棒性。

当智能体选择动作$a'_t$，然后根据人类当前指

导获得最终的输出$a_t$, 其中$a'_t$的取值范围为[0,1.0]叠加人类的经验$a_h$, 其取值为 0.2 或者−0.2, 最终输出的结果$a_t$区间为[0,1.0], 当$a_t$超出 1.0 时, 认为输出结果为 1.0, 当$a_t$小于 0 时, 认为输出结果为 0。当误差范围较大时, 人类经验认为可以增大比例参数, 此时$a_h = 0.2$, 进而加速无人机到达目标点, 当误差范围较小时, 人类经验认为需减小比例参数, 即取$a_h = -0.2$, 进而实现精准的降落, 对于积分和微分参数并没有使用人类的经验进行调节。

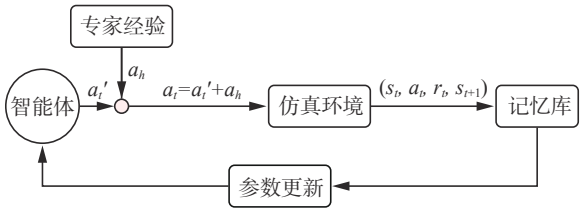



**图 4　使用矫正纠偏框架的强化学习方法**

**Fig. 4　RL with COACH**

控制参数自学习调节方法具体流程如图 5 所示, 由传感器获得无人机相对移动平台的坐标, 深度强化学习模块对状态进行处理, 输出底层控制器的 $x$、$y$ 方向上的控制参数, 之后底层控制模块根据当前误差和控制参数计算获得无人机位置控制指令并执行。无人机在降落过程中不断检测当前状态, 若无人机位置合适, 则旋翼停止运动, 无人机降落至目标区域, 否则无人机继续执行位置控制的步骤, 直至无人机降落至目标区域。

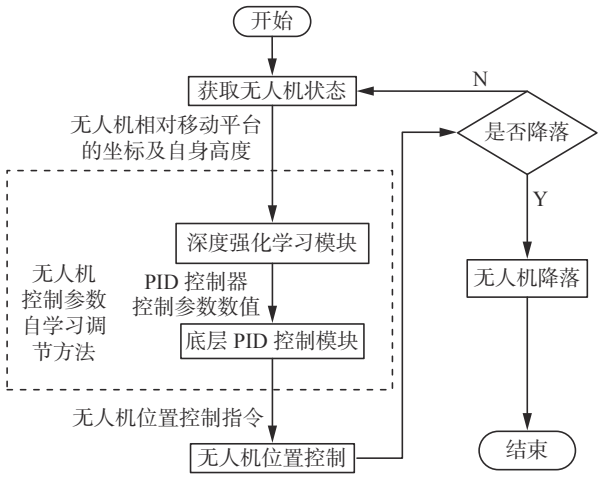
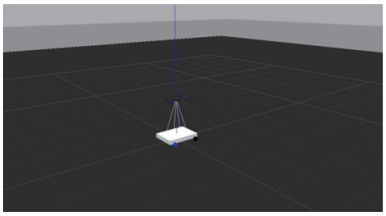



**图 5　无人机自主降落流程**

**Fig. 5　Autonomous landing process for an UAV**

## 4　实验及结果分析

本文提出的方法, 在 Gazebo 仿真环境中训练并在仿真和真实环境中进行测试。此外, PID 模块和强化学习模块之间的通信使用机器人操作系统 (robot operation system)[32], 如图 6 所示, 图 6(a) 给出的是仿真环境, 图 6(b) 给出的是真实降落的场景。



(a) Gazebo 仿真环境

(b) 实际环境

**图 6　降落环境搭建**

**Fig. 6　Training and testing environment**

### 4.1　无人机降落至静止平台

降落实验中, 静止平台比无人机稍大, 无人机具体尺寸为 0.4 m×0.4 m, 平台具体的尺寸为 0.6 m×0.8 m, 用于无人机降落。在仿真实验中, 搭建了一个简单的环境, 如图 6(a)所示。为了得到无人机与目标位置的相对信息, 在无人机的底部加装有摄像头传感器, 并通过 ROS 框架进行信息交互。无人机在这个仿真环境中训练和测试, 对于强化学习 PID 方法和应用矫正纠偏 COACH 方法的强化学习 PID 方法, 本实验对智能体进行了 200 回合的训练。当无人机降落至平台并保持静止后, 本实验认为无人机成功完成降落任务, 经过共 600 次仿真实验测试表明, 3 种方法都能够有效地 (>99%) 实现无人机降落任务。

3 种方法控制无人机降落的轨迹如图 7 所示。

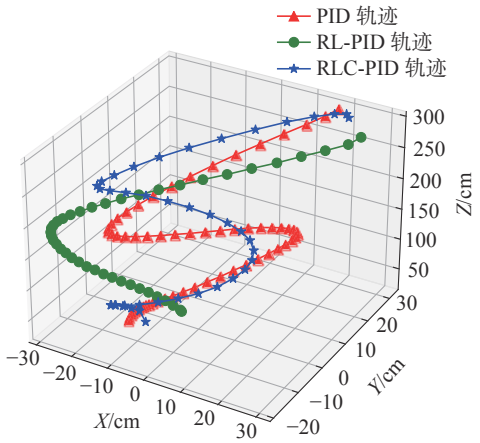



**图 7　3 种方法的降落轨迹（仿真）**

**Fig. 7　Trajectories of three approaches for landing in a simulated situation**

红色的轨迹为传统 PID 方法，本方法轨迹平滑，证明本方法的 3 个控制参数人工选择合理，能够有效完成无人机降落到固定平台的任务。绿色的轨迹是深度强化学习 PID 方法，本方法同样可以将无人机降落到规定的区域，但是在最终降落时，与静止平台中心的距离较大。蓝色轨迹曲线为应用矫正纠偏框架的深度强化学习 PID 方法，能迅速地对当前误差进行调整并且最终降落时与静止平台中心的距离较近，本方法有效应用了深度强化学习理论和 PID 控制方法，并使用矫正纠偏框架对两者进行结合，实现最佳的轨迹控制。无人机从坐标（0.3 m, 0.3 m, 3 m）出发，目标点为（0 m, 0 m, 0.2 m），3 种方法在各 200 次的测试实验中都能够有着较高的成功率（>99%），稳定完成无人机降落的任务。图 8 给出了 3 种方法控制下，无人机执行降落任务的时间，从结果上看，两种结合强化学习原理的方法能够有效地减少无人机降落的时间（传统 PID 方法时间平均值为 29 s，强化学习 PID 方法平均降落时间为 17 s，使用矫正纠偏框架的强化学习 PID 方法平均时间为 11 s），并且使用矫正纠偏框架，能够使得强化学习算法最终得到的策略控制效果更好，有效提升控制策略。最后，传统 PID 方法如果要达到控制要求，本身的参数是需要合理选择的，并且参数的选择是一个耗费时间的过程，在结合强化学习原理后，通过在虚拟环境中训练，可以实现智能体代替人类进行参数的选择，并且控制无人机降落效果比人类调节参数的 PID 方法效果更佳。

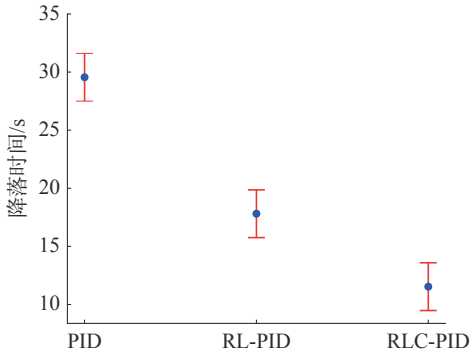



**图 8　3 种方法控制无人机至平台的时间（仿真）**

**Fig. 8　Time for UAV landing on a simulated static platform**

### 4.2　无人机降落至移动平台

实验设置从静止的平台改变为移动的平台，移动平台前进和后退并不断循环，其他的设置同之前静止实验相同，且使用了先前实验的训练模型和参数设置。

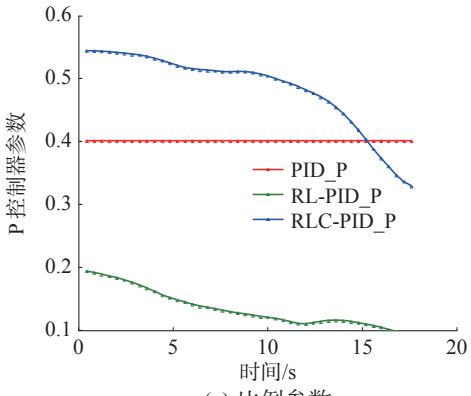
无人机初始坐标为（0.3 m, 0.3 m, 3 m），降落目标点坐标为（0 m, 0 m, 0.2 m），在移动平台降落实验中，3 种方法的成功率如表 1 所示。当平台移动后，由于各种不稳定性因素，无人机降落至平台的难度加大，固定参数 PID 方法的成功率在 99% 附近，结合强化学习原理的方法成功率分别是 89% 和 100%，本结果表明在训练合适的情况下，强化学习原理能够提高无人机降落的稳定性。

**表 1　无人机移动平台降落测试结果**

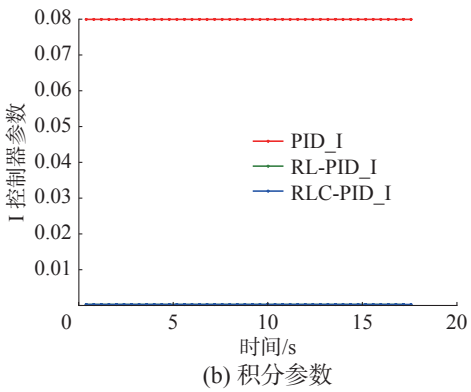
**Table 1　The result of UAV landing on a moving platform**

| 方法 | 成功率/% | 测试次数 |
|---|---|---|
| 传统PID方法 | 99 | 200 |
| 强化学习–PID方法 | 89 | 200 |
| 使用矫正纠偏框架的强化学习–PID方法 | 100 | 200 |

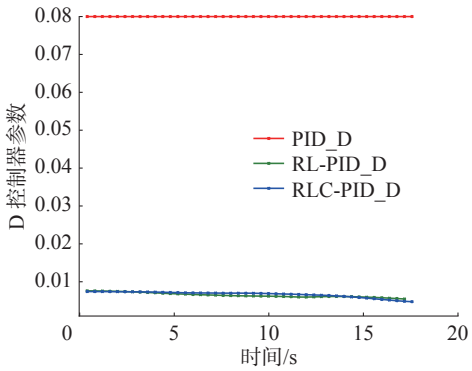
图 9 给出的是无人机降落过程中，3 种方法 PID 参数的变化情况，传统 PID 方法的参数是固定的，$k_P$、$k_I$、$k_D$分别是 0.4、0.08、0.08。使用矫正纠偏 COACH 方法的强化学习 PID 方法，$k_P$时刻改变，并且范围在区间（0.0, 0.6），$k_I$、$k_D$也在不断地更新来适应不同的环境。对于强化学习 PID 方法，$k_P$同样时刻改变，范围为区间（0.0, 0.2）。




(a) 比例参数




(b) 积分参数



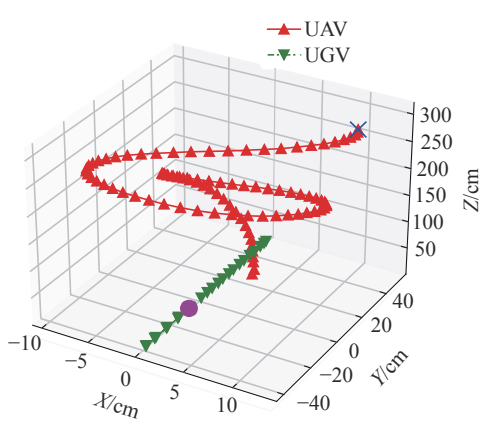

(c) 微分参数

**图 9　3 种方法的 PID 参数**

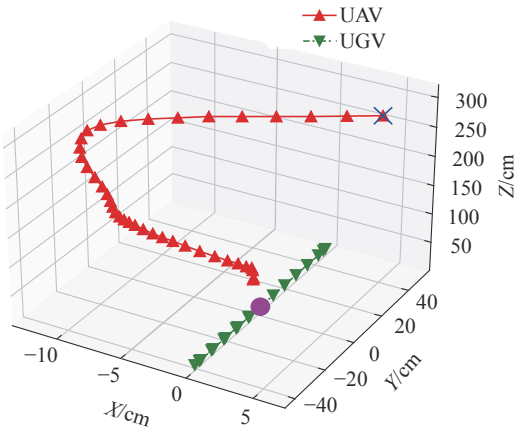
**Fig. 9　PID control gains in every sample time of three approaches**

两种基于强化学习原理的方法面对相同的降落问题, 同样的训练过程, 具有不同的效果, 主要的原因在于输出的 PID 控制参数不同, 应用矫正纠偏框架的强化学习方法能够在无人机处于高处时, 输出大的比例参数, 有助于跟随目标平台, 在处于较低的高度时, 输出较低的比例参数, 有助于无人机实现精准降落。

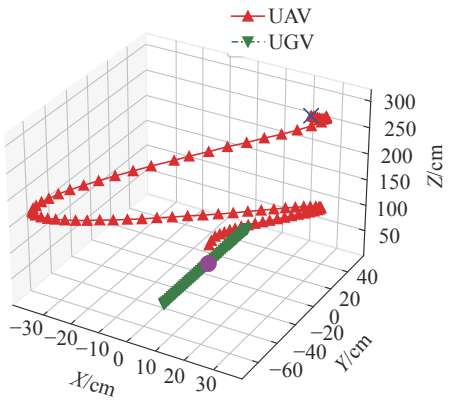
图 10 给出的仿真环境中, 无人机降落至移动平台的轨迹。如图 10(a)、(b)、(c) 所示, 蓝色“×”表示无人机降落的初始位置, 红色点线是 3 种方法控制下, 无人机降落至移动平台的轨迹, 紫色点表示无人机降落时, 平台的位置, 绿色线为移动平台的移动轨迹。无人机降落轨迹的终点与移动平台的终点距离是评判无人机降落效果的重要标准, 两者距离近则认为降落的效果优。图 10(b) 和 (c) 中无人机轨迹终点与移动平台的距离更近, 因此两种应用强化学习原理的方法有着更佳的控制效果。



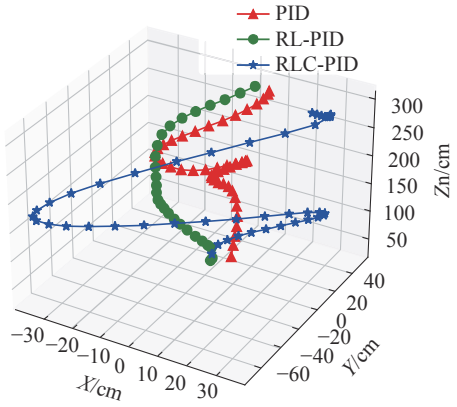

(a) PID 方法轨迹




(b) 强化学习方法轨迹




(c) 使用矫正纠偏框架的强化学习 PID 方法轨迹




(d) 3 种方法的轨迹对比

**图 10　3 种方法降落至移动目标**

**Fig. 10　Trajectories of three approaches for landing on a moving platform**

图 10(d) 是 3 种方法控制下无人机的降落轨迹对比图, 在高处, 使用矫正纠偏方法的强化学习 PID 方法控制无人机, 控制无人机的动作幅度大, 有助于无人机在高处跟随平台进行移动; 在较低高度, 本方法控制无人机的动作幅度小, 从而实现精准降落。实验结果表明应用矫正纠偏框架的强化学习 PID 方法既能及时跟随平台移动, 也能有比其他两种方法更高的降落精度, 因此, 使用矫正纠偏方法的强化学习 PID 方法在仿真环境中控制无人机降落的效果最佳。

### 4.3　无人机实体实验

在真实世界的设置与仿真环境中一致，并且模型的输入和输出也与仿真时相同。无人机降落的初始高度为 3 m，并且初始位置偏离目标平台 0.3 m。首先测试在静止平台上的降落，基于强化学习的方法不断输出动作，直到无人机降落至目标平台，具体轨迹如图 11 所示，无人机可以稳定地降落至目标平台的中心附近，在 20 次的测试中，无人机可以全部降落至目标点 0.3 m 半径内，认为无人机降落任务执行成功。

**图 11　无人机（实体）降落至静止平台**

**Fig. 11　Trajectories of our approach for landing on a static platform**

在协同系统执行任务时，大部分情况是移动或者不稳定的降落平台，因此，本文设置了无人机降落至移动平台的实验。实体实验设置为无人机降落的初始高度为 3 m，并且初始位置偏离目标平台 0.3 m，并且降落的平台在前后循环移动，速度在 0.05~0.1 m/s 范围内变化。如图 12 所示，实验结果证明使用矫正纠偏框架的深度强化学习方法能够控制无人机降落至移动的目标平台，有效完成移动平台在不稳定情况下的无人机降落任务。通过 20 次的测试，无人机都能够在平台移动的情况下，降落至偏离目标平台 0.4 m 半径内，认为无人机降落成功。综上，静止平台和移动平台的实验都证明了本文提出方法的有效性和稳定性。

**图 12　无人机降落至速度为 0.05~0.1m/s 的平台**

**Fig. 12　Trajectories of our approach for landing on a moving platform**

## 5　结束语

本文提出一种用于无人机降落的深度强化学习方法。上层使用在矫正纠偏框架下的深度确定性策略梯度方法，用于不断输出 PID 参数，提高 PID 方法的实用性，在底层使用 PID 方法，直接输出控制量，用于控制无人机实现降落任务。强化学习模型在环境中不断训练，不断输出 PID 参数值，区别于固定 PID 方法，获得的模型有更优的控制效果。矫正纠偏框架将人类经验应用于强化学习模型训练中，在人类的指导下，得到的强化学习模型控制无人机降落时间更短，降落成功率更高。仿真实验和真实实验的实验结果都表明本文提出的结合矫正纠偏 COACH 框架的深度强化学习 PID 方法能有效完成无人机移动平台降落任务。

## 参考文献：


[1] LIU P, CHEN A Y, HUANG Yinnan, et al. A review of rotorcraft Unmanned Aerial Vehicle (UAV) developments and applications in civil engineering[J]. Smart structures and systems, 2014, 13(6): 1065–1094.

[2] TSOUROS D, BIBI S, SARIGIANNIDIS P. A review on UAV-based applications for precision agriculture[J]. Information (Switzerland), 2019, 10(11): 349.

[3] REN H, ZHAO Y, XIAO W, et al. A review of UAV monitoring in mining areas: current status and future perspectives[J]. International journal of coal science & technology, 2019, 6(3): 320–333.

[4] MICHAEL N, SHEN Shaojie, MOHTA K, et al. Collaborative mapping of an earthquake-damaged building via ground and aerial robots[J]. Journal of field robotics, 2012, 29(5): 832–841.

[5] 王华鲜, 华容, 刘华平, 等. 无人机群多目标协同主动感知的自组织映射方法 [J]. 智能系统学报, 2020, 15(3): 609−614.
WANG Huaxian, HUA Rong, LIU Huaping, et al. Self-organizing feature map method for multi-target active perception of unmanned aerial vehicle systems[J]. CAAI transactions on intelligent systems, 2020, 15(3): 609−614.

[6] BACA T, STEPAN P, SPURNY V, et al. Autonomous landing on a moving vehicle with an unmanned aerial vehicle[J]. Journal of field robotics, 2019, 36(5): 874–891.

[7] TALHA M, ASGHAR F, ROHAN A, et al. Fuzzy logic-based robust and autonomous safe landing for UAV quadcopter[J]. Arabian journal for science and engineering, 2019, 44(3): 2627–2639.

[8] FENG Yi, ZHANG Cong, BAEK S, et al. Autonomous landing of a UAV on a moving platform using model predictive control[J]. Drones, 2018, 2(4): 34.

[9] RODRIGUEZ-RAMOS A, SAMPEDRO C, BAVLE H, et al. A deep reinforcement learning technique for vision-

based autonomous multirotor landing on a moving platform[C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Madrid, IEEE, 2018: 1010−1017.

[10] SHAKER M, SMITH M N R, YUE Shigang, et al. Vision-based landing of a simulated unmanned aerial vehicle with fast reinforcement learning[C]//2010 International Conference on Emerging Security Technologies. Canterbury, IEEE, 2010: 183−188.

[11] RODRIGUEZ-RAMOS A, SAMPEDRO C, BAVLE H, et al. A deep reinforcement learning strategy for UAV autonomous landing on a moving platform[J]. Journal of intelligent & robotic systems, 2019, 93(1/2): 351–366.

[12] LEE S, SHIM T, KIM S, et al. Vision-based autonomous landing of a multi-copter unmanned aerial vehicle using reinforcement learning[C]//2018 International Conference on Unmanned Aircraft Systems (ICUAS). Dallas, IEEE, 2018: 108−114.

[13] ARULKUMARAN K, DEISENROTH M, BRUNDAGE M, et al. Deep reinforcement learning: a brief survey[J]. IEEE signal processing magazine, 2017, 34: 26–38.

[14] HESSEL M, SOYER H, ESPEHOLT L, et al. Multi-task deep reinforcement learning with PopArt[J]. Proceedings of the AAAI conference on artificial intelligence, 2019, 33: 3796–3803.

[15] SEDIGHIZADEH M, REZAZADEH A. Adaptive PID controller based on reinforcement learning for wind turbine control[J]. World academy of science, engineering and technology, international journal of computer, electrical, automation, control and information engineering, 2008, 2: 124–129.

[16] WANG Shuti, YIN Xunhe, LI Peng, et al. Trajectory tracking control for mobile robots using reinforcement learning and PID[J]. Iranian journal of science and technology, transactions of electrical engineering, 2020, 44(3): 1059–1068.

[17] ASADI K, KALKUNTE SURESH A, ENDER A, et al. An integrated UGV-UAV system for construction site data collection[J]. Automation in construction, 2020, 112: 103068.

[18] ERGINER B, ALTUG E. Modeling and PD control of a quadrotor VTOL vehicle[C]//2007 IEEE Intelligent Vehicles Symposium. Istanbul, IEEE, 2007: 894−899.

[19] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning[EB/OL]. (2015−01−01)[2021−01−01]. https: //arxiv. org/abs/1509.02971.

[20] CARLUCHO I, DE PAULA M, VILLAR S A, et al. Incremental Q-learning strategy for adaptive PID control of mobile robots[J]. Expert systems with applications, 2017, 80: 183–199.

[21] CHOI J, CHEON D, LEE J. Robust landing control of a quadcopter on a slanted surface[J]. International journal of precision engineering and manufacturing, 2021, 22(6): 1147–1156.

[22] KIM J, JUNG Y, LEE D, et al. Landing control on a mobile platform for multi-copters using an omnidirectional image sensor[J]. Journal of intelligent & robotic systems, 2016, 84(1/2/3/4): 529–541.

[23] CELEMIN C, RUIZ-DEL-SOLAR J. An interactive framework for learning continuous actions policies based on corrective feedback[J]. Journal of intelligent & robotic systems, 2019, 95(1): 77–97.

[24] SILVER D, HUANG A, MADDISON C J, et al. Mastering the game of Go with deep neural networks and tree search[J]. Nature, 2016, 529(7587): 484−489.

[25] GRIGORESCU S, TRASNEA B, COCIAS T, et al. A survey of deep learning techniques for autonomous driving[J]. Journal of field robotics, 2020, 37(3): 362−386.

[26] HESSEL M, MODAYIL J, VAN HASSELT H, et al. Rainbow: combining improvements in deep reinforcement learning[EB/OL].(2017−01−01)[2021−01−01]. https: //arxiv. org/abs/1710.02298.

[27] SUTTON R S, BARTO A G. Reinforcement learning: an introduction[M]. Cambridge, Mass: MIT Press, 1998.

[28] WATKINS C J C H, DAYAN P. Q-learning[J]. Machine learning, 1992, 8(3/4): 279–292.

[29] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529−533.

[30] VAN HASSELT H, GUEZ A, SILVER D. Deep reinforcement learning with double Q-learning[EB/OL]. (2015−05−01)[2020−12−20].https://arxiv. org/abs/1509.06461v3.

[31] WANG Z, SCHAUL T, HESSEL M, et al. Dueling network architectures for deep reinforcement learning[C]// International conference on machine learning. PMLR, 2016: 1995−2003.

[32] KOUBAA A. Robot operating system (ROS): The complete reference[M]. volume 1. Cham: Springer, 2016.

## 作者简介:

张鹏鹏, 硕士研究生, 主要研究方向为空地协同系统、智能无人系统。

魏长赟，副教授，博士，博士毕业于荷兰代尔夫特理工大学人工智能专业，英国卡迪夫大学机器人及自主系统实验室访问学者，主要研究方向是自主智能无人系统。以第一作者发表学术论文 20 余篇，出版英文专著 1 本。

张恺睿，本科，主要研究方向为智能无人系统。

# 第四届国际高性能大数据暨智能系统会议
# The 4th International Conference on High Performance Big Data and Intelligent Systems

第四届国际高性能大数据暨智能系统会议（The 4th International Conference on High Performance Big Data and Intelligent Systems, HDIS 2022）拟于 2022 年 12 月 9 日至 12 月 11 日在中国天津举办。

会议旨在搭建高性能计算、大数据及人工智能领域高端前沿交流平台，促进海内外专家学者的交流与合作，推动智能技术进步和智能产业发展。本次会议将汇聚全球顶级专家、学者和产业界优秀人才，共同围绕国际热点话题、核心关键技术、产业发展及挑战等进行开放式研讨。

会议由中国计算机学会 (CCF)、中国人工智能学会 (CAAI) 联合主办，IEEE Computer Society 技术支持，天津理工大学、澳门大学、中国科学院半导体研究所、中国科学院深圳先进技术研究院、CCF 高性能计算专业委员会、CAAI 神经网络与计算智能专业委员会、CAA 模式识别与机器智能专业委员会、中国智能计算产业联盟共同承办。会议论文集将由 IEEE Xplore®出版，EI 收录，优秀论文将会推荐至 SCI/EI 期刊发表。热忱欢迎广大同仁踊跃投稿并莅临本届会议！

**投稿要求：**

1.论文未曾在国内外杂志或会议上发表。

2.稿件写作必须使用英文，并严格按照模板要求排版。

3.所有论文采用网上投稿，请访问会议官网进行投稿。

https://www.hdis.world/public/portal/list/index/id/9.html

**会议报名：**

请登录会议官网 http://www.hdis.world/，报名注册。

**重要日期：**

论文投稿截止日期：2022 年 9 月 15 日（已延期）

论文录用通知日期：2022 年 10 月 15 日

论文提交截止日期：2022 年 10 月 31 日

早鸟注册截止日期：2022 年 11 月 09 日

**联系方式：**

李老师，010-82304554，hpbdis@semi.ac.cn

薛老师，13920254011，xuewanli@email.tjut.edu.cn