PERSONAL RESUME

王立坤

年　　龄：24岁　　性　　别：男

籍　　贯：四川　　联系电话：15313246251

联系邮箱：wlk23@mails.tsinghua.edu.cn

研究方向：扩散模型，强化学习，世界模型，具身智能

教育背景

2019-09 ~ 2023-06　　**北京航空航天大学**　　**交通运输（本科）**

专业成绩：GPA 3.8/4 （排名 1/89）
主修课程：数学分析，概率统计，线性代数，数据结构，运筹学，自动控制原理， 交通工程学， 车辆技术基础
荣誉：国家奖学金，北京市优秀毕业生，校级三好学生， 年度优秀生， 学习优秀奖学金
竞赛：第十七届全国交通科技竞赛一等奖，第七届互联网加创新创业大赛北京市二等奖

2023-09 ~ 至今　　**清华大学**　　**人工智能（硕士）**

专业成绩：GPA 3.85/4
主修课程：深度学习, 强化学习, 最优控制, 车辆控制工程, 数值分析
荣誉： 2023年度优秀奖学金，2024年清华-陕汽特等奖学金，2023 年第四季度小研之星

科研成果

Off-policy Reinforcement Learning with Model-based Exploration Augmentation　　**NeurIPS2025 (一作)**

提出了一种世界模型驱动的适用于off-policy算法的探索机制MPGE：通过直接生成关键数据以供算法进行更新:

- 设计了基于classifier-guidance 的扩散状态生成器，通过策略熵、td误差等指标引导关键状态的生成，并通过理论证明了状态的可访问性和环境一致性
- 设计了可预训练的一步转移世界模型来拟合环境转移，既可以用于扩散模型的引导指标计算，又可以将关键状态转化成transition，用于下游算法的更新
- 设计了一种将MPGE与现有强化学习的联合训练机制，通过近似重要性采样实现原有算法性能和样本效率的双重提升，在Gym和DMC多项实验中达到sota

Bootstrap Off-policy with World Model　　**NeurIPS2025 (通讯作者)**

提出一种通过自举将在线规划和off-policy强化学习紧密结合的框架，由共同学习的世界模型提供支持以促进策略提升和策略评估:

- 设计了一种无似然对齐机制，将策略的更新直接对齐到规划器的非参数化动作分布，从而避免策略与规划器的分布偏移问题
- 设计了在经验回放中对高回报的动作进行优先学习的策略更新方法，降低次优动作对策略训练的干扰，减少规划器生成的低质量样本对策略更新的负面影响，提升稳定性和收敛效率

Enhanced DACER Algorithm with High Diffusion Efficiency　　**ICLR2026（在投，共一）**

提出了DACER-Pro，一种高效的在线强化学习算法，通过引入Q梯度引导和时间步加权机制加速扩散策略，仅用五个扩散步骤便可在复杂控制任务上达到业界顶尖（SOTA）性能，并展现出更强的多模态能力。

Mind Your Entropy: From Maximum Entropy to Trajectory Entropy-Constrained RL　　**ICLR2026（在投，共一）**

提出了一种基于**轨迹熵控制**的新型强化学习框架 TECRL，通过在整条控制序列上分配与约束探索强度，实现了对策略随机性的全局化、时序自适应调节，为最大熵强化学习带来了新的优化视角和强大的跨领域泛化潜力，在gym任务上基于最大熵强化学习框架实现了性能增长。

IDC-Ped: An Extended Social Force Model for Urban Pedestrian Behavior Modeling　　**TITS (在投，一作)**

针对现有自动驾驶仿真中行人行为不智能、无法处理复杂人车交互场景的现状，提出了一种高保真的智能行人微观行为模型:

- 设计了一种特殊的人车交互机制，在保证安全性的情况下，使行人的行为更加真实，在自动驾驶仿真实验中表现良好
- 设计了一种针对城市道路自动驾驶仿真的行人微观行为建模模型，通过对传统社会力模型改进，增强行人于车辆的交互能力

DACER: Diffusion Actor-Critic with Entropy Regulator　　**NeurIPS2024 (二作)**

针对现有强化学习算法中的策略近似函数限制策略表达能力的现状，提出了一种利用扩散模型来表征策略的方法:

- 将反向扩散过程视为一种新型的策略近似函数， 通过实验表明该策略函数具有良好的多模态性质
- 设计了一种使用混合高斯模型的方法对扩散策略的熵进行估计，并通过熵调节器的方法来引导扩散策略的探索，在多项基准实验中达到 SOTA

Controllability Test for Nonlinear Datatic Systems

Communications in Transportation Research（IF=12.56，二作）

提出了一种适用于数据驱动控制系统的可控性检验方法，通过引入一种称为 邻域可控性的新概念，将可控性的定义从传统的点对点形式扩展到点对区域形式，使其更适用于那些由有限数量数据描述的动态系统。

ODE-based Smoothing Neural Network for Reinforcement Learning Tasks　　**ICLR2025（spotlight）**

针对深度强化学习技术在控制任务中应用时面临的一个主要挑战——控制动作的平滑性不足，提出了一种改进的神经常微分网络，增强了抗扰动和平滑处理能力。

项目经验

2025-02 ~ 至今　　基于强化学习的端到端自动驾驶算法开发　　算法开发

项目介绍：通过目标集特征的输入，使用强化学习作为策略训练端到端自动驾驶
开发职责：设计了plannet编码框架，使用transformer encoder进行多维度状态表征，将高维输入编码到稳定的特征隐空间上使用强化学习进行训练
结果：该模型在多车道和交叉路口仿真中表现良好

2025-07 ~ 2026-01　　基于世界模型的自动驾驶强化学习训练框架　　方案设计，代码开发

项目介绍：（滴滴出行实习项目）通过设计双耦合的RSSM世界模型，通过混合损失训练下游强化学习算法，从而实现面向自动驾驶场景的强化学习算法
开发职责：设计世界模型结构和联合策略损失，并进行代码开发
预期结果：可部署的世界模型强化学习自动驾驶方案

2024-08 ~ 至今　　基于强化学习的自动驾驶数据确权方法设计　　方案设计、主持开发

项目介绍：针对现有自动驾驶数据可复制下载和数据确权的矛盾，设计了一种基于强化学习的鲁棒水印嵌入方式，使得在对原始数据影响最小的情况下完成水印的嵌入和识别
开发职责：设计了一种基于ADP的强化学习水印嵌入框架，通过在同一循环中使用两套MDP描述水印的嵌入和识别过程，并引入还原误差和识别误差进行对抗生成训练
结果：在工业控制系统中表现良好，目前在带领同学进行面向自动驾驶数据的数据进行训练和调优

2022-09 ~ 2023-06　　面向城市道路场景的交通参与者行为建模　　算法开发

项目介绍：开发面向城市道路场景的自动驾驶仿真平台(LasVsim)
开发职责：仿真平台交通流模型的设计和编写
结果：通过使用改进的mobil模型编写了可用于交叉路口多向冲突的车辆微观行为模型，实现了车辆在交叉口的真实行为复现

2022-01 ~ 2022-06　　基于强化学习的分层式区域交通信号控制　　方案设计，主持开发

项目介绍：通过设计多智能体强化学习控制算法，实现区域交叉路口的信号配时方案设计与开发
开发职责：设计了双层优化结构，底层结构实现单交叉口的基础信号控制，上层结构实现区域联合优化
结果：获得2022年全国大学生交通科技大赛一等奖

学生工作

1.在本科期间担任北京航空航天大学交通学院191313小班班长，学生会办公室部长，参与举办多次活动
2.在研究生期间担任清华大学车辆与运载学院体育部部长，院学生会副主席以及硕23班的组织委员，负责学生体育工作，策划、参与、承办多项体育赛事和体育活动

技能特长

- **语言能力：**大学英语6级528分，GRE319分。有熟练进行英语交流和读写的能力。
- **计算机：**熟练使用python，pytorch，git，letax等开发工具
- **团队能力：**具有丰富的团队组建与扩充经验和项目管理与协调经验。