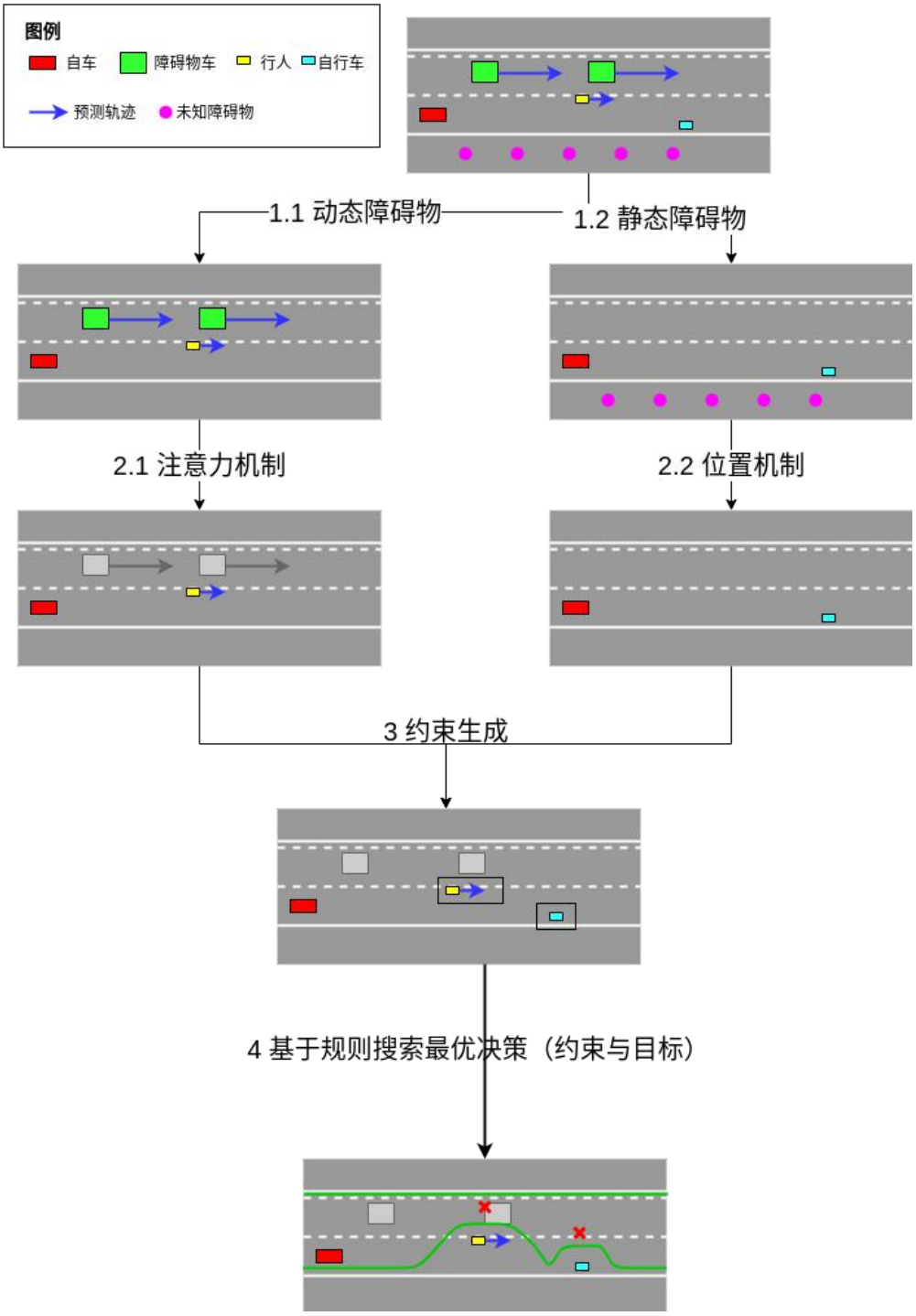
方案：基于监督学习的最优决策选择

目的：本文档给出一种最优决策选择的方案与计划的流程，供实验开展和实车部署测试

目录

一、背景：

1.1 障碍物器决策局限性：

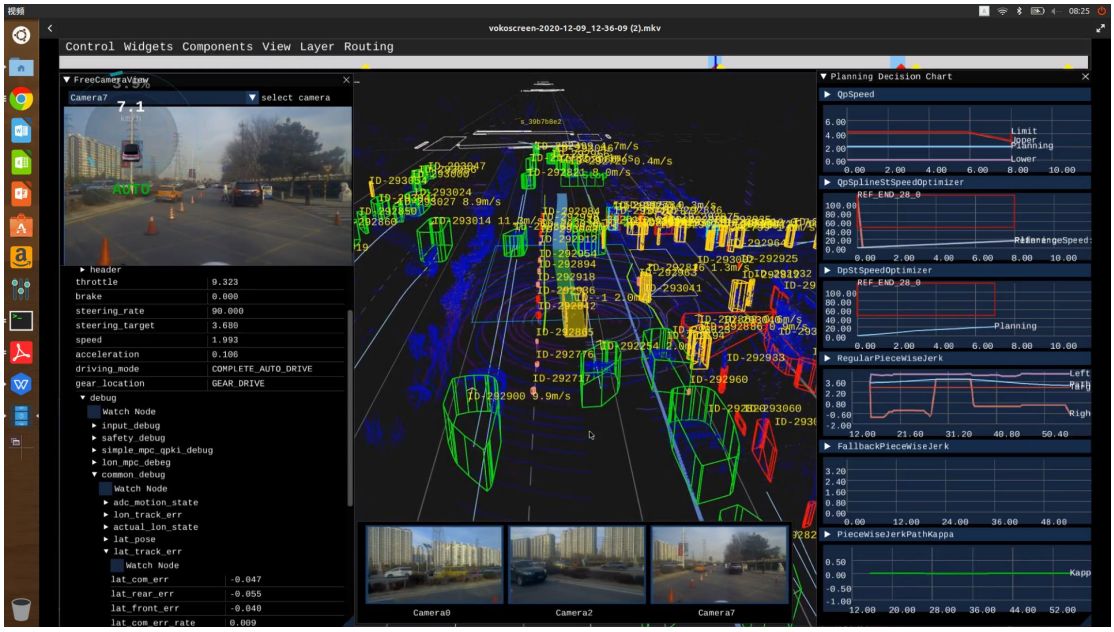


只能生成规则定义下的单一的“最优决策“

基于搜索的 Cost function 人工调整难

复杂场景中决策不像人，更多 cost 项待加入

1.2 问题案例：



决策不像人：复杂场景下 cost 设置不合理

二、目标

2.1 主要功能

目的：设计并训练出一个合适的 cost function 用于评估候选决策，并选择一个稳定、像人、少接管的决策序列。

2.2 预期收益

提高复杂场景中决策更像人程度

提高稳定性

减少相关接管数

免去人工调 cost 过程

三、模型方案

3.1 基于 max margin 的 IRL

IRL: 逆强化学习，从专家演示中学习内在的 reward/cost function，s.t.专家决策的 reward function 最大

Max margin：一种简单的实现 IRL 的方式 s.t.专家决策序列得分比其他决策序列得分高

s: state，自车+环境状态

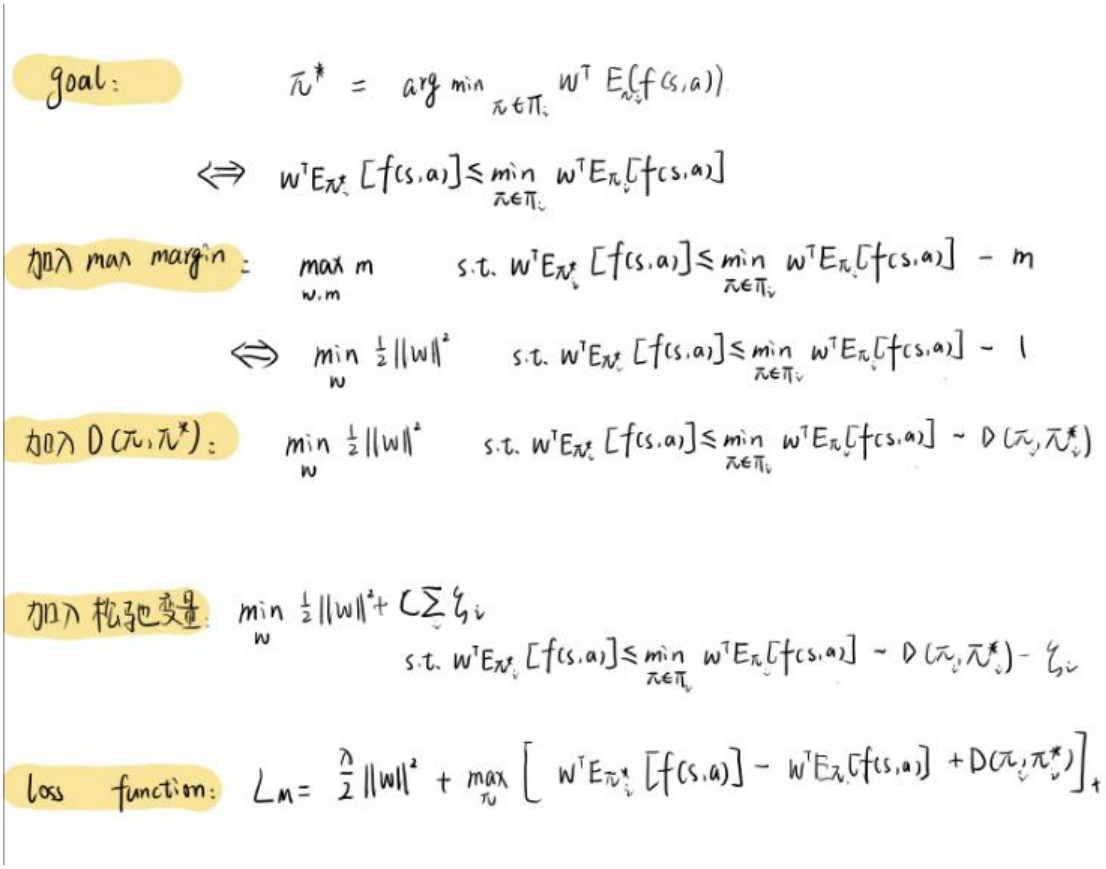
a: action，障碍物决策/anchor point

f: feature，1 x n vector，f(s,a)，由全局特征和 s,a 决定

w: weight，1 x n vector

c: cost = w∧t * f(s,a)

feature 项	描述	当前 baseline 是否已有
绕行是否有足够空间	bool	有
决策是否历史决策一致	-1,0,1	有
绕行方向是否与 centerline 一致	bool	有
绕行是否曲率超限	bool	有
绕行方向是否响应借道	bool	有
绕行可行空间大小	double	无
绕行曲率大小	double	无
决策与路口距离	double	无
决策与障碍物距离	double	无
自车 heading	double	无
偏离中心线距离	double	无
自车与障碍物 s 方向速度差	double	无
是否人行道区域	bool	无



Goals： 专家决策序列 cost 最小

加入 maximum margin：专家决策序列得分比其他决策序列 cost 越小越好

（变换同 SVM 的优化目标变换，margin 距离为 1/||w||∧2）

加入 D 衡量决策序列间的 difference：惩罚不像专家决策的，D 用 anchor point 拟合曲线的 L1 距离衡量

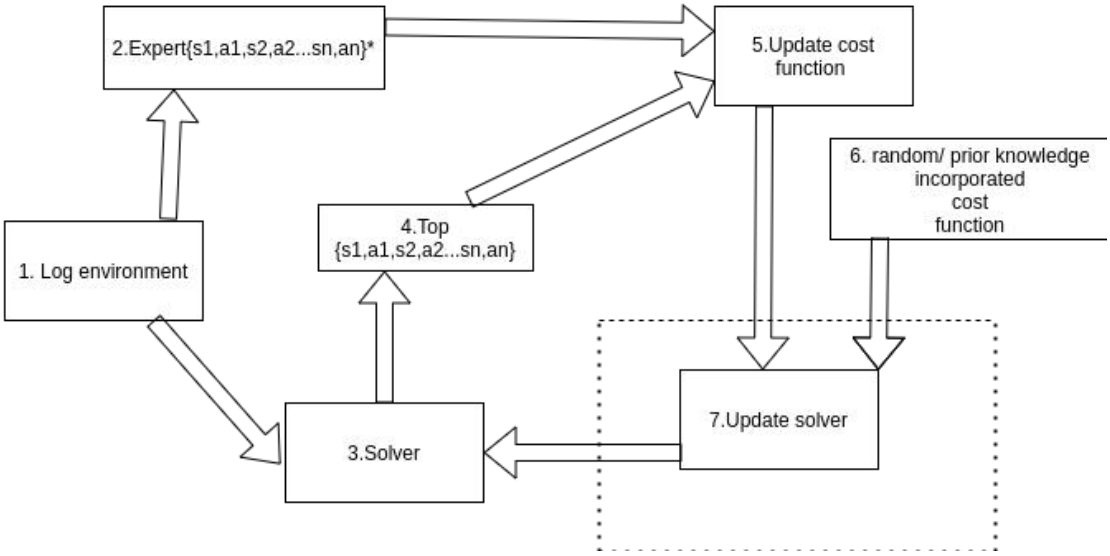
加入 slack variable：允许专家决策 suboptimal

loss fucntion： （类似 SVM 利用 hinge loss 求解），另一种角度去看是惩罚不像专家决策序列的 cost（经验风险），并加入 L2 正则项（结构风险）

3.2 模型训练

利用专家远遥数据，采用监督学习，输出 cost function 的 weight

3.2.1 Pipline



1.专家远遥数据提取

2.Expert 模块：根据专家轨迹 ground truth 自动标注，还原专家决策序列

3.Solver 模块：基于规则生成多条决策序列

4.Top 模块：选出得分最高（cost 最小）的决策序列

5.Update w：次梯度下降训练

6.初始化 w：可以用 baseline 权重初始化

可选：7.更新 solver 生成当前 weight 下最优的决策序列

3.2.2 数据集

3.2.3 特征工程

四、模型评估

4.1 评估步骤

4.1.1 模型验证

专家远遥数据建立验证集

4.1.2 仿真测试

跑过 badcase 收集制作的 logsim

4.1.3 影子模式

AB test: 和 baseline 分析比较

4.1.4 实车测试

进行实车测试评估

4.2 评估指标

专家决策/轨迹相似度

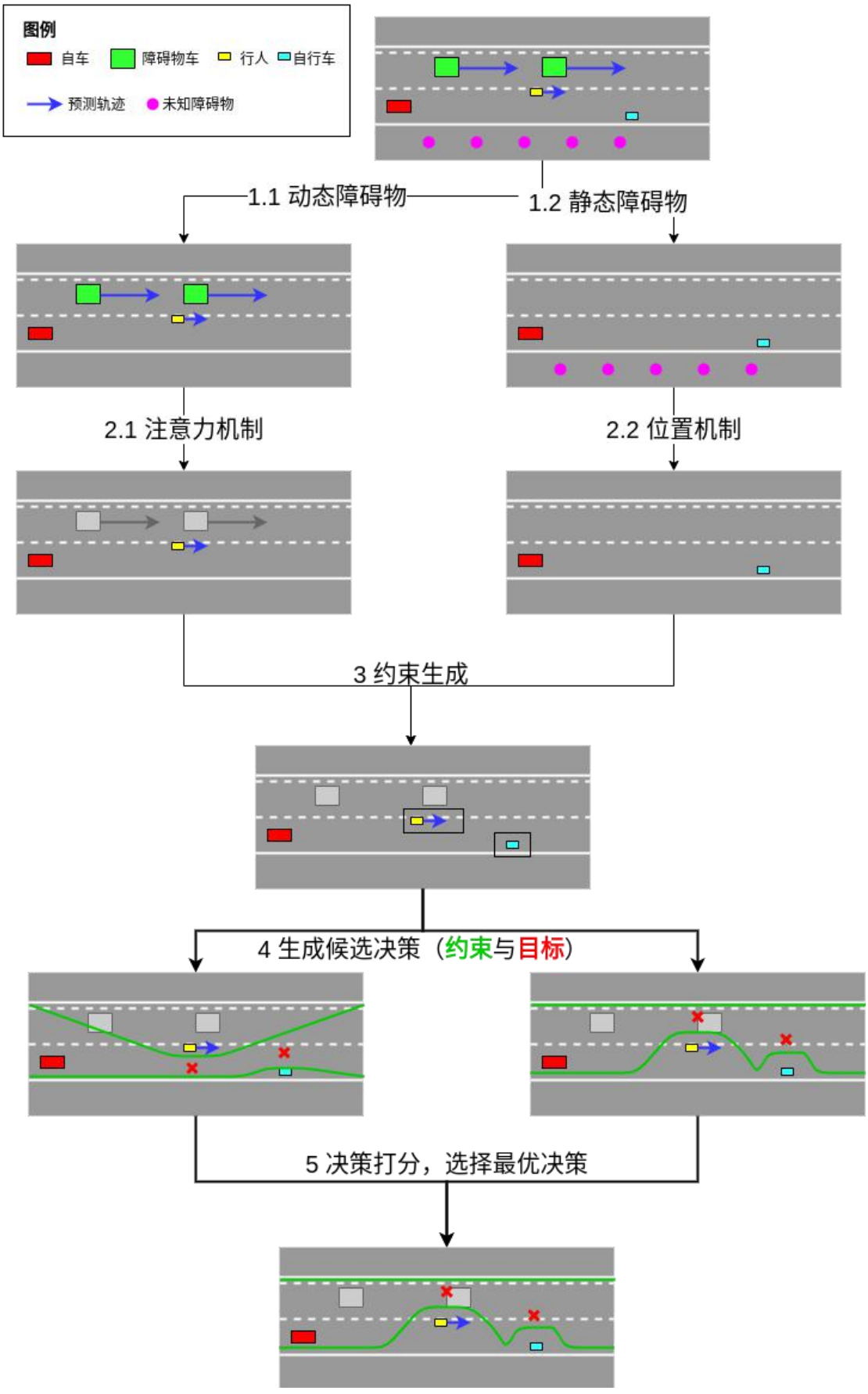
决策稳定性

绕障距离

五、模型部署

改造当前障碍物决策器，支持多个候选决策的输出

按训练出的 cost function 选择最优决策（对应下图障碍物决策流程中的 4,5）


图例
自车
障碍物车
行人
自行车
预测轨迹
未知障碍物
1.1 动态障碍物
1.2 静态障碍物
2.1 注意力机制
2.2 位置机制
3 约束生成
4 生成候选决策（约束与目标）
5 决策打分，选择最优决策

六、实验一览

内容	目标	任务
文档维护	能够快速让人上手， 能够管理记录每一次的实验	算法流程清单
		障碍物决策 baseline 逻辑
训练集	确保数据可用，准确	提取远摇数据，增加专家驾驶下的数据集
		数据集的分析，构建均衡的训练集
		自动标注，按专家轨迹还原决策
特征处理	设计并提取重要的特征	输出特征设计 wiki
		开发离线特征提取代码
模型设计及训练	调研合适的方法并设计方案	给出整体设计方案，并进行训练和实现
	跑通一个 demo	
	实现模型	
	训练模型	
前后处理	架构适配	支持多个决策的输出
	baseline 的实现	实现基于人工的代价函数的设计
	影子模式	开发影子模式，上线双跑采集数据
评测	探索如何根据评估标准找到优化方向	组织 evaluation 分析 输出分析问题的流程和迭代方向判读的方法
	logsim 评估	构建 30 个 logsim，并用 baseline 验收 根据分析方法，对模型进行分析
	worldsim 评估	补充一个泛化的 worldsim，并用 baseline 验收 根据分析方法，对模型进行分析

	影子模式 评估	根据线上数据，设计分析方法，进行分析 尝试从影子模式补充数据
	实车测试	