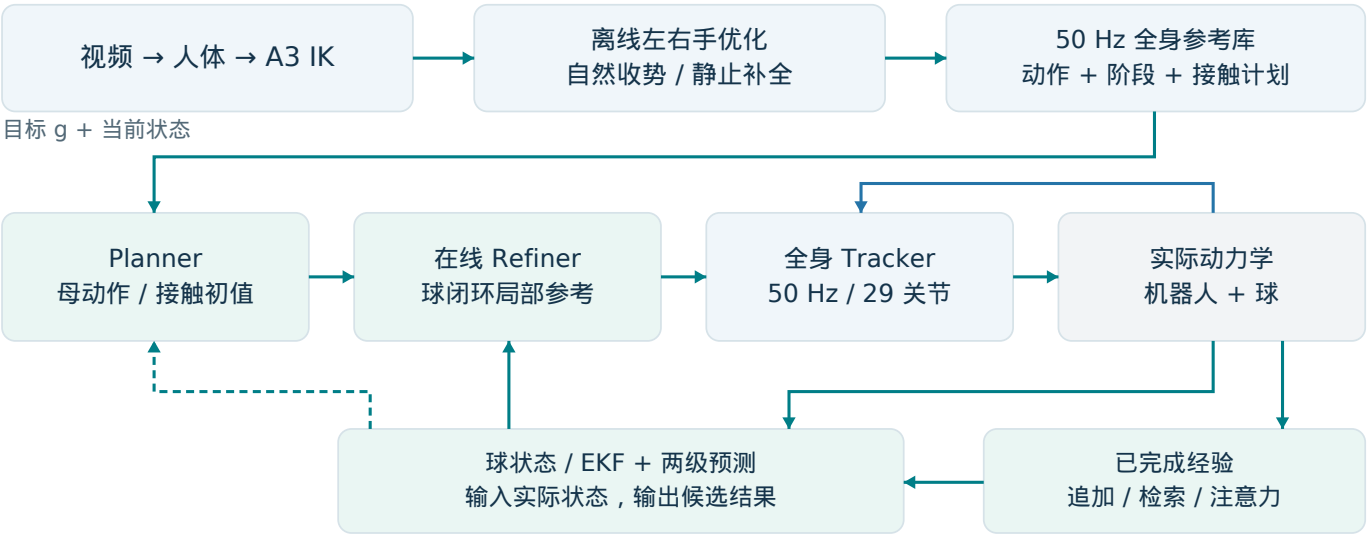


总体数据流与任务接口

实现边界：视频处理、离线优化与双分支 Tracker 已有实现；带球动态耦合、学习型 Planner、在线 Refiner 和经验注意力为后续设计。文中候选层宽是待测配置。



模块边界

模块	输入	输出
离线参考管线	视频、球桌几何、A3 模型、工具参数	50 Hz 全身参考库、阶段和接触标注
Planner	发球目标、机器人状态、参考库、已完成经验	母动作、站位/抛球参数、初始接触计划
在线 Refiner	当前球与机器人状态、参考、结果预测	接触附近的连续参考修正
全身 Tracker	实际本体状态、修正后的当前与未来参考	29 维关节位置目标，50 Hz
预测器与经验库	下发参考、实际执行、接触及终局结果	候选结果预测、检索上下文、监督样本

统一任务定义

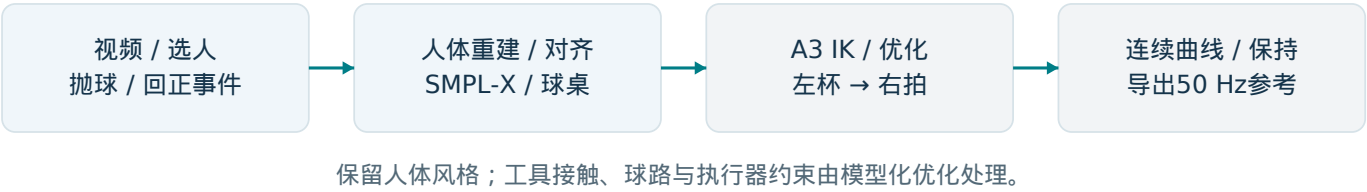
落点使用对方台面的连续矩形区域；旋转和速度在对方首次落台反弹后 20 ms 读取。己方先落台、过网、对方落台是事件约束。

字段	维数	定义
球路目标 g	16	希望落在哪里：矩形边界4；希望什么旋转：6个类别开关；希望多强：上下限2；球速上下限2；是否考核强度/速度的开关2
动作条件 g_pose	4	目标发球站位 x、y、yaw；离杯后的抛球高度。用于后续多目标 Planner
旋转类别	6 类	无旋、上旋、下旋、左侧旋、右侧旋、预定义混合旋；混合比例另由配置固定

任务空间以球桌中心地面投影为原点，+X 朝对方、+Y 向左、+Z 向上；单位 m、s、rad、rad/s。磁盘四元数 wxyz；网络朝向用旋转矩阵前两列的 6D 表示。侧旋符号与上下旋在球运动方向/重力构成的坐标系中统一计算。

视频到可执行参考库

离线数据处理。人体动作提供风格；球的释放、击球与结果由物理模型计算，不从人体重建中推定。



步骤	处理	产物
选片与选人	保留准备、真实抛球、挥拍、自然落地回正；裁去远距离走入与下一次回球。跟踪发球者，不跨人补帧	源帧映射、bbox、事件标注
人体重建	YOLO 连续选人；GVHMR 官方 ViTPose/HMR2 预处理；SMPL-X 姿态与根部轨迹。动作内缺失片段单独处理/过滤	30 Hz 人体动作
球桌对齐	四桌角与已知桌面尺寸估计相机；Y-up 转 Z-up；整段刚体对齐。左拍人体先统一镜像，A3 固定左杯右拍	桌坐标人体动作
A3 retarget	GMR 映射与缩放；Mink/DAQP IK；骨盆、躯干、腕、脚与整臂姿态目标；关节限位/限速、自碰与桌地间隙	根部与 29 关节轨迹
左手优化	标注准备姿态；杯口朝上；杯内放球一次；优化上抬、撤杯及离杯时序；自然释放，不给球注入速度	可用抛球轨迹/击球窗口
右手优化	在来球窗口内搜索接触时刻、拍面方向与接触点速度；中央双面拍接触；全接近段禁止提前蹭边	满足球路目标的局部变种
连续性与力矩	保留整臂姿态；连续曲线重拟合/重定时；检查位置、速度、 $M(q)qdd + bias - passive$ 与电机力矩限制	连续参考及约束记录
自然收势与保持	保留原视频回正；从已落稳姿态平滑减速；末帧速度清零。准备等待独立插入并平滑接回原动作	完整参考及阶段时钟

导出合同

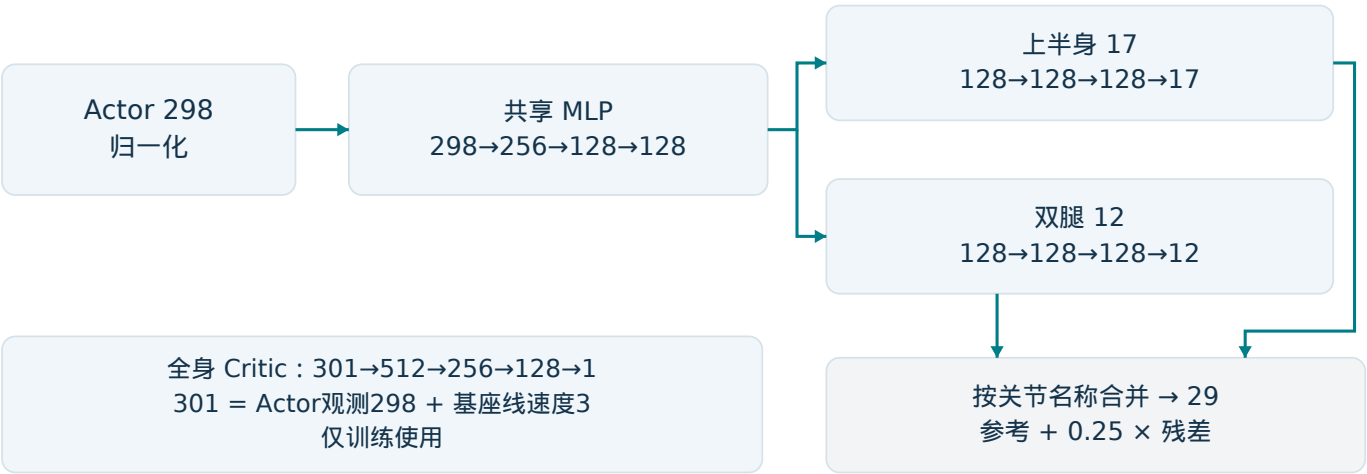
每条参考包含 $time[T]$ 、 $q[T,29]$ 、 $qd[T,29]$ 、 $root_pose[T,7]$ 、身体 FK、左右工具位姿/速度、阶段、接触窗口与 $policy_clock[T,2]$ 。按名称排列关节和身体；统一重采样到 50 Hz，再检查插值区间。

球轨迹单独保存 p 、 v 、 ω 、接触/离杯/过网/落台事件及有效性；不混入无球人体风格标签。参考重新定时或修改手臂后，重新计算抛球与击球配合。

离线优化复用已有 S1-S5 方法；自然恢复来自视频，并在其后补零速保持。不重新合成逐脚回 initial。

全身 Tracker 网络

当前实现。一个 Actor、两个同时工作的分支、一个全身 Critic。后续训练 Refiner 时冻结 Tracker 与归一化统计。



隐藏层 ELU；两分支始终同时工作；上半身独占腰部3关节。

Actor 输入	维数	字段
当前参考	67	29个关节希望到达的角度29、希望的角速度29；机器人相对参考根部的位置3、朝向6
本体/上次动作	93	身体当前转动速度3、重力在身体坐标中的方向3；29个关节当前角度29、当前角速度29；上一次策略残差29
上半身预览	136	腰和双臂未来4个时刻的目标角度/角速度：4×17×2；时刻为0/40/100/200 ms
参考时钟	2	动作进度、剩余动作时间；不累计静止等待时间
合计	298	Critic 另加仿真基座线速度 3，合计 301

隐藏层 ELU，动作头线性。上半身 17 = 腰 3 + 双臂 14；下半身 12 = 双腿。按关节名称合并，关节只有一个输出者；无路由器、GRU 或持久隐状态。

输出定义：Actor的29维是“每个关节相对参考再偏多少”的无量纲残差；乘0.25 rad并加参考后，才是29维电机目标角度，不是速度或力矩。Critic的1维是未来回报估计，不发给电机。

```
q_target(t) = q_ref_next(t) + 0.25 × a_tracker(t)
训练：29 维状态无关高斯；执行：取均值。
```

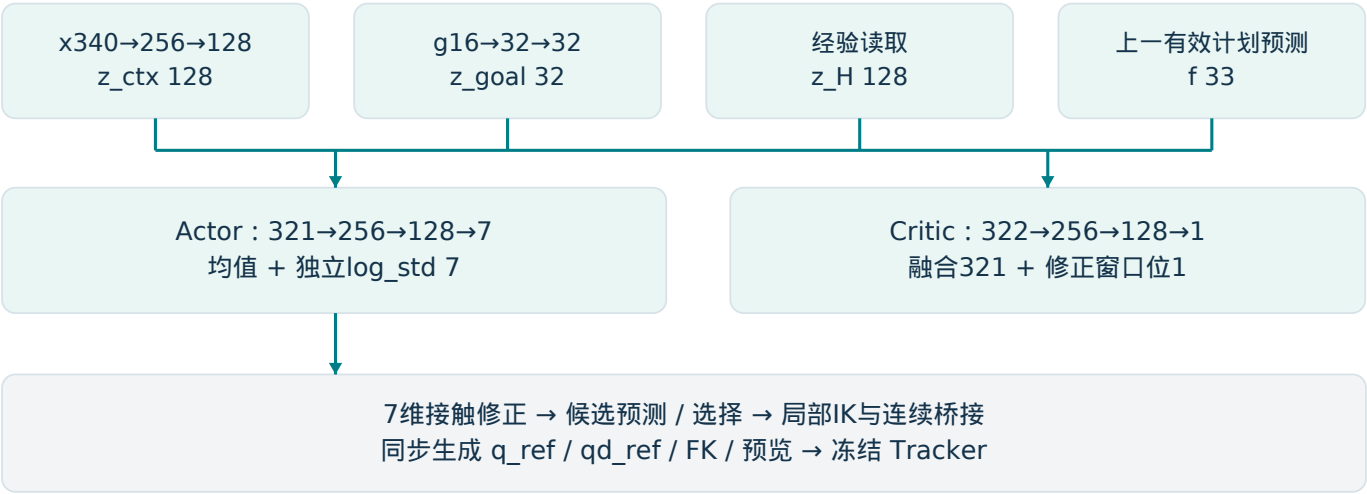
静止与动作时钟

准备等待保持第一帧及零速，进度固定为 0；释放等待后平滑恢复动作时钟。自然收势结束后固定末帧、速度归零、时钟 [1,0]，同一 Tracker 继续控制，不循环、不因参考播完重置。

准备时保持侧身、杯口向上；末段躯干正对球桌。静止段以低速度、低漂移、支撑和不倒为主，允许固定姿态偏差；挥拍段仍约束右拍位置、朝向与速度。

在线 Refiner 网络与参考生成

拟新增。首版只修正右臂接触附近的参考；保留原左手抛球、腰部与双腿参考。Tracker 仍控制全身。



输入	维数	定义
o_T	298	上页的机器人当前状态、参考动作、未来预览和动作进度
球估计 b	22	球在哪里3、飞多快3、绕什么轴转多快3；各分量的不确定性9；距最近观测多久1；位置/速度/旋转是否有效3
计划接触 c	13	还多久碰球1；在哪里碰3；拍面朝向6；接触处拍面的移动速度3
u_prev	7	上一次实际选中的：触球时间调整1、拍面转动调整3、接触点速度调整3
物理上下文 x	340	将机器人、球、接触计划和上次修正拼接；E_ctx：340→256→128，ELU
目标 g	16	E_goal：16→32→32，ELU；位于 Actor/Critic 模型内

Actor 融合 z_ctx128、z_goal32、z_H128、上一有效计划预测 f33，共 321 维。Actor：321→256→128→7；独立 Critic：321 + 修正窗口位 1→256→128→1。两者目标编码器独立；首版 E_ctx、经验与预测模块在 PPO 内固定。

z_ctx128、z_goal32、z_H128是网络压缩后的状态/目标/经验特征，不是新增传感器。f33表示“预计球路结果15 + 结果不确定性15 + 三个事件概率”；Critic窗口位表示现在是否允许球相关修正。

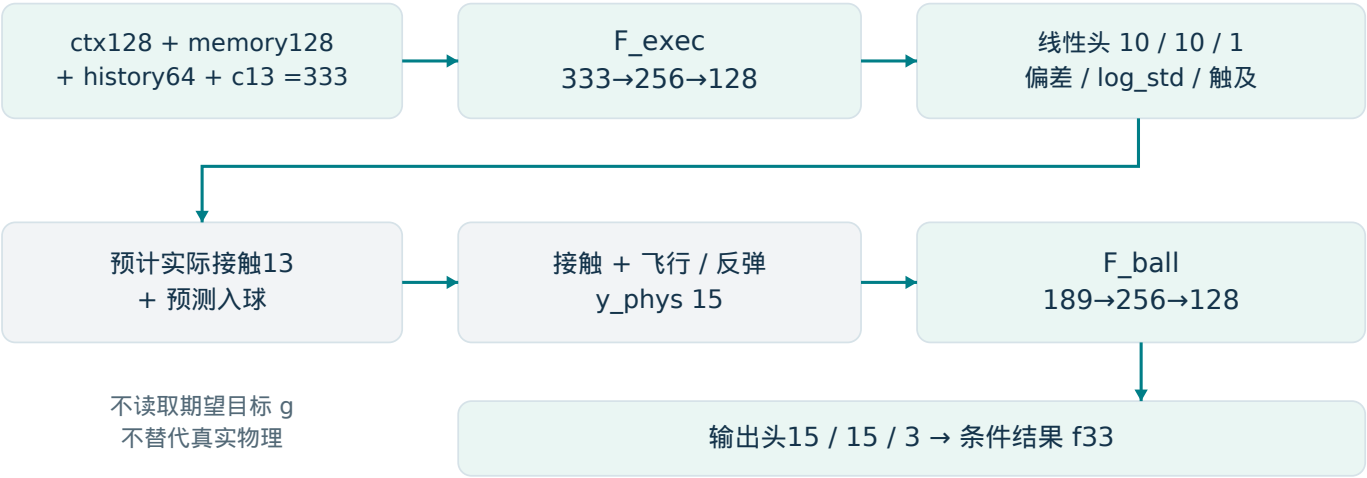
输出 u 共7维：触球提前/推迟多少秒1、拍面转多少弧度3、接触点速度改变多少m/s 3。朝向用旋转向量复合；接触位置由预测来球与时刻求得。高斯变量z经scale×tanh(z)限幅，scale取离线可执行范围。

确定性 Refiner：u → 接触目标 → 局部 IK/连续轨迹桥接 → q_ref、qd_ref、FK 与预览同步刷新。保留整臂姿态、边界位置/速度/加速度及关节约束；修正始终相对母动作/现行计划定义，不逐帧无界累加。

候选首版设 8 个：零修正、Actor 提案及固定局部扰动；冻结预测器评分后只执行一个。Refiner 调用频率待测 25/50 Hz；Tracker 固定 50 Hz。

球状态与两级结果预测

拟新增网络，复用现有球物理。已知球位置与旋转可观测；EKF 融合测量并估计速度、协方差和短时未来轨迹。



模块	网络	输出/用途
有限误差历史	最近 8 帧×64 展平 512→128→64，ELU	每帧：29个关节角误差、29个关节角速度误差、身体角速度3、重力方向3；每回合重置
F_exec	输入 333→256→128；并联线性头 10/10/1	接触偏差均值10、log_std10、触及概率1
接触/飞行基线	预计实际接触 + 预测入球 → 瞬时冲量 + 飞行/反弹	物理结果 y_phys15，不包含期望目标 g
F_ball	输入 189→256→128；并联线性头 15/15/3	结果残差15、log_std15、触球/过网/有效发球 logits3

F_exec 输入 = z_ctx128 + z_H128 + z_hist64 + c13 = 333。接触偏差10 = 时间1 + 位置3 + 旋转向量3 + 接触点速度3。用偏差均值重建预计实际接触 c_actual13。

F_exec预测“计划和实际会差多少”；其log_std10表示这10项偏差的不确定性，触及概率1表示能否碰到球。F_ball预测“预计会打出怎样的球”，不把用户希望的结果当观测。

F_ball 输入 = c_actual13 + 接触不确定性10 + b22 + z_H128 + y_phys15 + 触及概率1 = 189。连续均值为 y_phys + delta_y；输出 f = 均值15 + log_std15 + 事件概率3，共33维。

15维结果与接触接口

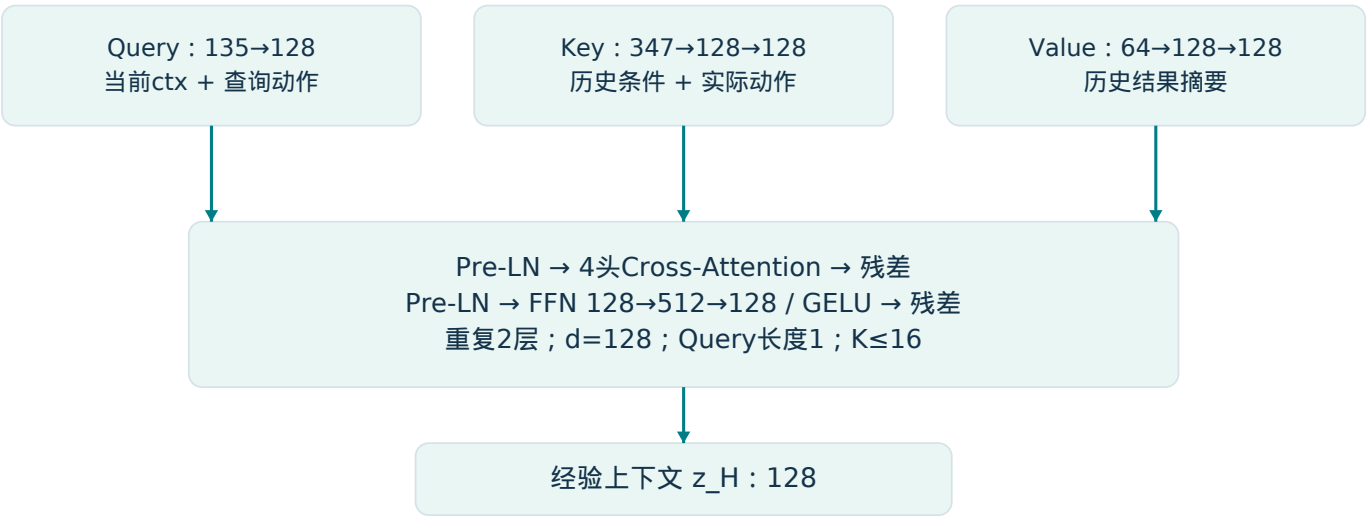
y = 出球速度3 + 出球旋转3 + 对方首次落台XY2 + 该次反弹后20 ms速度3 + 旋转3 + 过网净空1。不存在的落点或事件用掩码，不以零值代替。

球拍接触点速度 = 球拍线速度 + omega_racket × 接触点相对拍体质心的位置。双面拍使用朝向来球的法向；速度与旋转由同一冲量联合计算，拍边单独处理。

球：直径40 mm、质量2.7 g、薄壳惯量 I=2mr²/3。真实仿真保留球杯、球拍、球台、球网、飞行和反弹；预测结果不写回球状态。全身现有物理步5 ms；带球接触子步另行确定。

显式经验库与注意力读取

拟新增。磁盘保存完整尝试，内存保存近期摘要；新经验入库与网络参数更新分开执行。



张量	编码	语义
Query	[z_ctx128,u_query7] : 135→128	“现在是什么状态、准备怎样修正”，用于提问
Key	[历史x340,实际u7] : 347→128→128	“过去在什么状态下，实际用了什么修正”，用于匹配
Value	结果摘要64 : 64→128→128	“那次碰球/球路偏了多少，哪些事件发生了”，用于读取经验
读取	2层 Cross-Attention，d=128，4头，每头32；FFN 128→512→128，GELU	单查询长度1，历史长度K≤16；输出 z_H128

每层：Pre-LN → 交叉注意力 → 残差 → Pre-LN → FFN → 残差。读取最多8条近期记录，其余按状态/动作相似度检索补齐并去重；空历史直接返回零向量。

输出z_H128是读完这些经历后得到的128维经验特征，交给Planner、Refiner和预测器；它本身不是关节动作。mask表示某项数据是否真的被观察到，age表示经验距当前有多久。

摘要64 = 实际修正7 + 接触偏差10 + 相对物理基线的结果残差15 + 事件3 + 各标签有效位28 + 经验年龄1。原始目标、完整执行轨迹和所有决策时刻另存，不压缩丢弃。

写入、读取与参数更新

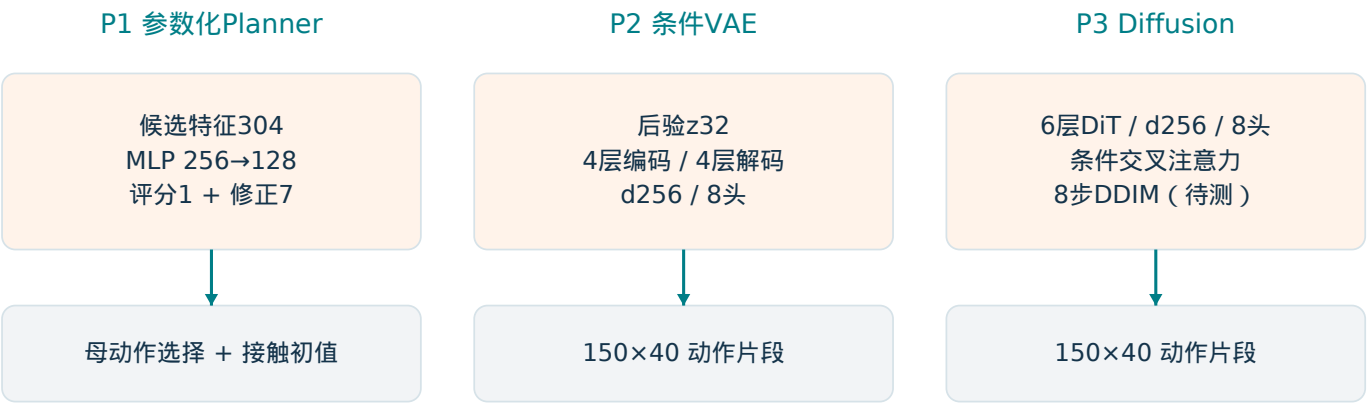
一球内：有限滑窗每步更新；模型权重固定。击球完成并分离后停止新的球相关修正，沿连续后缀自然恢复；球物理继续运行并产生终局标签。

一球后：把已完成试球加入近期缓存和磁盘库，再用于下一球。不能读取本球未来结果；无接触只监督事件，实际成功/失败都保留。输入同样的状态和显式记忆得到同样的条件，不维护持续漂移的 GRU 状态。

待测配置：M0 固定库/固定预测器；M1 新经验入库/固定预测器；M2 新经验入库，并在球间更新预测器均值头。M2只更新 F_exec 的128→10和 F_ball 的128→15，共3225参数；每球间1-5个小批量步，混合近期与基础样本。没有可靠接触标签时仅更新球路头。

Planner：三种待测结构

Planner 输出母动作及接触初值，或直接输出动作片段；之后都经连续 Refiner 和同一 Tracker。静止等待/末端保持由统一协议处理。



候选	具体结构	输出
P1 参数化规划	检索最多8条母动作；逐候选拼接 ctx128 + memory128 + goal32 + pose16 =304；共享 MLP 304→256→128；并联评分1、修正7	softmax选择母动作；初始7维接触修正。完整轨迹由母动作+确定性 Refiner 得到
P2 条件VAE	动作编码器4层 Transformer，d256/8头/FFN1024；潜变量z32。条件304→256；解码器4层 Transformer，d256/8头/FFN1024	H×40参考；训练均值/方差各32，重建+KL；推理采样z
P3 Diffusion Transformer	噪声动作 Linear40→256；6层，d256/8头/FFN1024；时间步嵌入+条件交叉注意力；Linear256→40	H×40去噪参考；首个候选配置8步DDIM采样

P1 中 pose16 来自 g_pose4→16→16；球路目标仍为16→32→32。当前物理上下文按候选母动作构建；查询经验不包含尚未执行的结果。

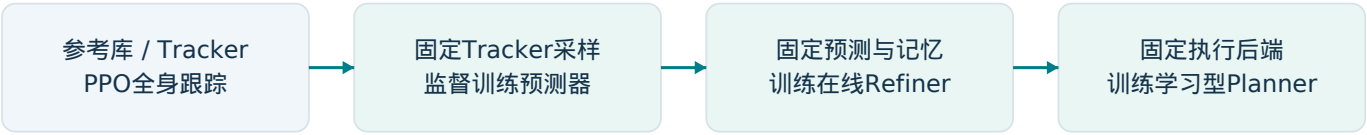
P2/P3首个候选H=150、50 Hz，即未来3秒。每帧40维：29个目标关节角；骨盆目标XYZ 3；身体整体朝向的6数编码；左右脚预计是否支撑2。它们是参考目标，不是实际观测。关节目标速度、身体/工具FK和时钟随后统一生成。

生成条件包含当前状态、目标、候选母动作上下文与历史摘要；接触计划/阶段时间随母动作重定时。P2后验池化特征256与条件256拼接，经512→128生成mu32/log_var32；解码器以150个时间查询和[z32,condition256]投影后的上下文生成动作。

首轮每球规划一次。滚动重规划作为待测变种：频率2/5 Hz，保留已执行前缀并用0.2秒连续桥接；实际接触结束后不再更新击球计划。站位改变需要连续走位参考，不直接平移正在执行的机器人。

训练顺序、标签与执行协议

只规定模块训练关系与数据合同，不启动新训练。已有无球 Tracker、离线球回放与后续带球闭环使用不同结果标签。



实际结果回写经验；一球内模型固定，球间可选更新预测器均值头。

阶段	更新模块	目标与数据
参考与Tracker	全身Actor/Critic ; RSL-RL5 PPO	跟踪、准备低速、挥拍工具误差、支撑与自然恢复；DR覆盖物理/电机/延迟/噪声/初态/外力/参考与球桌
带球数据与预测	E_ctx/E_hist、经验编码、F_exec/F_ball	冻结Tracker；优化器附近候选实际执行。有效连续标签用Huber/NLL，事件用BCE
在线Refiner	7维Actor、目标编码器、独立Critic	优化器候选监督初始化，再PPO；固定Tracker、预测器、经验编码与归一化
学习型Planner	P1或P2/P3及其条件编码器	P1动作选择+修正回归；P2重建+KL；P3噪声回归；补充学习者实际历史上的教师标签

执行时序

准备等待 → 达到可放球姿态 → 杯内放球一次/落稳 → 自然抛球 → 离杯后球闭环修正 → 实际触球并分离 → 自然收势/持续保持。漏球在可达窗口结束后转恢复。球路评分与身体恢复并行。

球闭环退出时，不把已采用修正突变清零；沿当前修正轨迹接回自然恢复。原始人类视频只约束风格，任务结果由同一次动态执行的球状态决定。

经验标签与损失约束

每球保存：母动作ID、目标、机器人/球初态、逐步x和原采样z、实际采用u、下发参考、实际接触、y15、事件/有效位、身体完成结果。落点区域损失在区域内为0；旋转按类别/强度、速度按启用区间评分。

优化器候选提供动作监督；未执行候选不提供成功结果标签。成功执行可按实际结果重标注目标；失败保留为结果/事件样本，不当作达到原目标的正确动作。

PPO记录原采样z及其log_prob；候选生成/重排是固定确定性动作映射，另存实际u。Refiner Actor/熵损失仅覆盖击球修正窗口，Critic保留至球路/身体结果结束的时间序列；rollout保存当时z_H和f，不用后续经验重算。

同次执行的所有帧与重标注副本属于同一数据划分；当前查询不检索自身终局。连续量按训练统计/任务容差归一化，缺失保留mask。新增PPO差异在A3包内继承RSL-RL5，不修改Isaac Lab或依赖源码。