双脑协调机制

03132015/SQ

目录

序

无论是双脑协调（图 1 中的 CCA），还是复合型世界模型（图 2 中的 “价值模型”），都需要在多个通道之间取得平衡。

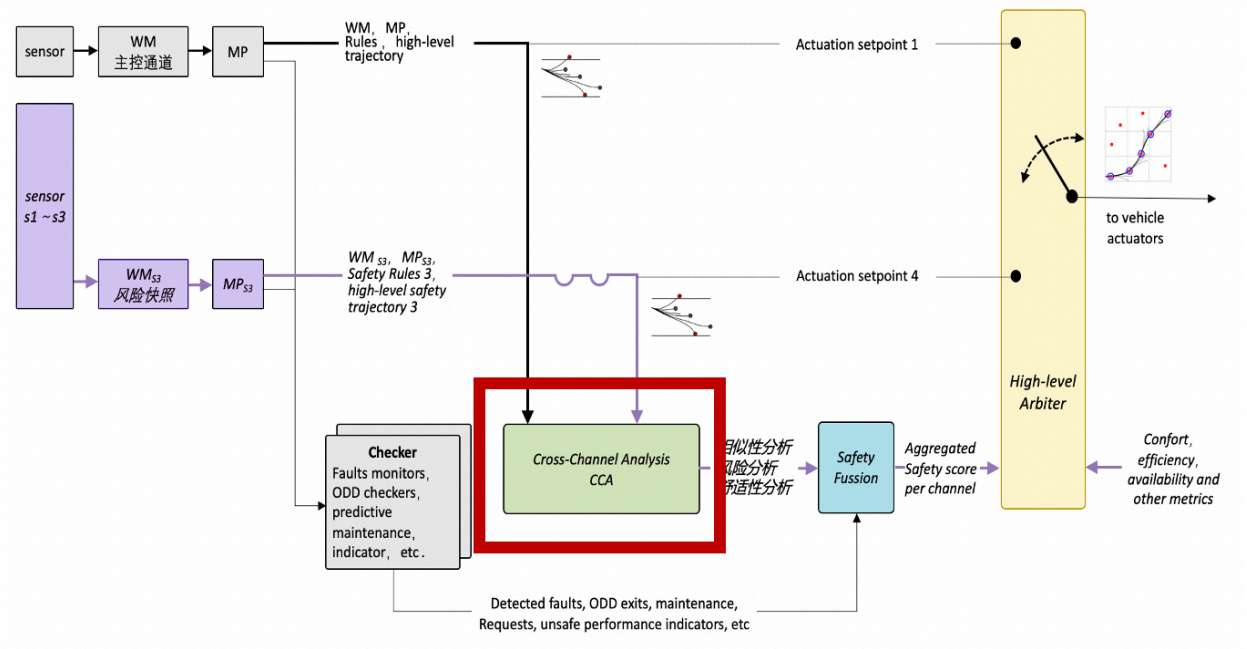

图 1

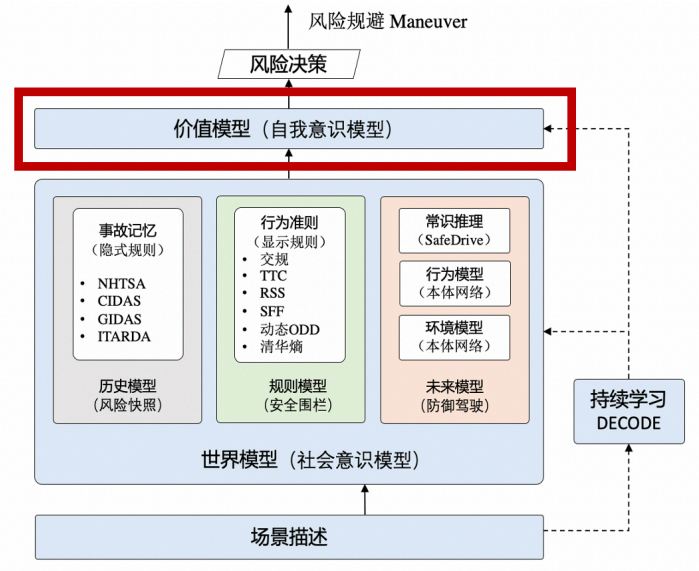

图 2

由于左右脑的思维方式、任务不同，会产生多方面的冲突，如果不化解这些冲突，右脑不但不会发挥作用，反倒会因为右脑介入而带来额外的危险。冲突类别包括（不限于），见表 1

表 1 左右脑冲突举例

冲突类别		举例	
		左脑	右脑
目标冲突		追求效率，优化驾驶体验	追求安全，避免风险
		基于数据优化驾驶策略	基于知识图谱判断安全性
		允许一定程度的试探性驾驶	需要严格遵守安全标准，避免危险
行为策略冲突	变道与超车	认为变道可行，能优化行驶时间	判断盲区可能有隐匿车辆，认为应等待更安全的时机
	跟车与车距控制	希望保持较小车距，以优化通行效率	认为前车可能突然刹车，应保持更大安全车距
	交叉路口通行	希望在无红灯但有行人等待的情况下提前通过，以减少通行时间	认为行人可能随时横穿，应降低速度或停车观察
风险评估冲突	突发事件处理	不一定能处理罕见情况，如突然掉落的障碍物	认为应立即制动或变道避让
	环境适应性	在特定城市或道路上优化驾驶风格	不允许因地域性驾驶习惯调整安全标准
伦理与法规冲突	碰撞不可避免时的决策	可能会选择对自身伤害最小的方案	可能会选择尽量保护行人或弱势交通参与者
	自动驾驶 vs. 车规合规性	可能学习到某些驾驶习惯（如跟随其他车压线超车），以优化效率	坚守法规，认为不应违反任何交通规则
计算资源冲突	实时性 vs. 计算复杂度	实时计算最优路径，需要高算力支持	计算量较大，在复杂环境下可能影响反应速度

参考文献[1]提出了 safe shell 架构和相应的多世界模型融合方案，提供了一个解决问题的入手点，但是对安全大脑的多个子模型而言，判据和裁决方法过于单一简陋，留有太多的变量需要在试验中摸索，可能导致大量试验资源投入，存在失败的风险。

双脑协同机制的设计也许比功能单元的开发难度更大，决定了终局的成败。

除了 safe shell 方案以外，本文提出多种可能的协调机制，不代表工程方案，只探索可行的路线，供制定研发计划参考。

本文所述双脑融合功能开发完成以后，能实现如下效果：

1）右脑不追求高精度，风险单元预测精度 > 60~70% 即为可接受的起点基础水平，即可发挥双脑效应

2）右脑输出的结论的同时需要附带有置信度信息（也就要求有自知之明）；

3）只推送可靠信息（不轻言、不信谣不传谣）；

4）在特定环境下，能够知道左脑和右脑的高阶规划哪个更可信。

1 协调

在尽可能的范围内，左右脑的差异最好能得到调和。调和方法有三种，分别为：权重动态调整机制、双向博弈机制、和层级仲裁机制

1.1 权重动态调整机制

协调的基本原则是：在不同场景下，左右脑的决策权重不同：

- 低风险环境（如高速公路顺畅路段）：左脑主导决策，右脑仅做背景监督。
- 高风险环境（如城市复杂路况、交叉路口）：右脑提高权重，可能调整左脑的决策边界。因为 KG 有场景针对性设计，所以在某些环境下置信度提高。
- 极端情况（如突发事故）：右脑直接接管，采取强制制动等操作。这个过程与“左脑放弃-fallback”的交接过程类似，但是改成为“左脑放弃-右脑接管-fallback”三段式接管

1.1.1 基于风险评估的自适应权重调整

核心思想是右脑实时计算当前驾驶场景的风险评分（Risk Score, P_risk），对低风险场景 → 左脑自主权更高；对高风险场景 → 右脑决策权重增加，甚至强制接管。

P_risk 方法包括确定性评估方法、概率性评估方法、基于势场理论的方法、基于可达集的方法等，然后对评估结果进行加权综合，得到最终的 P_risk[3][4][5][6]。

应用：设定风险阈值（Thresholds）策略（仅供示例）：

P_risk	风险等级	左脑权重	右脑权重
0.3	低	80%	20%
0.3 ≤P_risk < 0.7	中等	50%	50%
≥ 0.7	高	20%	80%
P_risk ≥ 0.9	极端	0%	100% 接管

案例：

- 高速公路（P_risk = 0.2）：左脑权重 80%，右脑主要提供辅助监督，不干涉常规驾驶决策。
- 交叉路口左转（P_risk = 0.8）：右脑权重 80%，若检测到盲区来车，右脑可直接干预或强制停车

1.1.2 基于环境复杂度的动态权重调整

核心思想是右脑根据环境感知的复杂度调整左右脑决策权重：简单场景（少量动态目标）→ 左脑决策主导；复杂场景（大量动态交互目标） → 右脑增强控制权，保证 OS 安全。

实施策略：设定环境复杂度量化指标（C_env）（仅供示例）：

C_env	复杂度	左脑权重	右脑权重
< 0.3	低，如高速公路顺畅路段	90%	10%
0.3 ≤C_env < 0.7	中，如郊区道路	70%	30%
≥ 0.7	高，如市中心拥堵路口	20%	80%

$$C_{env} = \text{f（动态目标数量, 道路类型, 交通信号, 交互强度）}$$

案例：

场景：普通直线路段 vs. 市区复杂十字路口

- 直线路段（$C_{env} = 0.2$） → 左脑执行决策，右脑仅监督突发风险（如前方车辆急刹）。
- 市中心路口（$C_{env} = 0.9$） → 右脑增强控制权，确保左脑不会做出激进决策。

1.1.3 基于基于驾驶行为稳定性的权重调整

核心思想是：右脑持续监测左脑的驾驶稳定性（S_stable），如果左脑在一段时间内没有错误决策，右脑逐步降低干预强度；如果左脑出现多次高风险决策，右脑重新提升决策权重。

可以注意到，safe shell 里面也强调了历史稳定性，但是方法比较简单。

实施策略举例，利用驾驶行为评分 S_stable（根据过去 N 次决策进行评估）：

S_stable	稳定度	右脑权重调整
> 0.8	高：转向平稳、加速度变化小、跟车稳定	-20%
0.6 ≤S_{Stable} ≤ 0.8	中，存在轻微不稳定行为，如较急的变道或刹车	保持默认权重
0.4 ≤S_{Stable} ≤ 0.6	低，经常出现急刹、急加速、剧烈变道	+20%
S_{stable} < 0.4	极不稳定，明显存在激进或鲁莽驾驶行为	+30%

案例：左脑在变道时表现稳定 vs. 频繁出现不安全变道

- 过去 100 次变道，安全率 95%（S_stable = 0.95） → 右脑干预减少，左脑更自由。
- 过去 100 次变道，误判率 30%（S_stable = 0.4） → 右脑增加权重，甚至直接否决部分左脑变道决策。

上述稳定指数评分方法如下[7][8][9]:

数据采集内容为：

评估指标	定义	单位
转向平稳性（Steering Smoothness）	方向盘角度变化速率（jerk），过快或剧烈调整可能导致不稳定	°/s²
车速稳定性（Speed Stability）	速度变化率（acceleration/deceleration），过大则不稳定	m/s²
跟车稳定性（Following Stability）	车距变化率，是否出现急刹、急加速等情况	m/s
变道稳定性（Lane Change Stability）	变道时的角速度变化、加速度变化是否平稳	m/s², °/s²
制动平稳性（Braking Smoothness）	制动力度的变化，急刹会降低稳定性	m/s²
驾驶一致性（Driving Consistency）	车辆在类似场景下做出的决策是否一致	-

驾驶行为稳定性（S_{stable}）计算方法

驾驶行为稳定性 S_{stable} 通过多个子指标的加权计算得到：

$$S_{stable} = \alpha S_{steering} + \beta S_{speed} + \gamma S_{following} + \delta S_{braking} + \epsilon S_{consistency}$$

- $S_{steering}$：转向平稳性得分，基于方向盘角度变化速率计算。
- S_{speed}：速度稳定性得分，基于车辆加速、减速变化计算。
- $S_{following}$：跟车稳定性得分，基于车距波动计算。
- $S_{braking}$：制动平稳性得分，基于制动力度的变化计算。
- $S_{consistency}$ ：驾驶一致性得分，基于过去相似场景的决策一致性计算。

计算举例：

方向盘角度变化速率计算：

$$S_{steering} = \exp(-k_1 \times \text{Jerk})$$

其中，Jerk 代表方向盘角度的时间变化速率（°/s²），k_1 为超参数

速度变化率计算：

$$S_{speed} = \exp(-k_2 \times |\text{Acceleration}|)$$

其中，Acceleration 代表车辆加速度的变化（m/s²）

车距变化率计算：

$$S_{following} = \exp(-k_3 \times |\Delta D|)$$

其中，ΔD 代表前车与本车的车距变化（m）

综合计算：

$$S_{stable} = 0.25 S_{steering} + 0.25 S_{speed} + 0.2 S_{following} + 0.15 S_{braking} + 0.15 S_{consistency}$$

1.1.4 基于法规与伦理约束的权重调整

核心思想是：某些情况下，右脑强制用法规约束否决左脑决策；在事故不可避免时，右脑接管，选择符合社会伦理的方案。

案例：黄灯时左脑希望加速通过，右脑认为应减速

- 法规判断：通过黄灯风险 > 50% → 右脑权重 100%，强制减速。
- 通过黄灯安全性高（无侧方车辆） → 左脑权重 80%，允许通过。

1.1.5 最终决策权重 W_left & W_right 计算公式

$$W_{left} = \alpha(1 - P_{risk}) + \beta(1 - C_{env}) + \gamma S_{stable}$$

$$W_{right} = 1 - W_{left}$$

其中

- α、β、γ 是调整系数，可根据实验优化
- 风险评分、环境复杂度、驾驶稳定性共同决定权重分配

1.2 双向博弈机制

采用类似于多智能体博弈（Multi-Agent Game Theory）的方法

- 用右脑对左脑的决策进行风险打分，若超过阈值则要求调整决策
- 左脑可以提供反证数据，表明某些情况下可以接受更激进的策略
- 采用强化学习 + 安全约束机制，使左右脑在长期博弈中逐步达成最优权衡点

1.2.1 Multi-Agent Game Theory（多智能体博弈论）

Multi-Agent Game Theory（多智能体博弈论） 结合博弈论（Game Theory） 和 人工智能（Multi-Agent Systems, MAS），主要研究多个智能体（Agents） 在共享环境中的决策交互。

在安全大脑中，每个小世界模型都可以被看作是一个智能体 Agent，都有自主决策能力，互相协作的过程也是多智能体博弈博弈的过程，左右脑之间是多智能体关系，复合子模型之间也是多智能体关系。每个世界模型（副通道）之间的关系可能是：

- 合作型（Cooperative）：左右脑之间、各个子模型之间共享目标，相互协作以达成最优解
- 竞争型（Competitive）：左右脑的目标互相冲突，例如自动驾驶中的车流竞争

- 混合型（Mixed）：智能体既有合作成分，也有竞争成分，例如自动驾驶中的双脑协同架构

1.2.1.1 博弈类型主要有如下几种

- 合作博弈（Cooperative Game）
- 非合作博弈（Non-Cooperative Game）
- 零和博弈（Zero-Sum Game）
- 非零和博弈（Non-Zero-Sum Game）

在双脑架构下，左脑（执行驾驶） 和 右脑（安全监督） 之间的冲突可以通过多智能体博弈论进行优化：

- 左脑希望提高通行效率（如快速通过黄灯）。
- 右脑希望最大化安全性（如减速等待绿灯）。
- 通过非零和博弈机制，左右脑在仿真环境中进行策略博弈训练，最终找到最优平衡点，避免极端激进或极端保守驾驶。

1.2.1.2 可以考虑采用的关键优化方法

1）纳什均衡（Nash Equilibrium）

- 任何一方都无法单方面改变策略而获得更好的收益，达到相对稳定的平衡状态。
- 例如，左脑与右脑达到稳定驾驶策略，既安全又高效。

2）强化学习 + 博弈论（Reinforcement Learning & Game Theory）

- 结合深度强化学习（Deep Reinforcement Learning, DRL），使左右脑在仿真环境中不断对抗 & 协同，最终找到最优解。

3）演化博弈（Evolutionary Game Theory）

- 通过模拟进化，逐步筛选出更优的策略，使自动驾驶系统能够不断适应新环境。

多智能体博弈理论提供了处理复杂决策冲突的数学基础，特别适用于双脑架构的自动驾驶系统。通过非零和博弈 + 强化学习，可以让左脑（执行）与右脑（监督） 在安全性和驾驶效率之间达到最优平衡[10]~[14]。

1.3 层级式仲裁系统

采用三层决策逻辑，主要思想是：

1）左脑执行主导决策，右脑提供策略引导（默认状态）。

2）右脑发现潜在风险时，左脑需进行风险再评估（如重新计算轨迹）。

3）若左右脑意见仍冲突，“协调模式”结束，系统进入“仲裁模式”，按照以下判据进行**裁决**。

2 仲裁

当左右脑对同一场景的决策存在冲突时，需要依赖**一套裁决标准**来决定采取哪一方的策略。关键判据包括：风险阈值判据、统计对比判据、规则优先级判据、机器学习反馈判据

2.1 风险阈值判据

右脑使用**风险评估模型**（基于知识图谱 + 数据分析）评估左脑方案的风险分值：

- 若风险评分 < 预设安全阈值，左脑决策有效。
- 若风险评分 ≥ 预设安全阈值，右脑可**否决左脑方案**并提出替代方案。

风险系数计算见 1.1.1.1，参考文献另见[15][16][17]。

2.2 统计对比判据

统计对比判据（Statistical Comparison Criteria）通过历史驾驶数据和仿真训练数据进行统计分析，主要依赖于历史数据、事故统计、驾驶安全性评估来评估左右脑的决策优劣[18][19]，比如：

- 在相似场景下，左脑的决策事故率是否高于右脑推荐方案？
- 在过去驾驶数据中，左脑方案是否导致了更多的急刹、急变道等危险行为？
- 右脑推荐的方案是否显著降低了历史数据中的风险？

2.2.1 计算流程

1）构建历史驾驶数据库

示例数据库：

场景编号	环境条件	左脑方案（L）	右脑方案（R）	结果（事故率）	最优方案
1	高速公路变道	变道超车（L）	保持车道（R）	L 方案事故率 3%	R
2	城市交叉口	加速通过（L）	减速等待（R）	L 方案事故率 15%	R
3	普通道路行驶	60km/h 行驶（L）	55km/h 行驶（R）	L 方案事故率 0.5%	L

2）定义统计判据

对于每个当前驾驶场景，统计以下 a), b), c), d) 指标：

a) 事故率对比

计算历史数据中左脑（L）和右脑（R）方案的事故率：

$$P_{accident}^{L} = \frac{左脑方案导致的事故次数}{左脑方案执行次数}$$

$$P_{accident}^{R} = \frac{右脑方案导致的事故次数}{右脑方案执行次数}$$

若 $P_{accident}^{L} > P_{accident}^{R}$，则右脑方案优先

若$P_{accident}^{L} \approx P_{accident}^{R}$，则采用其它仲裁标准。

b) 急刹车 & 急加速比例

计算左脑与右脑方案的急刹/急加速比例：

$$P_{hard\ brake}^{L} = \frac{左脑方案的急刹次数}{左脑方案执行次数}$$

$$P_{hard\ brake}^{R} = \frac{右脑方案的急刹次数}{右脑方案执行次数}$$

若 $P_{hard\ brake}^{L} > P_{hard\ brake}^{R}$，说明左脑方案较激进，右脑方案更稳定。

c) 平均车速差异

$\Delta V = V_L - V_R$

- 若 ΔV>10km/h 且事故率无明显变化，则左脑方案优先（提升效率）。
- 若 ΔV>10km/h 且事故率显著增加，则右脑方案优先（提升安全性）。

d）通过时间对比

$$T_{pass}^{L}，T_{pass}^{R}$$

若 $T_{pass}^{L} < T_{pass}^{R}$ 且风险无明显增加，则左脑方案优先（提高通行效率）。
若T_{pass}^{L}显著降低但事故率升高，则右脑方案优先（减少安全风险）

3）最终仲裁规则

结合所有统计指标，计算仲裁得分：

$$S = w_1 P_{\text{accident}} + w_2 P_{\text{hard brake}} + w_3 \Delta V + w_4 T_{\text{pass}}$$

其中：w_1, w_2, w_3, w_4为不同指标的权重（可通过机器学习优化）

设定阈值：

- $S_L < S_R$ → 选择左脑方案
- $S_L > S_R$ → 选择右脑方案
- $S_L \approx S_R$ → 结合其它决策机制（如知识图谱）

2.2.2 案例 1

黄灯通过 vs. 停止等待

左脑方案：加速通过黄灯（节省 5s）。
右脑方案：减速等待绿灯。

历史数据：

- 加速通过事故率：15%（$P_{accident}^{L} = 0.15$）
- 停止等待事故率：2%（$P_{accident}^{R} = 0.02$）
- 急刹比例：加速方案 30% > 停止等待 5%。
- 通行时间：加速方案 5s，停止等待 25s。

仲裁结果：

$S_L = 2.165$

$S_R = -1.975$

$S_L > S_R$，右脑方案更安全，停止等待黄灯。

2.3 规则优先级判据

右脑基于**法规和安全规则**进行判定：

- 若左脑的决策违反交通法规（如闯红灯、危险超车），则右脑直接否决。
- 若左脑决策符合规则，右脑的策略仅作为优化建议。

2.4 机器学习反馈判据

通过强化学习+监督学习：

- 记录左脑和右脑在历史场景中的冲突情况，并通过反馈调整决策策略。
- 允许右脑在部分场景下“放手”让左脑试错，若实际风险较低，则降低干预频率。

3 成长与提高

双脑系统还可以通过**自适应仲裁系统**、**对抗训练（Adversarial Training）**、和 **AlphaGo Zero 方法**进一步提高性能。

自适应仲裁系统

利用自适应神经网络调整左脑-右脑的权重分配，使其在**不同驾驶情境**下自动优化协调方式。

对抗训练（Adversarial Training）

让左脑和右脑在仿真环境（如 NVIDIA Omniverse）中进行**自我对抗学习**，在冲突博弈中找到**最优合作策略**。

AlphaGo Zero 方法

AlphaGo Zero 采用了双网络结构：Policy Network（策略网络）负责选择最佳行动，即在当前局面下做出决策；Value Network（价值网络）评估当前局势的胜率，即判断该决策是否有利。左右脑架构借鉴 AlphaGo Zero 的协调机制，以优化左右脑的合作与冲突处理。

3.1 自适应仲裁系统

上述所有根据环境动态调整权重的方法都属于自适应仲裁（Adaptive Arbitration System, AAS），参考文献另见[20][21][22]

需要强调的技术方法有：

3.1.1 基于强化学习的动态优化

使用多智能体强化学习（MARL, Multi-Agent Reinforcement Learning），让左右脑在模拟环境中相互博弈，不断调整权重。

训练目标：

- 最大化安全性（P_{risk} 最小）
- 保持驾驶效率（T_{pass} 最短）
- 避免过度保守驾驶

3.1.2 多模态数据融合

结合视觉（Camera）+ 激光雷达（LiDAR）+ V2X（车路协同） 数据，提供更准确的风险评估。通过历史数据分析，动态调整不同权重因子的影响力。

[]][

- 若右脑决策频繁触发但未出现实际危险，降低右脑干预强度
- 若左脑方案频繁导致高风险操作，则右脑接管权重增加

3.2 对抗训练（Adversarial Training for Dual-Brain Architecture）

为了优化左右脑之间的的协作，可将对抗训练（Adversarial Training）用于：

- 提升左脑的安全性（减少激进行为）
- 增强右脑的适应性（减少过度保守）
- 优化整体驾驶策略，找到安全与效率的最佳平衡点

主要方法参考[23][24]。

3.2.1. 对抗训练的核心机制

对抗训练（Adversarial Training）是一种博弈机制，其中：

- 左脑（执行者，Actor） 负责优化驾驶效率，追求更快、更流畅的驾驶策略
- 右脑（监督者，Critic） 负责评估左脑的方案，制造挑战，确保安全性

二者形成对抗博弈（Adversarial Game），类似于生成对抗网络（GAN, Generative Adversarial Network），最终达到动态均衡（Nash Equilibrium）。

3.2.2. 具体对抗训练方法

对抗训练可以通过以下三种方式进行优化：

3.2.2.1 基于强化学习（Reinforcement Learning, RL）的对抗训练

训练目的是：

- 让左脑使用深度强化学习（Deep Reinforcement Learning, DRL），优化驾驶策略。
- 让右脑使用价值函数（Value Function），评估左脑决策的风险，并生成对抗环境。

训练步骤

1）左脑训练（强化学习）

- 目标：最大化驾驶效率（如最短通行时间、最少刹车）
- 方法：使用深度 Q 网络（DQN）或策略梯度方法（PPO）
- 奖励函数

$$R_{left} = \alpha(-T_{pass}) + \beta(-Jerk) - \gamma P_{risk}$$

其中：
T_{pass}：通行时间（越短越好）
$Jerk$：加速度变化率（减少突兀操作）
P_{risk}：风险评分（减少危险行为）

2）右脑训练（风险对抗评估）

- 目标：最大化安全性，最小化事故风险
- 方法：使用基于知识图谱 KG 的推理模型 + 强化学习监督（Critic）
- 生成对抗性环境：
    - 模拟突发状况（如前车急刹、侧方车辆突然并线）。
    - 构造边界挑战（如在变道可行性边界内调整对抗力度）。
- 奖励函数：

$$R_{right} = -\delta P_{risk} - \eta T_{pass}$$

其中：
P_{risk}：事故风险
T_{pass}：通行时间（避免极端保守）

3）对抗训练

- 左脑尝试找到最优驾驶策略，右脑不断制造更难的挑战
- 右脑基于对抗性环境生成（Adversarial Environment Generation），不断优化挑战机制

示例

- 左脑方案：检测到黄灯，决定加速通过。
- 右脑对抗：右脑模拟行人突然穿越、侧方来车逼近，强制左脑重新评估决策。
- 优化结果：左脑学会在黄灯场景中适度减速，避免高风险通过。

3.2.2.2 基于对抗性数据增强（Adversarial Data Augmentation）

目的是让右脑制造对抗性驾驶数据，用于训练左脑，使其在更具挑战性的场景下学习更稳健的驾驶策略。

1）构造对抗性样本

基于知识图谱推理，筛选高风险驾驶案例：

- 事故多发路段
- 复杂交通场景（如交叉路口、隧道出口）
- 突发事件（如前方事故、行人违规穿越）

生成对抗性数据，如：

- 模拟强烈阳光干扰传感器
- 增加道路湿滑、低能见度等特殊天气条件
- 制造拥堵、急刹车等场景

2）左脑适应训练

- 将右脑生成的对抗性数据**加入左脑训练数据集中**，使左脑逐步适应极端环境。
- 使用**对抗性数据回放（Adversarial Replay Buffer）**，让左脑在最具挑战性的驾驶场景中进行自我修正。

示例

- 左脑方案：在高速公路上变道加速
- 右脑对抗数据：模拟对向车辆突然变道进入盲区
- 优化结果：左脑学会在变道前加大盲区检测范围，提高安全性

3.2.2.3 基于博弈论的动态权重优化

让左右脑**在驾驶策略与安全性之间博弈**，通过**动态权重优化**，调整决策权重。

1）建立动态仲裁系统

计算左脑（效率驱动）与右脑（安全驱动）的得分：

$$S_{left} = f(T_{pass}, Jerk)$$

$$S_{right} = f(T_{pass}, RuleCompliance)$$

使用非零和博弈（Non-Zero Sum Game）优化权重：

$$W_{left} = \frac{1}{1 + e^{-(S_{left} - S_{right})}}$$

$$W_{right} = 1 - W_{left}$$

2）动态调整决策权重

- 当左脑方案风险评分较低（安全驾驶）：提升左脑权重，减少右脑干预
- 当左脑策略风险较高（可能导致事故）：提升右脑权重，减少左脑自由度

示例

- 左脑策略：超车变道，提高通行效率。
- 右脑策略：检测到侧方车辆接近，建议保持当前车道。
- 博弈计算：
    - 左脑得分 $S_{left} = 0.8$（通行效率高）
    - 右脑得分 $S_{right} = 0.7$（风险较低）
    - 计算权重： $W_{left} = 0.58$, $W_{right} = 0.42$
- 最终仲裁：左脑方案优先，但右脑设定额外监测。

注：对抗学习与博弈平衡的区别见“**附录 I 对抗 RL 与博弈平衡的应用区别**”

3.3 AlphaGo Zero 方法

在围棋博弈中：策略网络（Policy Network） 通过强化学习预测当前局面的最优落子策略（概率分布）。价值网络（Value Network） 评估当前局面的优劣，并预测胜率。蒙特卡洛树搜索（MCTS, Monte Carlo Tree Search）：结合策略网络提供的行动建议和价值网络提供的局势评估，进行模拟搜索，选出综合最优解。

如果策略网络与价值网络冲突（即策略网络推荐 A，但价值网络评估 A 胜率较低），MCTS 通过多次模拟和统计反馈调整最终决策。

3.3.1 方法移植借鉴

原理：

左脑充当 AlphaGo 里 Policy Network 的角色, 强化学习执行网络，负责驾驶执行，优化速度、路径、超车等操作；右脑知识图谱推理网络充当 Value Network, 评估驾驶方案的安全性和稳定性，基于规则推理约束高风险操作。MCTS（蒙特卡洛树搜索）或动态仲裁系统负责协调决策冲突。

1）冲突检测

若左脑提议的方案 P_{left} 与右脑的安全评估 P_{right} 发生冲突：

- 左脑：期望优化通行效率（如高速变道）。
- 右脑：评估此举可能导致碰撞（如盲区车辆靠近）。

2）冲突协调方法

借鉴 AlphaGo Zero，采用“MCTS + 强化学习”进行协调。

（1）对指定的端到端左脑，搜索最优驾驶方案（Policy Tree）：

- 在仿真环境中，模拟左脑方案的执行情况。
- 计算不同驾驶操作的成功率（如变道成功率、刹车安全性）。

（2）引入右脑的风险评估（Value Function）：

- 使用风险快照、安全围栏、防御驾驶三个单元评估该方案的安全性
- 计算每种驾驶方案的安全得分 S_{risk}：

$$S_{risk} = w_1 P_{风险快照} + w_2 P_{安全围栏} + w_2 P_{防御驾驶}$$

- 若 S_{risk} 过高，则降低该方案的优先级

（3）最终决策：

- 选择综合评分最高的驾驶方案（结合左脑效率 & 右脑安全评估）

3.3.2 示例

高速公路变道

1）左脑（Policy Network）提议：变道以提高通行效率
2）右脑（Value Network）评估：检测盲区有快速接近的车辆，风险评分 $S_{risk} = 0.8$（高风险）
3）MCTS 决策：
- 运行 100 次 Monte Carlo 模拟
- 发现变道方案在 30% 情况下导致急刹或碰撞，最终否决变道

需要注意的是，在每次蒙特卡洛树搜索（MCTS）时，都要分别计算策略和价值[25]。双网结构也存在一些问题，主要是训练效率低，后来又改成了网结构[26]。如果在时间中计算开销过大无法承受，则应改用上述**动态权重调整机制**。

另外，自动驾驶与围棋应用的环境不一样，自动驾驶的复杂性远高于围棋，具体表现在：

- 围棋是离散决策问题，而自动驾驶涉及连续时空、多智能体互动，场景复杂度更高。
- 自动驾驶需要实时性，而 AlphaGo Zero 可计算大量模拟
- 围棋是零和博弈（敌对），而自动驾驶是非零和博弈（需合作，如避让行人）。

所以自动驾驶挑战更大。MCTS 方法能否应用到双脑协调，需要进一步实践尝试才能有答案。

3.4 预期改进

3.4.1 引入 Transformer 结构

- 结合 Vision Transformer（ViT）+ Transformer-based RL，提升左脑对复杂交通环境的理解能力。

3.4.2 采用 Graph Neural Network（GNN）优化知识图谱推理

- 让右脑更精准地建模道路场景，提高对事故风险的预测能力。

3.4.3 云端仿真训练（Sim2Real）

- 通过大规模仿真训练（如 NVIDIA Omniverse），提升左右脑在不同场景下的适应能力。

4 团队建设

假设：在已有 K2D 团队的基础上。

团队类别	岗位	人数（最小）
强化学习 & 博弈论	强化学习工程师（RL Engineer）	2
	多智能体博弈研究员（Multi-Agent Game Theory Researcher）	1
感知 & 预测	自动驾驶感知算法工程师（Perception Engineer）	2
对抗训练 & 安全	强化学习安全工程师（Safe RL Engineer）	1
	对抗训练专家（Adversarial Training Specialist）	1
仿真 & 系统优化	大规模仿真测试工程师（Simulation Engineer）	2
	系统架构师（AI System Architect）	1
总计		**10 人**

人员素质要求：

1）强化学习工程师（Reinforcement Learning Engineer）

核心任务：开发左脑的端到端强化学习模型，优化驾驶策略，使其兼顾安全性与效率。

技术要求：

- 强化学习（Deep Q-Learning, PPO, SAC, TD3）
- 强化学习平台（RLlib, Stable-Baselines3）
- 运动控制（Model Predictive Control, MPC）

学历/经验：

- 硕士及以上，计算机科学、机器学习、自动驾驶相关专业。
- 3 年以上强化学习研究或工业应用经验，熟悉自动驾驶决策优先。

2）多智能体博弈研究员（Multi-Agent Game Theory Researcher）

核心任务：研究左右脑的博弈机制，设计最优仲裁算法。

技术要求：

- 进化博弈（Evolutionary Game Theory）
- 非零和博弈（Non-Zero-Sum Games）
- MARL（MADDPG, QMIX, Independent Q-Learning）

学历/经验：

- 博士或硕士（有顶级期刊论文者优先）
- 3-5 年博弈论应用经验，有自动驾驶/机器人应用背景优先

3）对抗训练专家（Adversarial Training Specialist）

核心任务：开发对抗性训练方法，增强左右脑的适应能力。

技术要求：

- 对抗性训练（Adversarial Training, GANs）
- 自适应测试（Adaptive Testing）
- 逆向强化学习（Inverse RL）

学历/经验：

- 硕士及以上，AI 研究员背景优先。
- 3 年以上对抗训练研究或工业应用经验。

4）大规模仿真测试工程师（Simulation Engineer）

核心任务：建立自动驾驶仿真环境，测试左右脑的协调效果。

技术要求：

- 车辆仿真（CARLA, SUMO, LGSVL）
- 强化学习环境设计（Gym, Isaac Sim）
- 自动驾驶系统仿真（Apollo, Autoware）

学历/经验：

- 硕士及以上，计算机仿真/自动驾驶相关领域。
- 3 年以上仿真开发经验。

5）系统架构师（AI System Architect）

核心任务：优化 AI 模型推理性能，设计高效计算框架。

技术要求：

- 分布式计算（Kubernetes, Ray）
- 高效推理（TensorRT, ONNX Runtime）
- 云端训练 & 部署（AWS/GCP/Azure）

学历/经验：

- 硕士及以上，计算机科学/AI 工程。
- 5 年以上 AI 架构经验，熟悉自动驾驶系统架构者优先。

5 总结

- 单独依靠上述协调-仲裁方法中的一个，也许不容易解决左右脑的协同运作问题，可能需要多种方法并用。至于最佳组合方式需要由试验调试结果来确定。价值协调系统也许是安全大脑里最难的一部分。
- 最大的挑战是最终设计方案要足够简单又足够可信。
- 团队建设是关键。

参考文献

[2] **CHARACTERIZATION AND MITIGATION OF INSUFFICIENCIES IN AUTOMATED DRIVING SYSTEMS**

[3] 侯涛、丁伟平、黄佳霜、鞠恒荣,《DE-NNs: 基于动态证据神经网络的脑网络分析算法》

[4] Steven L. Tu，Dynamic Neural Networks: A Survey

[5] Anil Ranjitbhai Patel, Peter Liggesmeyer, LADRI: LeArning-based Dynamic Risk Indicator in Automated Driving System》

[6] Alessandro Zanardi, Andrea Censi, Margherita Atzei, Luigi Di Lillo, Emilio Frazzoli , A Counterfactual Safety Margin Perspective on the Scoring of Autonomous Vehicles' Riskiness

[7] Chen, Y., Zhang, X., & Huang, J., Assessing Driving Behavior Stability Using Motion Data and Deep Learning

[8] Andersson, L., & Yu, W, A Risk-Based Model for Evaluating Autonomous Vehicle Behavior Consistency

[9] Kim, H., & Park, J., Motion Smoothness Evaluation in Autonomous Driving: A Multi-Metric Approach

[10] Y. Shi, Y. Chu, Y. Liu, et al., "Multi-Agent Reinforcement Learning for Autonomous Driving: A Survey"

[11] M. Oroojlooyjadid, L. S. Snyder , A Comprehensive Survey on Multi-Agent Reinforcement Learning

[12] A. Brito, P. Alvito, P. Santana, Multi-Agent Deep Reinforcement Learning for Autonomous Driving

[13] T. Kim, R. Langari, Game Theory Based Autonomous Vehicles Operation

[14] R. Rabinowitz, J. F. Pfeiffer, D. P. Reichert, et al., Modeling Theory of Mind in Multi-Agent Games Using Adaptive Learning"

[15] L. Wu, H. Zhang, X. Li, A Knowledge Graph Approach for Risk Assessment in Autonomous Driving

[16] J. Smith, T. Miller,Risk-Aware Decision Making in Autonomous Vehicles Using Knowledge Graphs

[17] A. Gupta, P. Kumar, A Hybrid Approach for Traffic Risk Prediction: Combining Knowledge Graphs and Machine Learning

[18] M. Anderson, K. Kim, Statistical Risk Assessment in Autonomous Driving: A Data-Driven Approach

[19] H. Zhao, P. Kumar, Comparative Study of Autonomous Driving Strategies Using Statistical Decision Theory"

[20] J. Lee, M. Patel, Adaptive Arbitration Strategies in Autonomous Driving Using Multi-Agent Reinforcement Learning

[21] T. Kim, R. Langari, Risk-Aware Decision Making in Autonomous Vehicles: A Knowledge Graph Approach

[22] A. Gupta, K. Tanaka, Learning-Based Arbitration Between Model-Based and Model-Free Controllers for Safe Autonomous Driving

[23] J. Wang, R. Patel, Adversarial Reinforcement Learning for Safe and Efficient Autonomous Driving

[24] H. Zhang, P. Kumar, Game-Theoretic Approaches to Autonomous Driving: A Survey

[25] Silver et al., Mastering the game of Go with deep neural networks and tree search

[26] Silver et al., Mastering the Game of Go without Human Knowledge

---

附录 I 对抗 RL 与博弈平衡的应用区别

双脑对抗训练（Adversarial Training）与博弈平衡（Game Equilibrium）虽然都涉及对抗和博弈的概念，但在目标、方法和应用上存在显著区别，不是同一回事，但可以相互结合。

1. 双脑对抗训练（Dual-Brain Adversarial Training）

核心概念：
双脑对抗训练是指的是在强化学习（RL）或对抗训练（Adversarial Training）框架下，引入两个互相竞争或合作的智能体（双脑），通过相互对抗或协作进行学习优化。这种方式能够提升模型的泛化能力、鲁棒性和安全性。

特点：

- **对抗性：**可将左脑（主智能体）作为主导，右脑作为干扰者或对手，通过持续进化提高主智能体的对抗能力
- **强化学习（RL）结合：**使用强化学习框架，通过策略梯度、Q-learning 或 Actor-Critic 机制让两个智能体在博弈过程中学习最优策略
- **动态优化：**不断调整策略，使左脑（主智能体）对干扰因素（如环境变化或右脑策略）具有更强的适应性，类似于 GAN（生成对抗网络）中的 Generator 与 Discriminator。

应用：

- **AV 安全测试**（对抗式自动驾驶模拟环境）
- **安全智能体训练**（用于防御性驾驶或攻击检测）
- **机器人对抗训练**（如多智能体协作或竞争）

2. 博弈平衡（Game Equilibrium）

核心概念：

博弈平衡是博弈论（Game Theory）中的一个概念，指在博弈过程中，所有参与者都达到了一个策略稳定的状态，即每个参与者都没有动力单方面改变自己的策略。

常见的博弈平衡概念：

- **纳什均衡（Nash Equilibrium）：**在多人博弈中，如果每个玩家的策略都是对其他玩家策略的最优响应，则达成了纳什均衡。在这种情况下，单个玩家无论如何改变策略都无法获得更好的结果
- **极小极大（Minimax）均衡：**常见于零和博弈，如棋类游戏（围棋、象棋）。一个玩家的损失等于另一个玩家的收益，因此策略是基于最坏情况进行优化
- **演化稳定策略（ESS, Evolutionarily Stable Strategy）：**在进化博弈论中，若一个群体采用某种策略，个别偏离该策略的个体无法获得优势，则该策略是演化稳定的

特点：

- **关注策略稳定性：**不同于双脑对抗训练的动态演化，博弈平衡主要研究的是一个稳定的策略组合，即博弈中的均衡点
- **数学分析驱动：**博弈论的均衡概念主要基于数学模型，而强化学习和对抗训练通常基于数据驱动的学习方法
- **强调长期策略：**博弈平衡关注的是长期稳定性，而对抗训练更注重持续优化过程

应用：

- **自动驾驶策略优化**（如交叉路口多车博弈）
- **经济市场均衡建模**（如竞争对手定价策略）
- **军事和网络安全**（如攻防博弈策略）

3. 二者关系与区别

对比项	双脑对抗训练 （Adversarial Training）	博弈平衡 （Game Equilibrium）
目标	通过对抗训练提升系统的适应性与鲁棒性	寻找稳定的策略组合，不可被单方面优化
是否动态	是，智能体在训练过程中不断优化策略	否，均衡状态下策略不变
方法	强化学习、对抗生成、演化优化	纯数学建模、纳什均衡、极小极大优化

是否涉及博弈	涉及，但主要用于训练	核心是博弈均衡
自动驾驶应用	提升对抗性和安全性	优化驾驶策略
主要应用	机器人对抗训练、自动驾驶安全、对抗性 AI	经济、军事、安全策略、自动驾驶交互

4. 结合应用

安全大脑可以结合使用上述这两种方法。

1. **使用双脑对抗训练来提升安全性：**
   - 训练一个进攻性智能体（左脑）模拟极端驾驶情况（如故意变道、急刹车）
   - 另一个智能体（右脑）学习如何规避危险，提高防御性驾驶能力
2. **使用博弈平衡来优化自动驾驶策略：**
   - 在复杂交叉路口、多车交互等情境下，寻找最优驾驶策略，使所有车辆的策略达到稳定状态，提高行车安全性和通行效率

为了构建更鲁棒的自动驾驶安全系统，可以：

- 训练阶段使用双脑对抗训练，让 AI 在恶劣环境中学会生存；
- 部署阶段使用博弈平衡，确保 AI 与其他车辆的交互策略是稳定、最优的