

AI 桌面冰球机器人

摘要

桌面冰球作为一种流行的桌面游戏，是指双方利用“球拍”在较为光滑的桌面上通过击打冰球，从而进球得分的游戏，考验参与者的专注力和博弈思维。

本产品主要由负责通过摄像数据采集的计算机视觉系统、负责击球的运动控制系统和负责提醒和记录胜负的人机交互系统三大系统构成，并且辅以具有专用神经网络处理单元（NPU）的边缘 AI 开发平台 HiKey970，从而实现由人工智能操作的，能够与真人进行冰球对战的智能冰球机器人。

在人工智能的推行已然成为大势所趋，是科技发展的浪潮的时代，人工智能逐渐成为娱乐产业不可或缺的一部分。并且在“单人经济”出现的情况下，可以支持单人与机器对战而获取双人博弈快乐的娱乐项目，必然会成为娱乐产业的主流。

目录

摘要.....	1
一. 设计需求分析.....	3
1.1 <u>人工智能浪潮需求</u>	3
1.2 <u>娱乐产业智能化趋势需求</u>	3
1.3 <u>“单人经济”趋势需求</u>	3
二. 设计方案.....	5
三. 系统实现.....	6
3.1 主要结构.....	6
3.2 计算机视觉系统.....	6
3.3 运动控制系统.....	9
3.4 人机交互系统.....	11
3.5 内部硬件.....	13
四. 创新亮点.....	15
4.1 冰球追踪模型.....	15
4.2 Q-learning.....	16
五. 技术难度.....	18
5.1 技术难点分析及解决方案	
六. 成品展示 (见附件)	

一、设计需求分析

1.1 人工智能浪潮需求

自 2016 年以来，由于云计算、大数据等技术的突破，人工智能的发展正处于 1956 年以来的第三次高潮。人工智能是一门关联性极其广泛，适用于众多学科的重要科学问题。《国务院关于印发新一代人工智能发展规划的通知》中提出：面对新形势新需求，必须主动求变应变，牢牢把握人工智能发展的重大历史机遇，紧扣发展、研判大势、主动谋划、把握方向、抢占先机，引领世界人工智能发展新潮流，服务经济社会发展和支撑国家安全，带动国家竞争力整体跃升和跨越式发展。从此我们可以看出，人工智能的推行已然成为大势所趋，是科技发展的浪潮。

1.2 娱乐产业智能化趋势需求

2016 年 3 月 15 日，谷歌旗下的 DeepMind 公司开发的围棋人工智能程序 AlphaGo，以 4:1 的成绩战胜韩国棋士李世石。这种在博弈领域的“人机对战”，一来可以通过人工智能提供更强劲的对手，提高博弈游戏的博弈性，同时也可以为一般玩家提供一个可以博弈的同水准对手，增加游戏的乐趣。并且，人工智能的推行使得毫无感情的机器获得了独特的个性，这就增加了机器的独特魅力对于人们的吸引力。因此，在娱乐产业中，人工智能必然会成为不可或缺的一部分。

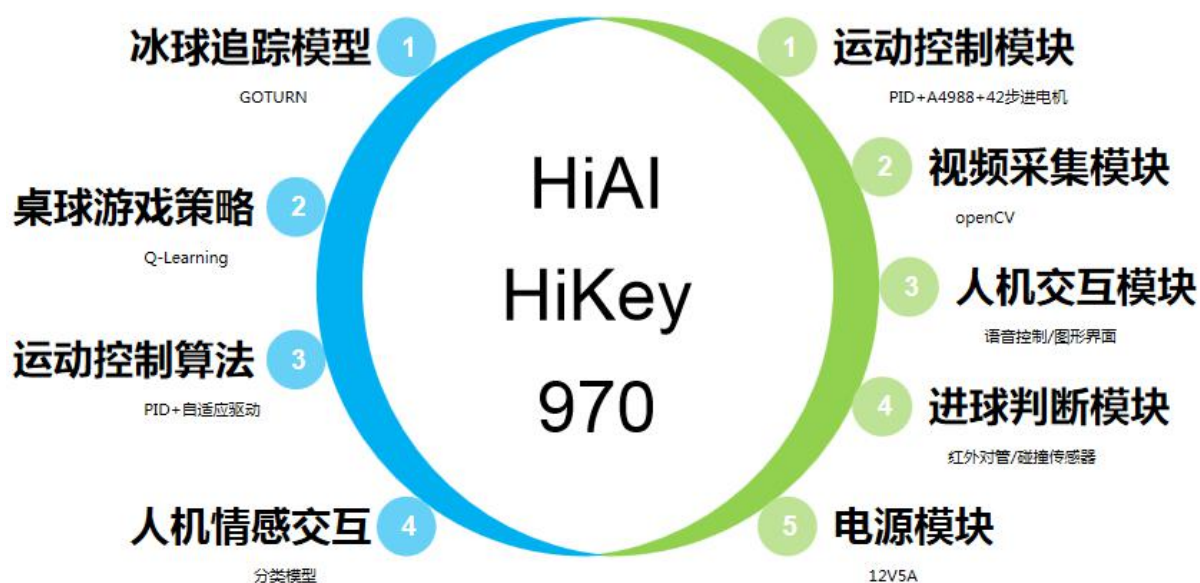
1.3 “单人经济”趋势需求

据民政部的数据显示：2018 年中国单身成年人口已经超过 2 亿，独居成年人口超过 7700 万。这使得一个叫做“单人经济”的概念出现——其内涵与孤独消费相似。举个例子，据和讯新闻报道：作为孤独消费代表的宠物经济，其所带动的宠物主粮、零食、保健品、用品以及医疗产品市场都包含了巨大的消费潜力。狗民网发布的《2018 年中国宠物行业白皮书》显示，2018 年中国宠物（猫

狗)消费市场规模达到 1708 亿,同比增长 27%。白皮书还显示,养宠省市的繁荣程度也和独居人口分布最多的城市成正比。而以冰球机器人为首的,可以支持单人与机器对战而获取双人博弈快乐的娱乐项目,在这个“单人经济”盛行的时代,必然会成为娱乐产业的主流。

二、设计方案

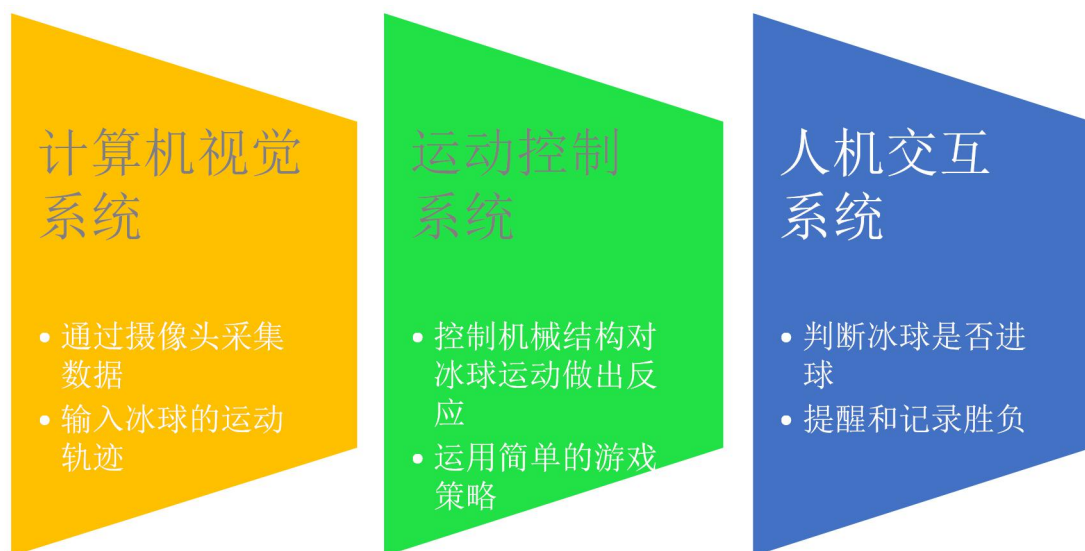
桌面冰球机器人主要由以下几个部分组成，负责完成冰球击打动作的运动控制模块，负责图像采集的计算机视觉模块，负责提醒和记录胜负的人机交互模块，还有控制核心，其主要负责感知、决策和控制功能，包括视频流下冰球的追踪、运动部分的自适应控制和基于 Q-learning 的游戏策略。



三、系统实现

3.1 主要结构

本系统主要由负责通过摄像数据采集的计算机视觉系统、负责击球的运动控制系统和负责提醒和记录胜负的人机交互系统三大系统构成，通过摄像数据采集系统对于冰球运动轨迹的相关数据采集，再使用相关算法加以计算得出冰球较为精准的运动轨迹甚至是折射路线，实现冰球的追踪；在得到分析后的数据后，击球系统迅速反应，以一定预设的方式运动，从而做到简单的击球回球等动作，在此基础上加入一定的游戏策略，提高该系统的可玩性与竞技性。在冰球游戏进行的同时通过人机交互系统对与游戏的结果进行统计，使整个系统更加完善。在此基础上我们将冰球追踪系统通过深度学习或者强化学习实现，使其在算法的优化和应用上有一定的创新。



3.2 计算机视觉系统

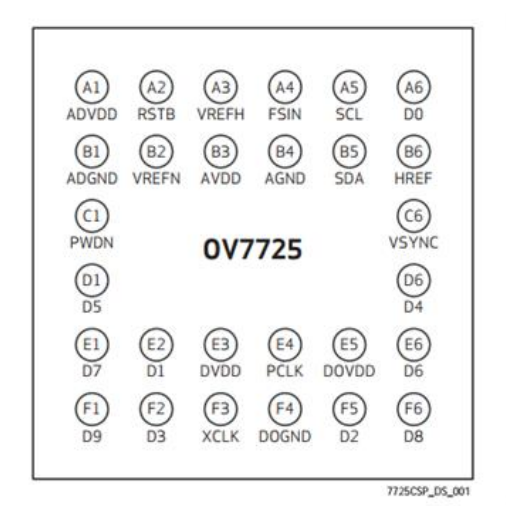
简单的我们可以利用摄像头所获取的两帧图片上冰球的坐标差计算出速度

并预测轨迹，再考虑反弹这些情况，预测出冰球的最终坐标，但这种方法较为简陋，可能导致最终误差较大，且冰球折射运动轨迹难以准确的判断会对整个系统造成极大的影响，所以对于冰球的实时追踪便显得很有必要的。

采用 OpenMV 机器视觉模块的方案。OpenMV 是一个开源，低成本，功能强大的机器视觉模块。以 STM32F427CPU 为核心，集成了 OV7725 摄像头芯片，在小巧的硬件模块上，用 C 语言高效地实现了核心机器视觉算法，提供 Python 编程接口，由于开放源代码，机器视觉算法有更自由的空间进行改进和增加；



OpenMV 上的机器视觉算法包括寻找色块、人脸检测、眼球跟踪、边缘检测、标志跟踪等。可以用来实现非法入侵检测、产品的残次品筛选、跟踪固定的标记物等。



OV7725

一方面图像处理的策略可以是：通过 **hsv** 颜色阈值在画面中寻找冰球和打击器坐标，通过帧差判断出冰球的运动速度和方向，当冰球朝着机器人运动时根据冰球的轨迹预测撞击点坐标（其中包括有反弹和无反弹的情况），将预测的终点坐标发送给下位机，下位机运动到终点视具体情况进行防御或主动进攻。

从创新的角度来说：我们可以利用深度学习来实现冰球追踪，机器学习技术

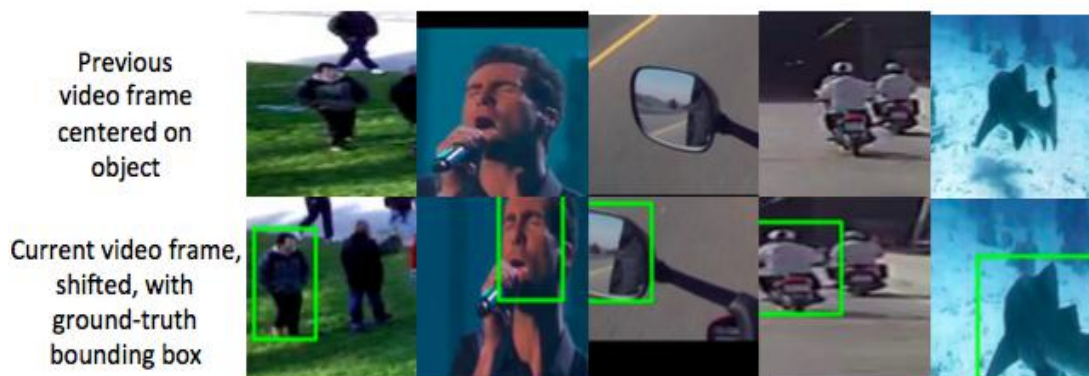
由于能够利用大量的训练数据来提高性能，因此经常用于计算机视觉。不幸的是，大多数通用对象跟踪器仍是从头开始在线培训的，无法从离线培训中容易获得的大量视频中受益。有人提出了一种使用神经网络跟踪通用对象的方法，该方法允许它们通过在带标签的视频上进行训练来提高性能。先前使用神经网络进行跟踪的尝试运行起来非常慢，并且对于实时应用而言不切实际。相比之下，我们的跟踪器使用简单的前馈网络，而无需进行在线培训，从而允许我们的跟踪器在测试期间以 100 fps 的速度运行。测试的跟踪器会从带有标签的视频以及大量图像中进行训练，这有助于防止过度拟合。跟踪器学习通用的对象运动，可用于跟踪未出现在训练集中的新颖对象。



基于神经网络的 GOTURN: Generic Object Tracking Using Regression Networks 是发表在 ECCV 2016 的一篇有关 Tracking 的文章，达到了 Tracking 中效果上的 state-of-the-art，尤其在检测速度上达到了 100FPS（第一个达到 100FPS 的深度学习方法），Code 用 Caffe 编写。该算法对遮挡敏感，但是对视角、形变、光照变化具有鲁棒性。由于冰球游戏场景很少有遮挡情况，故将这种理论模型应用于桌面冰球的球体追踪中。

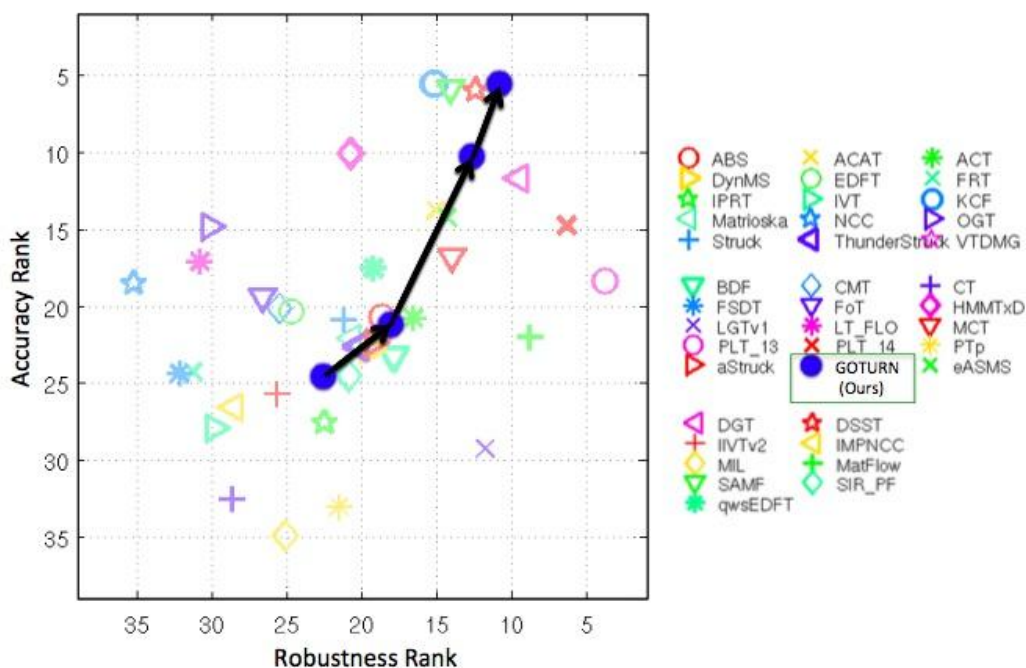
整个文章的关键点就是 Regression，回归的是什么？当然是 bounding-box 的坐标，那么回归的输入变量就是 current frame，输出为 bounding-box 的坐标。当然前提是知道 previous frame 中 object 的坐标在中心位置。那么这个 Regression Network 学习到的就是：object 在视频中前后帧的 motion 到 object 坐标的变化！知道了 object 在上一帧的中心，找到 object 在当前帧的位置。

利用视频和静态图片训练，优化 predicted bounding-box 和 ground truth bounding-box 之间的 L1 范数损失。训练集中 video 的某些 object 带有 bounding-box，按照前文所述将 search region 给 crop 出来，将两帧和当前帧 bounding-box 的坐标输入网络来优化。在训练的时候也可以在当前帧进行人为的 random crop，以增加 tracker 的鲁棒性、增广数据集。



因为冰球在视频中相邻两帧的运动具有相对平稳的性质，形变、位移等变化不会太大，所以可以对 **bounding-box** 的位移进行一个建模，实现对冰球的实时追踪。

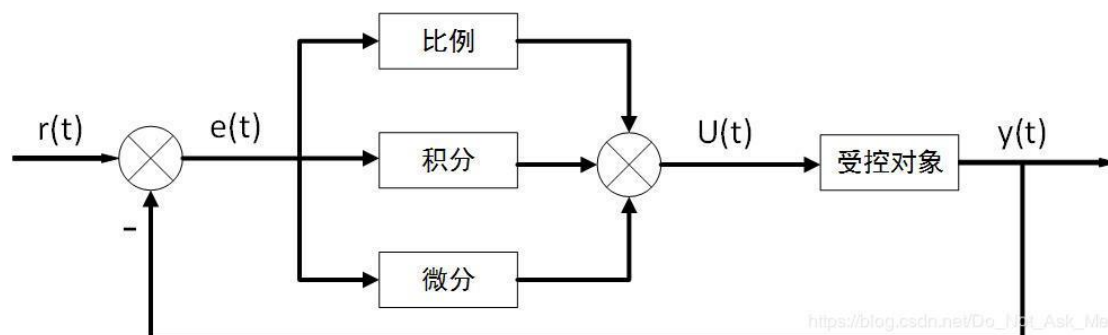
两个跟踪指标：**Accuracy** 和 **Robustness**，范围都是[0,1]，当然每个视频都有一些属性：遮挡、光照变化、运动变化、尺度变化、摄像机运动，**tracker** 在这几个指标上分别测量了 **Accuracy** 和 **Robustness** 并求了平均，以此作为测度指标。准确率和速度都是 **state-of-the-art**，效果如下



3.3 运动控制系统

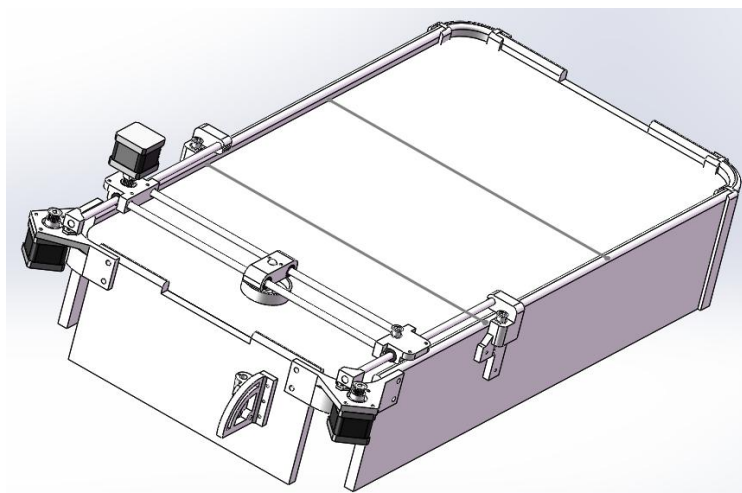
在对于击球运动控制系统的选择上我们采用了 **PID** 控制，**PID** 控制应该算是应用非常广泛的控制算法了。小到控制一个元件的温度，大到控制无人机的飞行姿态和飞行速度等等，都可以使用 **PID** 控制。

常规的模拟 PID 控制系统原理图如下，该系统由模拟 PID 控制器和被控对象组成。图中 $r(t)$ 是给定值， $y(t)$ 是系统实际输出值，两者构成偏差 $e(t) = r(t) - y(t)$ 。 $e(t)$ 作为 PID 控制器的输入， $U(t)$ 为 PID 输出。



PID 控制器是一个在工业控制应用中常见的反馈回路部件。这个控制器把收集到的数据和一个参考值进行比较，然后把这个差别用于计算新的输入值，这个新的输入值的目的是可以让系统的数据达到或者保持在参考值。PID 控制器可以根据历史数据和差别的出现率来调整输入值，使系统更加准确而稳定。通过由计算机视觉系统所传输的冰球运动数据，可以较为精准的做出击球运动反应。

同时我们采用 42 混合式步进电机作为执行组件，采用 A4988 电路模块作为驱动单元。使用自适应驱动，使用步进电机加减速算法和 DFB 插补算法，向步进电机驱动器发送控制脉冲信号，控制步进电机转动，进而控制球拍在水平竖直两个方向上的快速移动，使机械结构能够有较好的性能用以对冰球运动的快速反应。

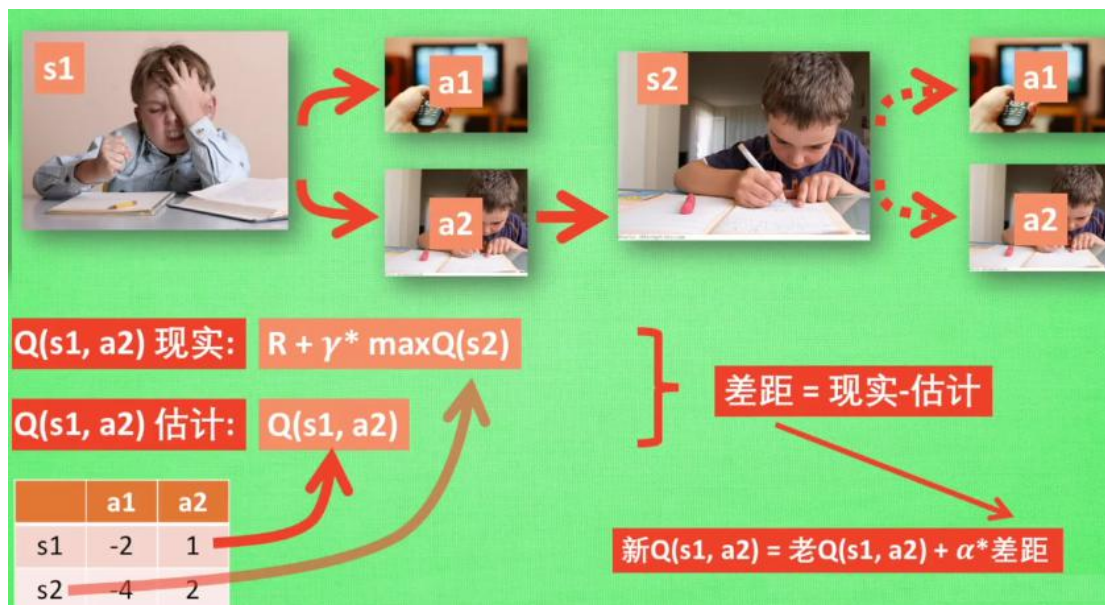


SolidWorks 模型

为实现与人对战的激烈感，我们还引入了游戏决策系统即基于 Q-learning 的游戏策略。

这是一种无模型的强化学习算法。Q-learning 则是强化学习领域的一个基本学习算法，通过与环境的交互和学习，可以得到每个状态下的评价函数 $Q(s, a)$

的值，根据这个值可以针对不同状态进行动作决策，是一种基于马尔可夫模型的策略搜索和迭代模型。我们将这种基本算法应用于冰球对战中的决策部分。



对于任何有限的马尔可夫决策过程（FMDP），Q-learning 都会发现一个最优的策略，即从当前状态开始，它将在所有和所有连续步骤中最大化总奖励的期望值。Q-learning 可以为任何给定的 FMDP 确定最佳的行动选择策略，给定无限的探索时间和部分随机的策略。“Q”表示返回用于提供强化的奖励的函数，可以说代表在给定状态下执行的动作的“质量”。

3.4 人机交互系统

人机交互（HCI）是一门研究系统与用户之间的交互关系的学问，是一个关注人和机器之间交互模式的多学科研究领域，有时也使用在制造业或过程控制系统有时更普通的术语的人机界面（HMI）被用于指在用户界面中的制造或过程控制系统。换句话说，HCI 学科关注有关人和计算机之间的界面设计和实现的所有问题。由于其性质和目标，人机交互自然而然地会涉及到计算机科学的多个学科方向（图像处理，计算机视觉，编程语言等）以及人文学科的多个方向（人体工程学，人因，认知心理学等）。人机交互研究首先关心能够改善人机互动的新型界面的设计、实现和评估。这里的改善 涉及多个方面，包括使用直观性和与接口健壮性。一个直观、自然、高效、健壮且可定制的界面可以明显缩小人类心智模式与计算机、机器或机器人完成既定任务之间的隔阂。虽然有关人机交互的研究可追溯到 1975 年，但近年来消费电子技术的进步更是打开了令人兴奋的新局

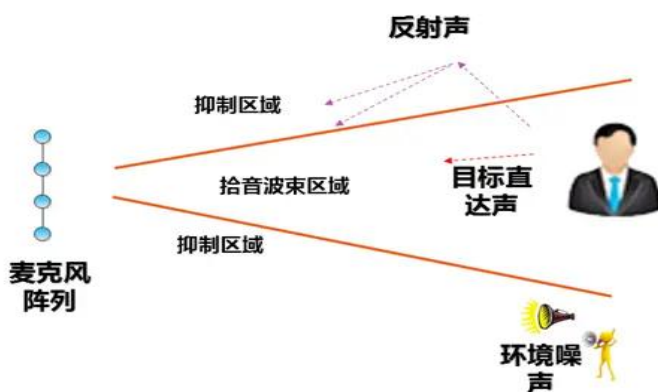
面：在设计经济实惠的自然用户界面（NUI，Natural User Interface）过程中，手势、手和身体姿态、言语、目光不过是多种自然互动模式中的几个例子。

系统可以是各种各样的机器，也可以是计算机化的系统和软件。人机交互界面通常是指用户可见的部分。用户通过人机交互界面与系统交流，并进行操作。人机交互界面的设计要包含用户对系统的理解（即心智模型），那是为了系统的可用性或者用户友好性。

除自然语言理解技术外，人机交互的另一困难就是信息维度的稀疏性。因此人机交互中情感分析应尽可能从多个维度入手，将缺失的情感带宽补上——多模态情感分析成为一个重要的研究方向。内部的多模态分析即 emoji+照片+表情包+文本长度等，外部的多模态即加入面部表情识别+提供的标签（性格，爱好，年龄等）+语音等。如竹间智能在构建基于自然语言理解技术的人机交互系统时，加入多模态情感分析技术（文字、语音、图像），打破了以往情感识别在单一文本维度上的局限性，让机器能进一步读懂、听懂、看懂人类，从而达到了更好的人机交互效果。

该系统需要多个模块共同协作来实现，如对局面和比赛结果的判断需要联动进球判断模块，通过其红外对管/碰撞感应器对进球数进行感应技术，而语音部分也是同理。

下面以语音识别为例：



语音交互属于人机交互的范畴，是人机交互发展到目前的一种最前沿的交互方式：就是用人类的自然语言给机器下指令，达成自己的目的这一过程。就目前而言语音识别仍然不那么完美，由于背景噪音的干扰，它并不能理解每一项请求。但出于创新和未来用户的需求，我们可以接受目前的不足，未来机器所做的贡献越是复杂，它们与用户之间进行高效的沟通就变得越为重要。在接口和传感器的

开发中必须始终将人的能力和局限性考虑进去。操作机器不能过于复杂或需要太多的熟悉过程。人与机器之间的顺畅沟通也要求命令与行动之间的响应时间要尽可能地短，否则用户不会认为这种交互是自然的。

该系统由局面识别和语音识别等部分组成主要作用是对冰球游戏的结果进行简单的识别与评判。通过红外对管或者膨胀传感器判断进球通过加入语音识别、语音理解等模块，实现与人交流进而达到最佳的陪护效果，并适用于各种不同的场景。该系统能够使人与整个冰球机器人之间产生简单的交互，提升整体可玩性。

3.5 内部硬件

为实现以上的各系统我们选择了 HUAWEI HiAI 能力开放平台，并使用 HiKey970。

HUAWEI HiAI 是面向智能终端的 AI 能力开放平台，基于“芯、端、云”三层开放架构，即芯片能力开放、应用能力开放、服务能力开放，构筑全面开放的智慧生态，让开发者能够快速地利用华为强大的 AI 处理能力，为用户提供更好的智慧应用体验。

三层能力开放



HiKey970 是世界上第一个具有专用神经网络处理单元（NPU）的边缘 AI 开发平台。HiKey970 提供了常用的人工智能功能 API，模型管理 API 和基本的运算符计算 API，因此开发人员可以专注于开发新的 AI 应用程序，而不必专注于性能调整。凭借它的专用 AI 处理器和丰富的影像和 I / O 集，HiKey970 使开发人员能够为智能城市中基于 AI 的智能相机构建高度集成的灵活产品。借助功能强大的 NPU 和 HiKey970 上的 HiAI SDK，开发人员可以轻松地将视频转化为有价值

的见解。凭借它出色的 AI 能力和 ROS，HiKey970 使开发人员能够构建能够自主感知，规划和决策的机器人。机器人甚至可以通过基于深度强化学习的决策来自动相互竞争。HiKey970 上功能强大的 NPU 使机器人可以显示一定水平的智能。

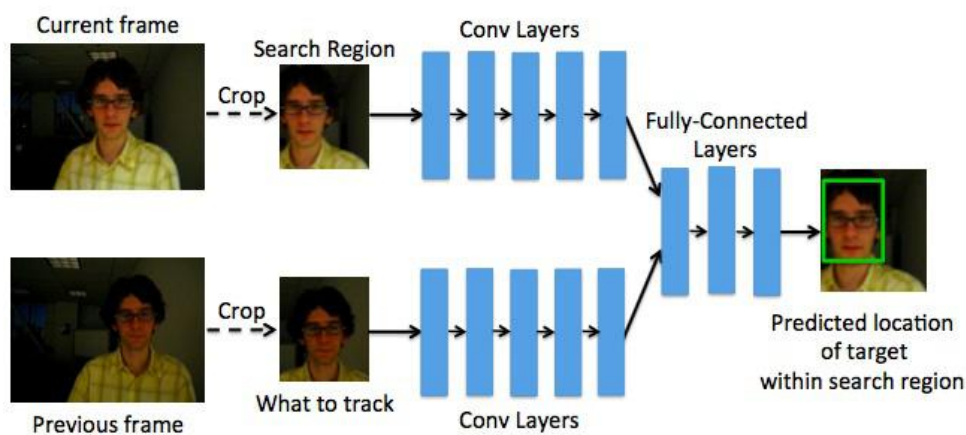
四、创新亮点



图 4 关键算法实现

作品核心在于冰球追踪和游戏策略两个功能均通过深度学习或者强化学习实现，在算法的优化和应用上有一定的创新。

4.1 冰球追踪模型



GOTURN 是一种深度学习追踪模型，通过裁剪实现对目标物体的追踪，将追踪能做到 100+fps。该算法对遮挡敏感，但是对视角、形变、光照变化具有鲁棒性。由于冰球游戏场景很少有遮挡情况，故将这种理论模型应用于桌面冰球的球体追踪中。

4.2 Q-learning

Q-learning 则是强化学习领域的一个基本学习算法，通过与环境的交互和学习，可以得到每个状态下的评价函数 $Q(s, a)$ 的值，根据这个值可以针对不同状态进行动作决策，是一种基于马尔可夫模型的策略搜索和迭代模型。我们将这种基本算法应用于冰球对战中的决策部分。

Deep Q-learning

具体算法如下：

Remember: want to find a Q-function that satisfies the Bellman Equation:

$$Q^*(s, a) = \mathbb{E}_{s' \sim \mathcal{E}} \left[r + \gamma \max_{a'} Q^*(s', a') | s, a \right]$$

Forward Pass

Loss function: $L_i(\theta_i) = \mathbb{E}_{s, a \sim \rho(\cdot)} [(y_i - Q(s, a; \theta_i))^2]$

where $y_i = \mathbb{E}_{s' \sim \mathcal{E}} \left[r + \gamma \max_{a'} Q(s', a'; \theta_{i-1}) | s, a \right]$

Backward Pass

Gradient update (with respect to Q-function parameters θ):

$$\nabla_{\theta_i} L_i(\theta_i) = \mathbb{E}_{s, a \sim \rho(\cdot); s' \sim \mathcal{E}} \left[r + \gamma \max_{a'} Q(s', a'; \theta_{i-1}) - Q(s, a; \theta_i) \right] \nabla_{\theta_i} Q(s, a; \theta_i)$$

如果Q-函数对应于最优值 Q^* ，则迭代地使Q值接近它应有的目标值 y_i

https://blog.csdn.net/weixin_44188264

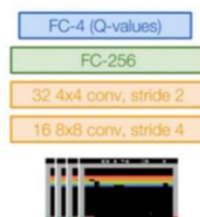
Q-Learning 算法的缺点：状态与状态之间是相互独立的，没有联系。例如：在 Astair 游戏中，再用“表格”来表示 Q 函数就有点不合适了。原因在于 Astair 游戏画面为 210×160 的 RGB 图像，假设用一帧图像表示一个状态，算在每个像素位置只用 0 和 1 里表示，产生的状态也会高达 2 的 210×160 次方种，这就意味着 Q 函数对应的表格有 2 的 210×160 行，根本无法存储下来，也无法进行训练。所以我们可以采用 Deep Q Network。

Deep Q Network

DQN 算法用一个深度卷积神经网络来表示 Q 函数。它的输入状态是 S，输出每个动作对应的 Q 函数值。假设一共有 4 用种动作，分别用 a1, a2, a3, a4 来表示，那么神经网络的输出是 $Q(s, a1), Q(s, a2), Q(s, a3), Q(s, a4)$ 这个神经网络叫做

Q-network Architecture

$Q(s, a; \theta)$:
neural network
with weights θ



Current state s_i : 84x84x4 stack of last 4 frames
(after RGB->grayscale conversion, downsampling, and cropping)

Deep Q Network。

如果说 GOTURN 是我们的眼睛，负责追踪和识别球的位置和状态，那么 Q-learning 就是我们的大脑，负责决策和思考价值，而我们设计的硬件则是我们的四肢，负责执行和反馈决策。

五、技术难度

核心技术大概分为 3 块：

一是识别算法的训练和应用，GOTURN 虽然是一个已经存在的算法模型，但是关于其应用还仅仅局限于人脸和人动作的追踪和识别，而我们通过自己训练，实现了该算法在桌面冰球的识别和追踪，这部分数据集和算法要求比较高，因此是一个技术难点。

二是 Q-learning 的训练，众所周知，强化学习的训练好坏高度依赖参数的调节和训练环境的交互，我们通过不断尝试和经验选择了一个合适的交互环境和奖赏值的设计方法，进而能够为最终决策提供一种优秀的算法基础，这部分的调节和实现是一个经验和理论的融合，因此也是一种技术壁垒。

三是硬件部分，相对算法，硬件相对开放，但是我们的控制结构和硬件模型通过 3D 打印零件和部分合金零件组合构成，形成了稳定的控制结构，这样的硬件设计难度较大，控制较为复杂。