arXiv:2302.04530v1 [cs.RO] 9 Feb 2023

基于注视的意图估计：人机交互中的原理、方法与应用

ANNA BELARDINELLI，本田欧洲研究所，德国

意图预测已成为人机交互领域的重要研究方向。任何与人协同运作、旨在帮助人类伙伴并与其协调动作的人工系统，都能从提前推断人当前的意图中获益。为了免去用户明确说出其目标所带来的认知负担，这种推断主要依赖被视为能够表明当前动作的行为线索。人们早已发现，眼动能够高度预示任务过程中即将展开的各个步骤，因此可作为一种非常早期且可靠的意图识别行为线索。本综述旨在梳理心理学文献中有关视觉运动控制的研究成果，以及技术领域中基于注视的意图识别的相关应用，重点关注远程操作和辅助机器人系统。本文从意图、眼动与动作之间关系背后的认知原理出发，探讨眼动追踪和基于注视的模型在人机交互意图识别中的应用，并介绍常用方法及其多样化应用。最后，本文特别关注设计此类系统时需要考虑的相关人因问题和当前局限。

CCS 概念：• **综合与参考** → 综述与概述；• **计算方法** → 合作与协调；
• **以人为中心的计算** → **人机交互理论、概念与模型**。

其他关键词与短语：意图识别、人机交互、基于注视的交互

1 引言

意图识别算法目前是快速发展的技术的核心，这些技术涉及人与人工系统之间的交互，例如遥操作机械臂、驾驶辅助系统、机器人轮椅，或是在与人类共享的环境中行动的社交机器人 [Vernon 等人 2016]。这些系统通常依赖多种机器学习技术，依据部分观测到的行为线索时间序列，尽早估计正在发生或即将发生的动作。此外，随着小巧廉价的可穿戴摄像头日益普及，利用自我中心视角可以有效观察意图如何演变为动作。近年来，人们逐渐认识到，在许多应用中，利用注视和眼动是一种自然、及时且便捷的意图预测方式。以身体动作中的意图为例，考虑到需要应对感觉运动延迟 [Miall 和 Reckess 2002]，，在基于任务的场景中，注视控制本身可被视为对诸多与动作相关方面具有内在预测性。事实上，移动眼睛时，我们会运用基于知识和感觉运动的经验与记忆 [Fiehler 等人，2019；Hayhoe，2017；Henderson 2017]，以迅速提取规划肢体运动所需的信息。在整个动作执行过程中，我们也会运用眼睛来启动并监督动作。其他人也会利用这类注视线索（以及其他线索）来推断我们的意图，并可能据此提供帮助或作出相应反应 [Emery，2000年；Huang 等人，2015；Tomasello 等人 2005]。注视确实是社会互动中的重要线索，也常用于人机交互（HRI） [Admoni 和 Scassellati 2017]。这些考量表明，意图识别是一项高度认知性的任务，因此它是智能协作系统的一项关键特征。一般而言，正如Bulling 和 Zander [2014], 所指出的，情境感知或上下文感知系统已经被开发出来，以主动帮助用户；然而，人们更多关注环境情境，直到近期才开始关注用户的认知情境。考虑到这一点，

作者地址：Anna Belardinelli，anna.belardinelli@honda-ri.de，本田欧洲研究所，Carl-Legien大街30号，德国奥芬巴赫，63073 .

例如，通过推断意图来推进认知感知计算。事实上，正如Zhang 等人[2020], 所指出的，虽然目前依设计依托以人为中心的人工智能的机器人系统正日益普遍地进入人类社会环境，但这类智能体需要感知并利用人类注视，才能更好地理解人类并与之互动。与人机交互中的其他认知功能和模型一样，基于注视的意图识别既可以采用纯数据驱动的方式，也可以采用更具假设导向或模型驱动的方式 [Tsotsos 等人 2019], ，而近年来在应用领域中，前一种方式往往更受青睐。如今，针对任何特定应用，大量采集注视数据已十分容易；不过，我们想强调的是，更深入地理解人类视觉运动系统，也能大大有助于优化此类数据的利用。在这方面，尽管机器学习方法可以不依赖先验地学习注视与意图之间的相关关系，但了解意图和动作如何在人类心智中产生，以及眼睛之间的协同如何促成这一过程，或许有助于识别相关特征，并根据具体情境和活动设计更好的系统。由于当前技术系统往往缺乏这类背景知识，或未能充分将其系统化，因此在综述基于注视的意图识别文献之前，我们将先概述认知科学研究，回顾意图如何形成、意图在动作生成中的作用，以及注视和视觉系统如何以整合的方式促进这种联系，并着重说明眼动在自然眼手协调中的预测作用。我们还将探讨通常会考量哪些注视特征、如何计算这些特征，以及它们能提供何种洞见，同时考察最常采用的预测模型。这也可能取决于意图识别具有重要应用价值的技术领域（见图1，概要见此）。近年来，利用注视追踪设备推断用户意图的做法日益增多，但其广泛的多用途应用仍受到限制：意图既被视为较高层次的活动（例如阅读、浏览，见第4.1节），也被视为原始动作层面的行为（例如拾取、放置或制动，见第4.2 节和第4.3节），但语义层次更高的有意操控意图似乎更难以捉摸，也更难捕捉。鉴于此，我们最后将探讨若干局限、挑战及相关未决问题；这些问题都需要纳入考量，才能使意图识别系统更加有效，并更容易被人类用户接受。

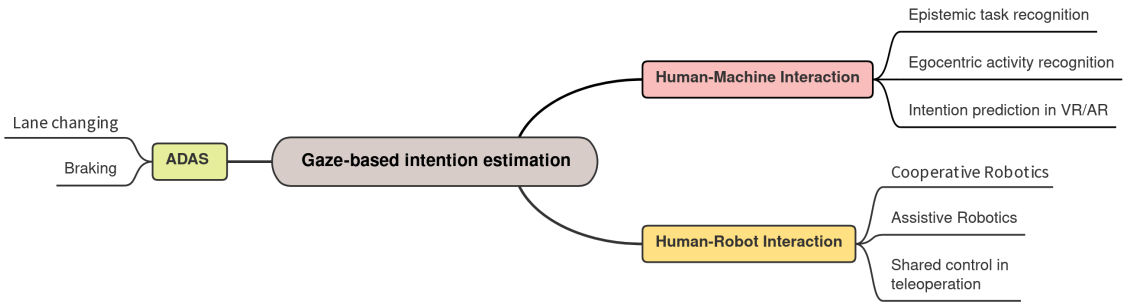


图 1. 本综述所调查的基于注视的意图估计应用领域及子领域。

2 意图、注视与动作

通过对技术文献的梳理，可以发现许多术语似乎与意图识别指向同一概念：意图估计、意图预测、意图推断、动作预测、活动

识别、任务预测，以及在某些情况下的计划识别。这些术语都假定人类行动者表现出有目的、有意识的行为，而人工系统能够依据不同的可观察线索对其进行识别/预测。这些线索包括由摄像机从外部追踪的身体运动、由穿戴式惯性测量单元或鼠标/操纵杆控制器检测到的动作、来自肌电活动或脑机接口的信号，以及注视控制；注视同样可以通过外部摄像机或以自我为中心的头戴式眼动追踪设备进行观察。对于计划识别而言，这些线索无法让系统对未来进行过于长远的预测，而且还需要具备某种心智理论，并结合规划器 [Baker 和 Tenenbaum 2014], 因此，我们不会评述这类系统。正如 Kong 和 Fu [2018], 所指出的，尽管有时这些术语的区分仍不够明确，动作识别 通常是在离线状态下进行的，且待识别的动作已经完成；而动作预测则依赖对上述线索的初步观察，以估计（可能是在线地）正在进行的动作。不过，在谈及意图时，这种区分几乎不适用。正如下一节将回顾的，意图是隐藏的、不可观察的状态，只有在转化为运动行为后才能被推断。从这个意义上说，意图总是在事后 *post-hoc*识别或推断，因为它们已经形成，而相应的动作仍可能在完成之前被预测。因此，本文将意图估计、预测、识别和推断视为同义词，并采用前者作为更通用的术语。

基于上述原因，本文综述聚焦于注视这一意图估计的关键线索；即使手部动作尚未开始，它也有助于估计意图。从这一视角来看，什么是意图？意图与动作有何关系？这种关系如何通过我们的视觉系统和注视控制系统来实现?

2.1 从意图到动作

长期以来，意图作为一种心理状态，一直被视为哲学行动理论中的关键要素。哲学家 Dennett 提出了意向立场，认为这是预测他人行为和动作的基础视角[Dennett 1989]。只要将信念、欲望和意图归于任何理性主体，我们就能够预测他们将如何实现这些意图。事实上，在运动认知领域，意图与动作生成和能动性密切相关，也是行动解释的核心 [Pacherie 2012]。早期理论[Davidson 1980; Goldman 1970] 仅将信念和欲望视为因果先行因素，更具体地说，将其视为动作的心理原因。然而，动作可以被看作一个连续谱，涵盖从低层次运动基元到经过周密计划、审慎执行的动作。从这一视角来看，除了信念和欲望之外，人们还引入了意图，以解释那些面向未来、却不会立即引发动作的意图，以及那些出于本能、几乎自动执行、却仍体现最低限度目的性的动作 [Bratman et al. 1987]。基于这一认识，人们认为意图具有多种功能，并根据其与后续动作之间的时间关系对其进行界定。具体而言，在任何实际动作开始之前，意图会促成计划，并协调计划活动的时间安排 [Bratman et al. 1987], ；此外，意图还发挥三项关键功能：启动、引导和控制动作直至其完成 [Brand 1984; Mele 1992]。这些区分催生了先行意图与行动中的意图区分 [Searle 1983] 、面向未来的意图与面向当下的意图区分 [Bratman et al. 1987]，以及远端意图与近端意向的区分 [Mele 1992]。

在这一思路下，Pacherie [2008] 提出了更细致的分类：D-意向、P-意向和 M-意向。这种区分基于这样一种考虑：远端意图不会在近端意向承担起监控和引导作用后便停止发挥作用。相反，这三种意向（远端、近端和运动意向）会在不同的表征内容层次上、以不同的时间尺度并行发挥作用。在这一框架中，

D-意向界定要追求的目标，并且在时间和情境方面具有灵活性，而对计划中的动作参数和约束则没有明确规定。P-意向接收来自 D-意向的行动计划，并触发动作的开始，将其置于特定的时间和情境中，同时依赖明确的感知、情境意识和思维。不过，P-意向仍可作用于较高层次的动作描述，而 M-意向则处理具体的动作原语和运动表征（见图 2）。这三个层次当然以不同的时间尺度运作：D-意向的时间尺度为数分钟乃至更长，P-意向为数秒，M-意向则为数百毫秒。

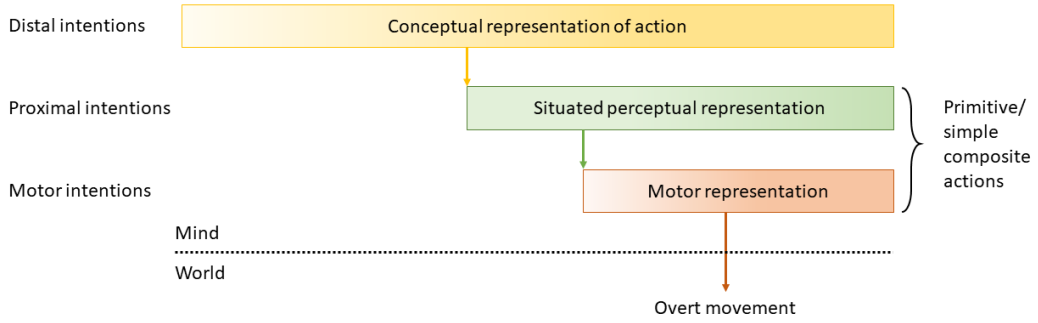


图 2. 意图层级级联的图示：从抽象的远端意图，到以情境化方式直接转化为运动动作的运动意向（改绘自 [Pacherie 2008]）。

这些形式化表述表明，动作源于计划与运动程序，它们结合在一个依赖前向模型和逆向模型的功能架构中。这意味着，任何能够与运动计划的形成相关联的行为观察，都有可能揭示与当前近端意向相对应的动作。从这个意义上说，尽管许多方法都依靠运动线索来推断意图 [Aarno and Kragic 2008；Hauser 2013；Javdani 等, 2015；Tahboub 2006；Tanwani and Calinon 2017；Yu 等人 2005], 还有一种更早出现的指标，它在行动准备阶段就已显现，那就是注视控制。为什么会这样？视觉系统和注视系统如何介导意图向运动动作的转化？

2.2 自然任务中的基于注视的预测与眼手协调

任务会强烈影响注意，进而塑造眼动模式：在一项早期眼动追踪实验中，Yarbus [1967] 发现，向受试者呈现同一张图片，每次都要求他们回答不同的问题（例如，估计图片中人物的年龄或财富），会产生稳定且不同的扫描路径。

如今，我们的大部分视觉活动可能都与被动处理文本、图像、视频及其他媒体中的信息有关。然而，我们的视觉器官（以及整体认知）实际上是在一个需要通过在环境中采取动作来适应并生存的运动有机体中发展和演化而来的。即使是在简单的视觉搜索过程中，我们的性能也会受到动作意图和可能性的影响 [Bekkering and Neggers 2002；Bub 等人2013]。从理论层面来看，注意的前运动理论早已提出了这一点 [Rizzolatti 等人 1987],认为，空间注意表现为对可能执行动作的位置进行增强的空间加工，这表明

注意与动作之间存在紧密联系并共享机制，尤其是知觉选择与动作选择之间 [Deubel 等人 1998]。当我们打算采取动作（或需要作出反应）时，正是信息寻求型知觉行为必须最有效地围绕最终目标进行调整的时刻。动作实际上大多是序列性的（我们几乎不会同时协调具有不同目标的运动），并且由于执行过程中的惯性，其速度相对较慢。因此，为了最大限度地提高动作成功的概率，运动规划需要及时利用感知所提供的确定信息。这对应于一种视觉运动行为策略，力求最大化信息增益，并最小化不确定性或意外 [Friston 等，2015；Itti 和 Baldi 2009]。为此，视觉系统需要依赖以往的经验和知识，从而根据当前的运动和子任务需求，在准备和控制特定目标导向动作时，预先判断需要提供的关键信息。尽管这一点在运动和导航任务中同样显而易见 [Einhäuser 等人，2007年；Werneke 和 Vollrath2012], ，但在与物体互动时，这一点表现得最为明显。过去二十年中，头戴式眼动追踪器的出现使得以生态学方式研究眼手协调成为可能，研究对象可以执行制作三明治等日常任务 [Hayhoe 等人 2003], 泡茶 [Land 等人 1999], 四处行走 ['t Hart 和 Einhäuser 2012] 或进行运动练习 [详尽综述见 Land 和 Tatler 2009]。具体而言，在眼手协调方面， [Johansson 等人 2001] 研究了注视与手部行为之间的时序相关性：研究对象需要抓取、提起并携带一个简单的木块，使其接触目标位置，同时避开一个凸出障碍物。

这些研究得出了不少经验，尤其是在考察与物体进行有组织的手动交互时：

- 注视通常只会指向任务相关物体 [Land 2006];
- 人们通常会在任务需要物体之前不久瞥向它们——即“前瞻性”和“即时”注视[Hayhoe 等，2003年； Pelz 和 Canosa 2001]——并持续注视，直到手即将抓取物体或任务完成。也就是说，即使动作任务尚未完全完成，眼睛也可能先行移向下一个子目标；

- 人们很少注视手部或手中拿着的物体 [Johansson 等 2001]。一般而言，眼睛相对于手部会提前约 0.5 至 1 秒（眼手时距），这表明可以利用注视来预测下一个操作步骤。

这表明，在执行眼手协调（操作）任务的过程中，眼球运动计划和手部运动计划如何展开，以及它们与所涉及物体之间的关系，为理解由意图统筹的动作表征与规划的认知结构提供了重要洞见。一旦我们确定要实现某个特定目标，针对两种行为（信息搜寻和物体操作）并行且紧密交织执行的指令便会形成，并分别传递给注视与视觉系统，以及运动系统。Land [2009]提出，这一过程涉及四个系统（见图 3）：

(1) 图式控制 设定目标，并通过规划任务序列中必要步骤的安排，制定实现目标的粗略计划。具体而言，需要向后续的其他系统传达有关每个子任务中物体的一般指引、物体的位置，以及交互所需的运动基元。

(2) 注视系统 负责控制身体各部位的运动，引导我们的眼睛转向最可能找到待操作物体的位置。因此，身体／躯干运动、头部运动和眼动都由注视系统统筹[1]。

(3) 视觉系统 负责处理视觉输入，并识别视觉场景中的物体及其情境。它还会监控动作是否顺利展开，从而实现每项任务的目标。
(4) 运动系统 负责在身体层面执行预期动作，并结合可用的运动基元（如伸手和抓握）与当前物体构型；相关信息由视觉系统提供。

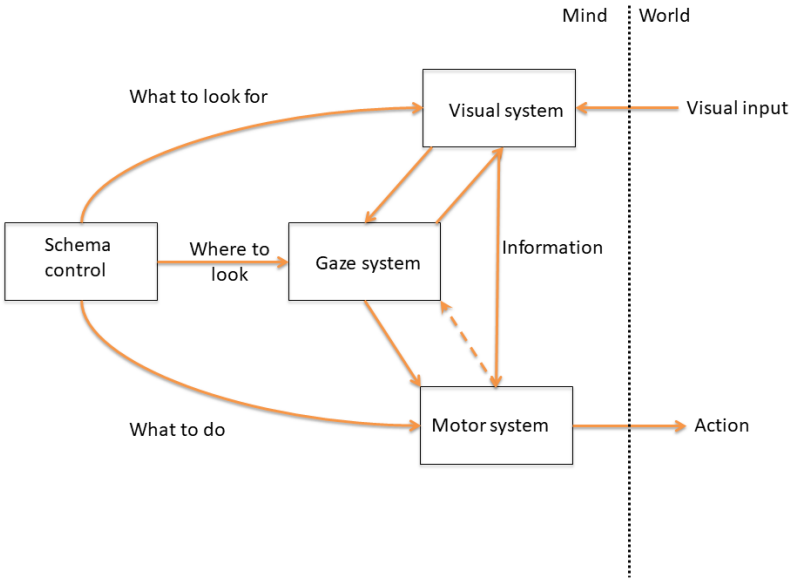


图3. 调节感知与动作的功能模块之间的信息交换：注视控制、视觉处理和动作控制之间的关系由图式系统统筹，该系统负责安排要注视的位置、要寻找的物体以及要计划的移动。实线箭头表示由 [Land 2009], 勾勒的信息流，此处重新绘制；虚线箭头表示本文提出并在 [Belardinelliet al. 2016a,b; Brouwer etal. 2009]所示的进一步信息交换。

显然，这些系统必须紧密协同工作，并且彼此交换信息，尤其是注视系统与视觉系统之间，以及这两者与运动系统之间。例如，如果视觉系统未能在当前视野中检测到我们正在寻找的物体，注视系统就必须知道还需要进一步探索。反之，如果已找到动作目标，运动系统就必须知道它在视野中的位置、相对于身体其他部分的位置，以及如何抓取该物体。这些信息交换在图 3 中以不同系统之间的实线箭头表示。

图式控制系统的活动无法通过行为直接观察，但它与视觉、注视和运动系统之间的关系，使我们能够通过观察任务中的注视位置、扫描路径，以及可能的手部运动及其时间同步情况，推断其意图。不仅是接下来要抓取或使用什么物体

[1]尽管有人可能会认为，这些移动实际上是由运动系统控制的，但在这种情况下，这些探索性移动具有认知价值（降低不确定性），而实际执行动作的移动则可被视为在伸手实现目标状态时具有实用价值 [Friston 等人 2015]。

也可以根据手伸向物体之前的注视来推断。许多研究表明，任务、预期的抓取方式以及用户对物体的使用经验，都会影响手伸向物体之前的注视[Belardinelli等人 2016年a，2015年，2016年b；Brouwer等人 2009]。

在规划伸手抓取动作时，眼睛通常会注视食指接触位置，因为食指通常是最先与物体接触的手指 [Brouwer 等，2009；Cavina-Pratesi 和 Hesse 2013]。一般而言，人们会用视觉瞄准物体上相关的可供性点；如果可能，也会同时瞄准，因此，要么以隐蔽方式注意这些点，要么将注视点对准中间位置。由此形成一种注意景观，增强与动作时空要求相关的知觉加工 [Baldauf 和 Deubel 2010]。也就是说，通过观察不同任务和条件下对物体的视觉探索及随后的操作，可以看出规划和执行每项任务所需的信息各不相同。

这些研究还表明，上述示意图中缺少另一条箭头，即从运动系统流向注视系统的信息。也就是说，只有在选定适当抓取方式的运动计划后，眼睛才能移动到最合适的位置来引导抓取。图中的虚线箭头表示这种信息交换 3。

眼睛先于手部运动，并满足手部的信息需求；与此同时，运动系统将其运动学需求传达给注视，这一切都发生在动作执行开始之前或开始过程中。这表明，感知与动作在很大程度上是具身且整合的过程，依赖于对预期感觉运动目标状态的学习和预期。

目前所述内容涉及抓握和操控过程中视觉注意与眼动控制的认知基础。虽然驾驶仍需要眼手（以及眼脚）协调，但像驾驶这样更复杂的活动还需单独考量；此外，驾驶也是许多应用的领域，这些应用需要依靠及时的意图估计来实现人机协作。正如上文所述，注视控制依据对当前场景中相关信息所在位置的空间预测 [Henderson 2017], 而这些预测往往基于情景经验或知识。结合自上而下的内源性注意，这种预测能力使视觉特别高效，能够在数百毫秒内聚焦于复杂场景中的相关信息。然而与此同时，我们必须能够应对意外刺激，尤其是那些与当前任务无关、但检测到它们对自身及他人安全至关重要的刺激。当然，这在动态环境中行动时尤为重要，因为错误的预测需要迅速反应。驾驶发生在这类环境中，需要扫描的场景通常信息丰富，且各类信息的相关程度不同。驾驶本身也是一种多任务处理：它不仅仅是将车辆转向预定方向，还由多个并行子任务（碰撞规避、导航规划、车道保持……）构成；每项任务的优先级都会随时间变化，因此需要持续重新调度注意资源分配 [Hayhoe 和 Ballard 2014；Rosner 等人 2019]。也就是说，在结构相对顺序化的任务中，注视行为可以视为当前认知议程的实时记录，并在一定程度上揭示当前子任务；而在多任务处理过程中，注视选择并非依据任务的线性展开来进行，而是取决于彼此竞争的模块。这些模块需要定期利用新信息更新，尤其是在无法假设世界状态静止不变时。更复杂的是，扫描行为可能因情境（城市、高速公路或乡村驾驶）、当前要执行的操作（超车或通过交叉路口）以及当前速度而改变。这些因素不仅会影响注视位置，也会影响整个视觉运动控制系统。驶过弯道时，眼睛会注视

视线与车道边缘相切处的切点 [Land 和 Lee 1994；Mars 2008], 并比转向的手提前约 0.8 秒。在城市交通中，注视可能每 0.5 秒转移一次，在前方车辆、车道边缘、行人、道路标志、交通信号灯或一般障碍物之间交替 [Land 和 Tatler 2009]。不过，很难脱离具体情境对注视分布进行概括。此外，已知认知负荷越高，注意就越集中于中央凹视觉，从而牺牲周边视觉 [Crundall 2005]。

人们尤其关注变道过程中的注视行为。变道需要情境意识、决策和运动控制之间紧密协调，因为这一过程会与其他移动中的车辆相互影响。 Salvucci 和 Liu [2002] 研究了这种情况下的注视停留时间，发现变道开始时（车道保持阶段），注视大多停留在起始车道上；从右侧车道变至左侧车道时，随着变道时刻临近，驾驶员会花更多时间注视后视镜以检查驶来的车辆，并注视目标车道。反之，从左侧车道变至右侧车道时，变道前不久对后视镜的注视会增加，但在变道过程中及之后，注视大部分时间都集中在目标车道上。作者指出，注视会预先配合这一操作：在变道开始前 1 至 2 秒，注视就已经集中在目标车道上，尽管车辆实际上仍完全处于起始车道内。这些结果表明，驾驶时有多种因素会影响注视的控制方式 [参见 Kotseruba 和 Tsotsos 2021年，可查阅全面综述]。这可能会使评估驾驶员意图变得格外棘手，因为触发一系列注视点的过程（有意获取信息还是注意被反射性地吸引）或许难以明确判断。不过，综合多种行为线索可以提供更丰富的信息。

最后需要指出的是，正如我们所见，意图可以具有多种时间粒度，从远端意图到近端意图，再到原始运动意图。然而，上述研究表明，身体动作中的“即时”注视体现了一种动作规划策略：根据这一策略，系统会在即将执行下一动作之前（即动作开始前数百毫秒）获取与该动作直接相关的视觉信息。随后，只要操作仍需要视觉引导，眼睛就会持续注视关注对象。在驾驶场景中，尽管眼睛收集机动操作信息与手部启动该操作之间的时间间隔通常更长，但仍在 1 到 5 秒的范围内。这样的时间跨度几乎无助于提取长期的远端意图。因此，接下来我们将主要关注处理近端意图识别和运动意图识别的模型与系统，以驱动单个、通常是原始的动作。我们的目标是说明，这类案例迄今确实是将注视用于意图识别的最有力应用，以及如何借助易于计算的特征轻松提取此类意图，下一节将对此进行介绍。

3 注视特征与意图预测模型

在探讨注视如何应用于不同应用领域之前，我们先简要概述常用的注视与眼动特征及测量指标，以及能够根据目标充分利用这些信息的机器学习模型。尽管有许多研究采用眼电图 [，参见 Bulling 等人，2010，综述] ，或仅使用由头部运动引起的光流特征[，例如 Kitani 等人 2011] ，或使用外部摄像机提取的注视方向 [，例如 Dermy 等人 2019], 。本文主要关注使用头戴式设备采集的注视与眼动追踪数据，因为这种技术能够更准确地确定注视位置，并将其定位到场景中，从而将眼动模式与任意时刻感兴趣的特定物体或位置联系起来。

下表列出了下文将使用的不同缩写 1。

缩写	含义	缩写	含义	缩写	含义
ADAS	高级驾驶辅助系统	**GMM**	高斯混合模型	**部分可观测马尔可夫决策过程**	部分可观测马尔可夫决策过程
AOI	兴趣区域	**HMI**	人机交互	**POR**	注视点
AUC	曲线下面积-The-Curve	**HRI**	人机交互	**RBF**	径向基函数
CRF	条件随机场	**k-NN**	k近邻邻居	**（循环）神经网络**	（循环）神经网络
（动态）贝叶斯网络	（动态）贝叶斯网络	**LDA**	线性判别分析	**SVM**	支持向量机
DTW	动态时间规整	**LSTM**	长短期记忆	**VR**	虚拟现实
（高斯）隐马尔可夫模型	（高斯）隐马尔可夫模型	**MAP**	最大后验概率概率		

表1。文中使用的缩写。

3.1 注视特征：位置、移动、序列和语义特征

无论采用何种特定硬件，所有眼动追踪设备都会将眼睛的位置作为原始数据输出：以二维坐标的形式表示在校准表面上，或根据头戴式场景摄像机的画面流，以帧坐标的形式表示在场景中。对于许多涉及与现实世界中的物体进行物理交互的应用，可能需要在三维空间中定位注视点（兴趣区域），为此已有一些专门的方法被提出，这些方法通常依赖基准标记 [例如，Haji Fathaliyan 等人，2018；Huang 和 Mutlu，2016；Takahashi 等人， 2018] 不过，新的解决方案正尝试摆脱对基准标记的依赖 [Liu 等人，2020；Weber 等人，2018]。

无论如何，获得注视坐标后，都可以直接使用这些坐标，也可以通过特定检测器计算感兴趣的注视事件，通常是注视点和扫视。注视点是指注视相对静止于某个点的一段时间，而扫视则表示注视点之间兴趣区域的快速位移。注视通常持续约 200–300 毫秒，而扫视可能持续 30 至 80 毫秒。有关眼动的更全面分类，请参阅 Holmqvist 等人 [2011] 。在扫视期间，视觉输入会变得模糊，几乎无法进行视觉处理，因此分析眼动数据时通常只考虑注视点，认为视觉信息仅在这些时刻被摄取。为此，可以通过基于离散度或速度的算法检测注视点 [Salvucci 和 Goldberg 2000], ，但这些算法仍基于预定义的阈值。近年来，有研究提出了一些基于机器学习的事件检测方案，其准确率更高 [Tafaj 等人，2012年；Van der Lans 等人，2011年；Zemblys 等人 2018]。不过，当系统需要在线运行以进行实时预测时，直接使用所有注视采样可能更为合理，从而避免在意图分类之前先通过在线注视检测器处理数据 [Coutrot 等人，2018年；Fuchs 和 Belardinelli 2021]。直接使用注视点确实同时考虑了关注对象的注视位置及其时间分布 [例如Keshava 等人 2020]。

在活动识别或任务预测中，有时可以不依赖底层视觉输入，单独使用注视转移模式 [例如，Bulling 等人，2010；Ogaki 等人， 2012]。在这方面， [Srivastava 等人 2018] 对这类特征作了区分：低级特征指单次注视／扫视的持续时间和位置统计，而基于形状和距离的模式（量化为字符串）的统计则被视为

中层特征。这类模式在不同桌面活动中的分布确实有所不同，例如 [Ogaki 等人，2012年；Srivastava 等人 2018]。通常会使用注视和扫视统计数据（例如注视时长或扫视幅度的均值和方差），以及注视落点处图像特征的统计数据（例如方向梯度直方图、显著性或 Gist 特征），来推断当前的认知任务 [参见Boisvert 和 Bruce，2016年，相关综述]。

然而，在预测身体交互时，需要考虑更高层次、某种程度上更具语义性的特征。尽管图像上的注视分布或方向性扫视转移的重复模式能够揭示特定的重复性活动，但这些特征很难与单一的动作意图联系起来（不过，它们至少能够揭示交互意图 [David-John 等人 2021]）。事实上，当前关注的物体，甚至该物体上的特定可供性，往往能够揭示下一个近端意向 [Belardinelli等人 2016a，2015], 而不受其在相机图像中的位置或注视点处特定低级特征的影响。因此，场景中的相关物体应定义为感兴趣区域（AOI），并对这些区域内的每个注视点进行处理并统计相关数据。在这种情况下，可以得出针对各兴趣区域的指标，例如每个兴趣区域内的注视点数量、进入各兴趣区域的时间以及总停留时间。更重要的是，考虑到意图和动作随时间的演变，可以将扫视过的兴趣区域序列用作离散符号的时间序列，从而揭示当前活动中注意力在不同相关物体之间的转移，例如Haji Fathaliyan等人 [2018]。为考虑眼球运动的不准确性以及旁中央凹视觉的策略性运用，在某些情况下，不会在每个时刻都将注视完全归于单一物体，而是将兴趣区域表示为高斯分布 [例如，如 Admoni 和 Srinivasa（2016）中所示；Fuchs 和 Belardinelli 2021] 也可能会考虑与不同兴趣区域相交的区域，或注视计数的时间区间，并据此计算兴趣区域上的离散分布 [Lengyal等人 2021]。在车辆中，兴趣区域常被用作具有语义意义的惯性区域。虽然在存在自我运动的动态场景中估计当前关注对象在技术上相当复杂 [，但可参见 Jiang 等人 2018年，其中提出了一个较新的解决方案], 。可以通过考察驾驶员当前关注车内视野的哪个部分来评估车道保持或变道等意图：通常将挡风玻璃、后视镜和侧视镜以及仪表板视为兴趣区域。在任何情境下，兴趣区域序列都可以视为字符串，其中典型的重复转移模式很适合通过马尔可夫模型进行学习，下一小节将对此进行讨论。

3.2 预测模型

通常，意图识别系统会将被认为能表明当前意图的注视特征输入某种分类器，由分类器输出最可能的意图。为此，研究中常使用支持向量机（SVM） [Huang 和 Mutlu，2016；Kanan 等人 2014], 随机森林 [Boisvert 和 Bruce 2016], 逻辑回归 [David-John 等人 2021], 以及神经网络 [Lethaus 等人 2013] 。在某些情况下，例如图像观看中的任务预测 [Henderson 等人2013] 或驾驶操作 [Lethaus 等人 2013], ，朴素贝叶斯分类器也能实现令人满意的意图识别，但大多数成熟的方法采用了更复杂的生成模型。事实上，如前所述，要预测意图，我们需要估计一种认知状态，而这种状态本身不可观测，也就是隐藏状态。因此，意图识别常通过概率图模型来建模，例如动态贝叶斯网络和隐马尔可夫模型 [Bader 等，2009；Boccignone，2019；Fuchs 和 Belardinelli，2021；Haji-Abolhassani 和 Clark，2014；Tahboub 2006；Yi 和 Ballard 2009]。一般而言，如果意图 I 代表隐藏状态，而可观测内容（感兴趣的眼部特征）表示为 E，则通常可以用贝叶斯形式来定义意图识别：

$$P(I|E) = \frac{P(E|I)P(I)}{P(E)} \tag{1}$$

如果我们针对每种意图都有一个模型 M_I，并且有一系列特征 $E_{0..t}$，那么当前意图 I_t 的推断结果就是使观测数据后验概率最大的模型：

$$I_t = \underset{I}{\operatorname{argmax}}\, P(M_I|E_{0..t}) = \frac{P(E_{0..t}|M_I)P(M_I)}{P(E_{0..t})} \tag{2}$$

近期不少方法也采用了循环神经网络，例如长短期记忆网络（LSTM）[González-Díaz 等，2019；Min 等人，2017；Schydlo 等人，2018；Wang 等人 2020] 以及部分可观测决策马尔可夫模型（POMD） [Admoni 和 Srinivasa 2016], 这些算法同样能够很好地处理长度不同的时间序列。

一般来说，采用能够反映场景中扫描路径随时间演变的模型是合理的。事实上，如果只考虑当前的注视点或兴趣区域，那么当注视落在任何与当前意图并无直接关联的对象上时，预测就会变得非常不稳定。即使是前文提到的依赖判别分类器的方法，也不会只考虑某个特定时刻的注视特征，而往往会计算一定时间窗口内的汇总统计量。这样虽然纳入了较长的历史信息，却丢失了时间信息。生成模型同样依赖在一定时间窗口内观测到的数据，但它们会纳入对任务演变过程的内部表征，因此也能更好地应对无关注视点。不过，必须谨慎选择所考虑的时间窗口，以捕捉待估计意图的时间尺度。对于伸手、抓握和简单操作，0.5 至 2 秒的时间窗口通常能取得最佳效果 [Fuchs 和 Belardinelli，2021；Huang 和 Mutlu 2016], 而在驾驶场景中，通常会考虑更长的时间范围，即 5 至 10 秒 [Lethaus 等人 2013]。

此外，尽管隐马尔可夫模型能够以相对灵活的方式处理长度不同的序列，但有时需要将时间序列缩减至可比较的长度，才能计算距离或相似度。动态时间规整主要用于解决这一问题 [Haji Fathaliyan 等人，2018；Wu 等人 2019] ，也可以通过编码器-解码器 LSTM 架构将序列编码为固定长度的表示 [Schydlo 等人 2018]。

最后还需考虑一点：在许多情况下，与其采用只返回最可能意图的分类方案，不如获得涵盖所有可能意图的概率分布。之后可以通过 argmax 函数进行选择（如公式2所示），同时也可以衡量预测的置信度，例如计算可能性最高的两个意图之间的差值 [Jain 和 Argall 2018]。这样一来，可以根据系统对所估计意图的确信程度，相应地调整后续辅助的力度。在下一节中，我们将根据具体用例和待识别意图的性质，介绍所述技术和特征在多个应用领域中的主要应用情况。

4 技术领域中的基于注视的意图估计

如图1 所示，注视特征已被用于三个主要领域，以广义方式进行意图识别。在计算机视觉领域，意图通常是在事后离线识别的，也就是说，根据收集到的数据离线推断认知任务。尽管可以设想将其用于电子游戏、虚拟现实或其他扩展现实（XR）应用等更具交互性的场景，但在这一领域，我们大多将意图识别视为活动/任务分类，并依赖与整体注视分布相关的特征。相反，在机器人系统（例如遥操作机械臂或驾驶辅助系统）需要迅速响应或支持人类现实世界活动时，

与关注物体相关的兴趣区和时间序列变得尤为重要，而识别的及时性也是一个重要参数。我们在此综述这三个领域中一些具有代表性的系统案例。

4.1 计算机视觉与人机界面

在过去二十年里，随着眼动追踪技术和适用的机器学习算法不断发展，研究人员常常利用眼动来逆向开展 Yarbus 实验，并以高于随机水平的准确率推断任务（被提问的问题） [例如，Borji 和 Itti 2014; Haji-Abolhassani 和 Clark 2014; Kanan 等人 2014]。除了 Yarbus 的问题之外，这些方法大多用于对静态图片上的被动信息搜寻任务进行分类，例如统计特定项目的数量、视觉搜索、场景记忆和阅读。通常，这些研究会利用一些特征，例如注视时长和扫视幅度分布的统计参数、注视点数量、转移矩阵，以及兴趣区域（面部、身体、物体）中的注视比例和持续时间。根据眼动及其可能的序列，计算给定任务发生概率的常用有效技术包括朴素贝叶斯分类器、隐马尔可夫模型、支持向量机、多变量模式分析和随机森林[更全面的综述请参见Boisvert 和 Bruce 2016]。

不过，在这些情况下，被识别（而非预测，这种情况很少见）的确是一般性的认知任务，而不是某种特定意图。考虑到观察者与（动态）场景之间涉及身体和手动交互的任务时，基于注视的意图识别有着更切实的应用。可穿戴摄像头的普及率大幅提升，推动了从第一人称视角观察日常活动并进行识别的研究 [Fathi 等, 2012; Ogaki 等人, 2012年; Yi 和 Ballard 2009], , 因此依赖眼睛、手、头部以及可能还有身体之间的协调 [参见 Nguyen 等, 2016，其中有完整综述]。

然而，这些方法关注的更多是整体活动识别，而非简单动作或近端／运动意图识别；后者在虚拟现实或交互式游戏中尤其有用。
例如，考虑虚拟现实中的简单手动动作，Keshava 等人 [2020] 最近表明，不同兴趣区中的注视点所占比例本身就构成一种可靠的特征，可以区分不同任务。在大屏幕桌面上操作虚拟物体时进行基于注视的意图识别，这一方法此前已由 Bader 等人 [2009] 在一项拾取与放置任务中提出。研究使用贝叶斯网络，根据主动注视来预测意图，以提高输入设备的稳健性或增加交互可能性。与此相反，Singh 等人 [2020]提出了一种旨在进行长期计划识别的方法。在一个数字化棋盘游戏版本中，游戏路线以图的形式呈现；对任意兴趣区域的注视次数和持续时间决定了用于为可能轨迹赋予先验的分数，这些轨迹会导向远端意图。作者表明，借助这类注视先验，可以更早、更准确地推断近端和远端意图。

在数字游戏领域，通过注视进行计划与目标识别，有助于根据玩家的意图量身定制游戏过程，呈现个性化叙事，或与非玩家角色互动。 [Min 等人 2017]展示了在开放世界游戏中，结合注视与游戏事件，借助长短期记忆网络架构识别玩家目标的效果优于仅使用游戏事件特征。

本节介绍的研究总结见表 2。这些研究提出的意图识别方法都颇具价值，但大多数实验设置既不涉及合作性的共享任务，也不涉及用户试图控制一个在现实世界中产生影响的物理系统；在这种情况下，及时识别意图对于顺畅的人机交互至关重要。下文将回顾与这一主题相关的研究。

研究	Task	特征	识别 方法	什么是 识别到的
[Borji 和 Itti 2014]	认知任务 + Yarbus任务	注视密度 NSS	kNN, 提升 分类器	任务标签 （例如，年龄猜测）
[Haji-Abolhassani 和 Clark 2014]	认知任务	注视位置序列	HMM	任务标签
[Kanan 等人 2014]	认知任务	注视次数 ， 持续时间和 Fisher 位置上的核特征	SVM	任务标签
[Boisvert 和 Bruce 2016]	自由观看和 认知任务	注视位置、密度， 方向梯度直方图、Gist 特征	随机 森林	任务标签
[Yi and Ballard 2009]	制作三明治 （包含子任务）	注视点、兴趣区序列	DBN	子任务 （离线和在线）
[Fathi 等人 2012]	备餐	注视位置， 基于物体的特征	支持向量机 + 生成 模型	子任务 （动词 +物体）
[Ogaki 等人 2012]	桌面操作	眼睛的 n元语法统计 和自我运动	SVM	活动标签 （例如，浏览）
[Keshava et al. 2020]	对齐立方体 在虚拟 现实中	兴趣区中的 POR 分布	径向基 SVM	对齐类型
[Bader 等人 2009]	虚拟拾取与 在屏幕上放置	注视位置与 反应式/主动式行为	贝叶斯网络	拾取/放置目标
[David-John 等人 2021]	食材取用 在虚拟 现实中	多个特征 基于注视动态	逻辑 回归	交互开始
[Singh 等人 2020]	棋盘游戏 *(Ticket to Ride)*	注视点的数量和持续时间 兴趣区	归一化 注视点 权重	前往 目标目的地
[Min 等人 2017]	开放世界 教育 电子游戏	对虚拟物体的注视点	长短期记忆网络 + CRF	目标/任务 在游戏世界中

表2。人机交互中任务与意图识别的方法。

4.2 人机交互

机器人学或许是最适合应用意图识别的领域，因为在这一领域中，物理系统能够知晓人类追求的目标，并主动提供支持：它可以在同一环境中采取行动（第 4.2.1 节），也可以在远程或无法进入的环境中采取行动（第 4.2.2 节）。

4.2.1 人机协作与社会机器人学。在人机协作中，机器人伙伴通常了解活动情境。为了实现有效协作，它只需检测人类伙伴当前的动作意图，以便协助对方。在最早尝试将眼动追踪应用于人机协作场景的研究之一中，Sakita 等人 [2004] 提出了一种协助人类完成装配任务的系统。如果机器人有把握地检测到用户意图执行的下一步，就会递上下一件待装配的部件，或者直接与人类并行执行该步骤。

研究	Task	眼部特征	识别 方法	什么是 已识别
[Dermy 等人 2019]	协作 拾取与放置	由头部朝向得到的注视向量 头部朝向	Max. 似然， RBF	动作原语 （轨迹）
[Haji Fathaliyan 等人 2018]	粉末饮料 冲调	注视对象序列 滑动窗口内	DTW 欧几里得 距离	子任务
[Huang 和 Mutlu 2016]	冰沙 制作	注视次数， 首次注视时长 总停留时间， 如果最后注视的项目	， SVM	预期配料
[Ravichandar 等人 2018]	伸手	注视图 根据头部朝向	NN	目标位置 轨迹的
[Sakita 等人 2004]	乐高积木装配	注视时间 功能上的注视点 功能点	， 基于规则， 投票方案	装配类型
[Schydlo 等人 2018]	放置、递给、 进食、倒液体	注视位置	LSTM	动作和位置

表 3. 人机协作中的任务和意图识别方法

在该现实世界设置中，注视似乎经过了定位，但关于实现方式和准确度的细节很少。

在大多数情况下，确实会将用户的注视映射到眼动追踪器前置摄像头采集的二维场景中。在研究与真实三维物体的交互时，这种做法相当不便。为解决这一问题，Haji Fathaliyan 等人[2018] 提出了一种方法：将注视向量投影到由某人准备粉末饮料时所操作物体的点云表示上，从而确定注视在三维物体上的位置。研究人员生成了三维热图，展示了在执行不同子任务时，每个物体上最受注视的位置。

在某些情况下，在给定预定义动作的前提下，可以利用注视来识别用户打算选择的物体。例如，Huang 和 Mutlu [2016]提出了一种预测性控制方法，机器人据此预测人类用户的意图，并在收到明确指令之前提前制定计划。在所研究的任务中，机械臂根据人类用户的语音选择，从一份配图清单中拾取配料来制作冰沙。机器人通过眼动追踪预测用户打算拾取的配料，从而加快任务完成速度。意图识别所依据的特征和训练集，与先前一项工作相同 [Huang 等人 2015], ；在该工作中，支持向量机根据每种配料的特征向量进行判断，其中包括注视次数、首次注视的持续时间、总持续时间，以及该物品是否为最近一次注视的物品，这些特征被用来预测用户当前意图选择的配料。尽管这种方法简单有效，但在这里，人类顾客有意且仅通过注视来选择目标，并没有同时进行可能产生无关注视点的视觉运动控制任务。

有些方法同时利用注视特征和身体运动来推断意图。 Dermy 等人 [2019] 利用头部/注视方向（根据机器人的摄像头估算）和/或动觉辅助，推断人类期望机器人执行哪种运动基元。与 Singh 等人 [2020]（见上一节）类似，Ravichandar 等人[2018] 将注视信息作为先验，用于计算人类伸手过程中多种可能轨迹的后验概率，这些轨迹是通过神经网络学习得到的。

不过，在这项研究中，注视也是通过外部摄像机估算的，主要依赖头部朝向。

考虑到不同长度的动作序列，Schydlo 等人 [2018]提出了一种架构，将注视和身体姿态特征输入编码器-解码器长短期记忆网络，并由解码器层展开 K 个最可能的动作序列。训练和测试使用了现有数据集，其中包含放置和递交动作[Duarte 等人 2018] 或进食和倒液体动作序列。与仅使用身体姿态相比，注视特征提高了结果，在预测准确率和及时性方面均有所提升。

人机协作相关方法总结见表3。

4.2.2 共享自主抓取系统：辅助技术与智能遥操作。操作任务中的机器人遥操作旨在将感知与动作分离，把运动动作的影响转移到远程位置或环境中，同时仍利用操作员精准的眼手协调。近年来，针对共享自主系统，尤其是辅助与遥操作场景，已有多种意图估计方法被开发出来。共享自主在执行远程任务时结合机器人与人类控制。随后，双方分别承担自己更擅长的部分：动作执行中较低层次的运动学方面通常交由机器人处理，而任务规划、应对意外事件等较高层次的认知技能，通常由人类同时承担。这可以通过机器人伙伴不同程度的自主性来实现 [Beer 等人 2014；Goodrich 等人 2013年；Schilling 等人 2016]。用户输入控制器（如操纵杆）与机器人执行器之间的自由度或运动学能力往往存在很大差异，因此，共享自主可以简化控制，减轻操作员的认知负荷并加快执行。用户设定目标及其实现方式，因此，机器人必须先推断当前的人类意图（意图预测），才能决定在执行过程中提供多少辅助（仲裁）。

Losey 等人 [2018] 回顾了实体人机交互中共享控制的大量文献，特别是康复领域的相关研究。为了审视任何与人类搭档共同执行任务的系统，他们提出了一种统一的观点，指出任何在身体上辅助人类的系统在设计时都应关注三个方面：

- **意图识别**。应尽早且尽可能自然地识别意图，从而让用户无需明确指挥机器人，并使机器人及时开始辅助动作。已有多种方法采用先预测后行动策略 [例如，Dragan 和 Srinivasa 2013；Yu 等人 2005], 预测一个目标并执行相应计划，但更灵活的解决方案是计算所有可能目标的概率分布，并随着动作的展开持续重新评估预测的意图，从而实现在线纠正和持续调整 [Hauser 2013；Javdani 等人 2015]。

- **仲裁**。 在完成意图检测后，必须设计一种策略来仲裁用户输入与机器人辅助在多大程度上进行融合并执行。这可以采用先预测后融合策略，即 $Arbitration = (1-\alpha)\cdot UserInput - \alpha \cdot RobotPrediction,\ \alpha \in [0,1]$ [例如，参见 Dragan 和 Srinivasa 2013],；也可以采用更平滑的辅助策略，对所有可能的用户目标计算概率分布，并将实现用户目标的代价降至最低 [Javdani 等人 2018, 2015]。

- **通信/反馈**。 在大多数场景中，机器人系统会替代人类手臂的运动功能；有时也会使用其传感器向用户反馈环境的属性。

鉴于这些前提，通过注视跟踪推断操作意图已成为一种极具吸引力的方法，尤其是在辅助技术领域。在此，目标用户是身体

研究	Task	特征	识别 方法	识别的内容
[Admoni 和 Srinivasa 2016]	倒液体	注视分布 在兴趣区上	部分可观测马尔可夫决策过程	抓取目标
[Razin and Feigh 2017]	拾取与 放置	注视位置 和速度	LDA, 决策树, k近邻、隐马尔可 夫模型	动作（'移动'， '准备'，'保持'）
[Webb等人 2016]	抓握 网球	注视位置 和运动指令	人工势场	目标位置
[Li等人 2017]	抓握	三维注视 位置	GMM	抓取目标和 相关接触点
[Li等人 2020]	用于抓取的 操作的抓握	注视与 运动	贝叶斯网络、支持向量机, NN	抓取目标和 后续操作 （转移、使用、 交接）
[Shafti等人 2019]	拾取与放置, 倒液体	物体上的注视停留时间	FSM	原始动作 （伸手、抓取、 倒液体） + 物体
[Jain and Argall 2018, 2019]	拾取与放置	/	贝叶斯滤波	目标物体
[Wang 等人 2020]	制作速溶咖啡， 制作粉末 饮料，准备清洁 海绵	当前感兴趣区域, AOI序列, 注视-物体角度/速度	RNN	动作原语, 目标物体
[Fuchs 与 Belardinelli 2021]	拾取与放置	时间序列 AOI上的注视分布	广义隐马尔 可夫模型	动作和 目标物体

表 4. 共享控制系统中的意图识别方法

有障碍的人或老年人，对他们而言，使用操纵杆控制器进行完全遥操作是一项更大的挑战。复杂的遥操作场景中也在一定程度上存在类似问题：用户需要在杂乱且可能有危险的环境中进行灵巧控制（例如处理核废料或拆弹）。事实上，即使是用户亲自用手控制仿人机械臂的情况，也有一些底层运动学方面可以交由机器人部分处理，而非由用户负责，前提是能够及时预测正确的意图。例如，机器人可以根据当前硬件和目标物体，自动选择最佳抓取方式和抓取位置。

一些研究首先探讨了基于注视的意图估计的理论表述。作为首次尝试整合来自用户的注视输入，Admoni 和 Srinivasa [2016]提出了一种基于Javdani框架的方法[Javdani 等人 2015] 在该方法中，通过在部分可观测马尔可夫决策过程中将用户的眼动和操纵杆指令都作为观测，并采用后见优化实时求解，来更新目标（隐藏状态）的概率分布。假设已知被操纵物体的位置（通过使用基准标记），则给定当前图像时，某个目标的概率被认为与当前注视位置和目标物体之间的距离成正比。不过，这项研究并未提供实验数据或实际实现。

在更常见的共享自主场景中，Jain 和 Argall [2018] 提出了一个在递归贝叶斯滤波框架下进行意图识别的数学表述。假设可用目标的数量有限，则这些目标的概率

分布是根据一系列观测（注视、运动输入命令等）计算得到的后验概率，并取决于马尔可夫模型中的先前目标估计。随后，将最大后验视为当前目标，其置信度取决于最有可能的目标与第二有可能的目标之间的差异。系统提供的辅助程度由目标预测的置信度调节。总体而言，这项研究还表明，确定早期提供辅助与对预测用户目标具有较高置信度之间的恰当权衡十分重要。

在另一项后续研究中，Jain 和 Argall [2019], 在相同的数学表述基础上，通过针对用户定制的理性模型优化进行了扩展，从而考虑了人类动作次优性。不过，在这两项研究中，注视都只是作为系统可能的输入提及，并未在实践中集成和测试。

迄今为止，大多数依赖眼动追踪进行意图识别的共享自主方法都比较新近，而且往往不包含上述三个组成部分。此类系统通常聚焦于第一个组成部分或其子部分，例如注视三维定位，或在共享自主情境下对注视模式的研究。以下研究考察了遥操作期间注视行为的特征，以及如何通过学习注视行为来进行意图估计。

对于一只辅助机器人手臂，它要从盘子中叉起食物块喂给行动不便的用户，Aronson 等人 [2018] 开展了一项初步眼动追踪实验，采用被试内设计，比较用户在不同操作模式下的行为：完全遥操作、完全自主、依据 Javdani 等人 [2015], 的混合自主，以及如Dragan 和 Srinivasa [2013]所述的混合策略自主。研究观察到两种注视模式：监测性扫视，用于检查手臂接近预期食物块时的平移行为；规划性扫视，则在启动手臂执行动作前选定目标食物块，这与自然眼手协调中的情况相同，见 [Hayhoe 等人，2003年；Johansson 等人2001]。然而，研究未发现操作条件对这类注视频率的影响，但观察到一些反复监测性扫视的模式，即注视在执行器与目标食物块之间切换。此外，眼动数据并未真正相对于场景内容进行定位，分析也仅聚焦于注视动态，因此部分结果的结论性受到限制。随后的一项研究进一步探讨了注视未能定位到关注对象以及数据质量可能下降的问题 [Aronson 和 Admoni2019], 采用语义标注加以处理；同一研究团队还提出在操作过程中使用注视跟踪进行错误检测 [Aronson 和 Admoni 2018]。Li 等人研究了 [2017]三维注视定位及其在意图预测中的应用。在此案例中，视觉运动抓取模型还可以预测用户打算如何抓取所选物体。后一项功能目前仅针对一个长方体物体进行了形式化，因此只能推断一个接触点和手的接近方向；具体结果取决于长方体被注视的位置，而模型是在训练阶段通过观察受试者亲自抓取物体来学习的。近期，Wang 等人沿用了Haji Fathaliyan 等人 [2018], 首先提出的方法， [2020] 提出了一个利用四种特征进行动作基元识别的框架：当前注视的物体、注视对象序列，以及作为补充的注视物体角度（注视向量与指向各物体的向量之间的角距离）和注视物体角速度。研究针对三项活动设想了四种动作基元（伸手、移动、放下、操作）：冲泡速溶咖啡、调制粉末饮料、准备清洁用海绵。该系统基于循环神经网络，具有良好的识别准确率和较低的观测延迟，也就是说，平均在观察到动作的10%后即可给出正确预测。尽管该系统旨在用于辅助应用场景中的共享自主，但它是在自然眼手协调数据上训练的，而在实际场景中，这种协调可能会受到视觉输入和机器人执行器的强烈调节与影响。Fuchs 和 Belardinelli [2021] 通过在虚拟现实头戴设备中呈现场景，并让用户

通过自然手臂运动（借助 Vive 控制器进行追踪）遥操作渲染的机器人手。在抓取与放置任务中，用户注视的兴趣区域序列被用于训练两个动作特定隐马尔可夫模型；这些模型能够将其识别准确率推广到新用户、不同数量的物体及不同物体构型。不过，该系统尚未在共享控制场景中进行测试。

与上一节一样，不少研究结合了注视与运动特征，以提高预测能力和鲁棒性。 Razin和Feigh [2017] 测试了多种模态、特征和分类技术，以预测机器人眼手协调过程中的意图。在使用机械爪执行拾取与放置任务的情境中，研究人员利用手部运动、眼睛和注视特征来预测三项任务（“移动”、“准备移动”和“保持”）。在这种情况下，作者发现，运动特征足以预测这三项任务；眼睛和注视特征虽能达到较好的预测水平，但与运动特征结合后并未进一步提升性能，这并不令人意外。决策树、k近邻和隐马尔可夫模型在平均准确率和曲线下面积（AUC）方面表现最佳。不过，这项研究旨在区分的仅是运动动作，并未涉及关注对象。但当用户面对杂乱场景时，与运动轨迹特征相比，注视能提供更精确、更早的信息；起初，运动轨迹特征可能对应多个目标。Aronson 等人Aronson 等人 [2021] 针对使用操纵杆的运动控制进行了这一展示，而 Belardinelli等人 [2022] 则针对虚拟现实中的自然运动控制进行了展示。

迄今为止回顾的研究中，仅预测即时动作和关注对象。在一项较新的研究中，Li等人 [2020] 探讨了除目标物体外，是否还能在伸手运动完成之前识别操作意图，以便机器人为后续任务选择最佳抓取方式。在这项研究中，研究人员利用手部跟踪控制一条配备三指夹爪的机械臂。由于相同的抓取方式可能对应不同的任务（有时，同一任务也会采用不同的抓取方式），因此将整个问题置于多标签分类框架下处理。研究人员通过贝叶斯网络（并以支持向量机和神经网络进行比较）对意图推断进行建模，根据物体以及注视停留其上的位置和抓取配置，计算意图的最大后验似然。实验涉及4个物体和3项任务；训练集采集自人类抓取实验，测试则使用真实抓取序列和遥操作序列。与Razin 和 Feigh [2017], 的研究类似，注视信息似乎能提升遥操作情况下的性能，但在真实抓取情况下提升较小。需要指出的是，该框架尚未实现共享控制。

其他系统已经在共享自主框架中集成了某种形式的意图检测。Webb等人 [2016] 考虑了远程遥操作场景，并采用操纵杆与注视混合控制策略，通过恢复一种更自然的眼手协调，来克服通过屏幕反馈操纵物体时受限的深度感知。系统利用“预视”注视点预测移动目标，并使用高斯势场将这一输入与直接操纵杆指令相结合，形成先预测后融合框架。与仅使用操纵杆的控制方式相比，该策略减少了抓取尝试次数和执行时间，也更受参与者欢迎。然而，在这一系统中，注视被主动用于指示目标。为了区分自主注视转移和非自主注视转移，用户在注视目标物体时必须按下操纵杆上的按钮；同时，抓握也通过按键触发，从而削弱了其关于自然眼手协调的主张。在少数将基于注视的意图识别集成到更复杂的康复辅助框架中的实例中，Shafti等人 [2019]提出了一个实例。该系统依赖三维注视定位、用于目标识别的卷积神经网络、头部跟踪，以及一个有限状态机（FSM）。该有限状态机根据场景中物体预先定义的“可抓取性”和“可倾倒性”来实现基本的动作语法。事实上，除了在

桌面上进行拾取与放置之外，还选择了在容器中进行拾取与放置以及向容器中倾倒作为任务。在这里，用户同样需要主动注视右侧的物体，以发出交互意图信号，而连接在用户手臂上的机械臂则执行相应动作。

此处介绍并在表 4 中总结的研究表明，在辅助人机交互和遥操作中使用注视的兴趣正在迅速增长；但与其他领域相比，这些研究也显示出诸多关键局限和挑战，将在第 4.3 节中进一步讨论。

4.3 推断高级驾驶辅助系统中的驾驶员意图

过去几十年来，人们一直在开发高级驾驶辅助系统（ADAS），以提高驾驶安全性。这些系统不仅像被动安全系统（安全带、安全气囊等）一样，在发生事故时保护乘员，还通过预先降低事故风险来提升安全性 [Koesdwiady 等人 2017]。在自动驾驶在多种场景中普及之前，继续投资于能够在低于 5 级的任意自动化水平下为驾驶员提供支持的系统十分重要。正如我们所见，驾驶是一种相当复杂且持续时间较长的行为，需要多项高级认知能力（持续注意、多任务处理、情境意识、决策、眼手足协调等）始终保持充分运作。即使其中一部分行为已经实现了一定程度的自动化，在某些情况下，认知负荷仍可能非常高，哪怕是极轻微的疲劳或注意力不集中，也可能导致致命后果。在此背景下，人们探索了多种解决方案：一方面，许多研究工作通过注视跟踪评估驾驶员的认知状态，并在其注意力分散时发出警告 [例如，Ahlstrom 等人 2013]；另一方面，尤其是在半自动驾驶（最高达到 3 级）的背景下，研究工作正致力于辅助执行困难的驾驶操作。在后一种情况下，意图估计输出的及时性当然可能至关重要，因为与遥操作相比，ADAS 根据该估计采取行动的时间范围更短，尤其是在考虑车辆速度与人类运动速度的差异时。这还取决于 ADAS 需要如何利用驾驶员意图信息：它可以仅发出警告（将人类反应时间纳入考虑），提供辅助（通过规划并执行支持性计划），或者仅仅避免做出与驾驶员意图相冲突的决策，例如在驾驶员确实有意变道时触发车道偏离预警。 [Xing 等人 2019]。

尽管利用注视推断驾驶意图这一想法可追溯到几十年前 [例如，见Oliver 和 Pentland 2000年，他们通过隐马尔可夫模型识别了多达 7 种驾驶操作], 但在随后的几年里，由于在汽车中进行眼动追踪存在技术难题——光照条件不断变化，场景也在运动——这一领域的意图估计主要采用了其他行为线索。

尤其是，Trivedi 及其同事长期致力于驾驶员意图估计，但他们主要使用头部动态数据，并指出在变道时，这类数据比注视数据更具显著的判别力 [Doshi 和 Trivedi 2009年；McCall 和 Trivedi 2007]。

Liebner 等人 [2013] 结合头部朝向角、车辆速度和转向灯激活情况，推断变道和转弯等意图序列。

聚焦于注视，Lethaus 等人 [2013]比较了神经网络、贝叶斯网络和朴素贝叶斯分类器在区分变道与车道保持（或向右变道与向左变道）方面的表现。通过在驾驶模拟器中收集注视数据，作者同时考察了模型进行预测所需的数据量，以及在驾驶操作开始前多久可以做出预测。模型输入采用了 5 个区域（后视镜、左右后视镜、挡风玻璃和速度计）上的视线时间，时间窗口分别为变道前 10 秒和 5 秒。正如预期，预测时间越接近驾驶操作开始时刻，所有模型的估计就越准确。

事实证明，5 秒的时间窗口能够产生更好的预测模型，因为其中包含的其他活动噪声更少；神经网络的结果略优于贝叶斯网络。Peng 等人 [2015]发现，在结合头部旋转标准差以及其他涉及车辆运动、指示灯状态和变道条件的特征时，同样需要一个 5 秒的时间窗口。

[Wu 等人 2019] 探讨了半自动驾驶车辆中的意图估计。在此情境下，如果系统在跟随一辆开始减速的前车时，无法确定最佳驾驶操作，就应提醒驾驶员接管控制。为了仅凭注视/凝视数据估计驾驶员制动或变道的意图，作者设计了一种结合概率动态时间规整的隐马尔可夫模型，以递归贝叶斯估计方案，根据此前的观测和上一次估计的意图，对每个新观测的概率进行建模。研究表明，该模型对 3 种意图的识别准确率超过 90%，且在驾驶操作实际开始前可提前 3 秒以上进行预期。尽管高级驾驶辅助系统中的大多数意图估计系统都致力于预测变道和制动， [Jiang 等人2018] 考虑了另一种应用：在此，用户被设定为自动驾驶车辆中的乘客，系统则试图在车辆行驶过程中识别其兴趣点，例如商店招牌或用户打算前往的位置，从而可能向其提供相关信息。该方案利用马尔可夫随机场估计动态兴趣点，并表现出较强的鲁棒性，能够应对眼动追踪不准确带来的噪声，以及运动场景和车辆颠簸运动带来的影响。不过，在这种情况下，与 [Huang 和 Mutlu 2016], 一样，系统检索到的并非精确的运动意图，而是关注对象。

本节所介绍的研究表明，尽管意图估计对高级驾驶辅助系统至关重要，但基于注视的意图预测仍处于相当早期的阶段。事实上，在这类预测错误可能造成致命后果的场景中，即便用于警告系统，其应用也仍非常有限。

5 当前局限、问题与尚待解决的挑战

前几节综述的文献表明，关于基于注视的意图估计的科学研究与技术进展已达到相当成熟的水平，足以让人预见其即将应用于技术产品，尤其是认知机器人。尽管眼动追踪功能正日益嵌入各类接口，但利用眼动实现稳健算法的进展却未能达到同样的速度。如前所述，现有应用与解决方案通常适用范围有限，并依赖精心设计的监督训练。我们希望，更深入地理解其认知基础能够有所助益；但与此同时，某些局限也源于尚未解决的问题与挑战：一方面，我们对视觉运动系统如何服务于图式控制的理解仍不充分；另一方面，技术系统如何利用眼动数据也存在难题。此外，尽管其他一些问题和主题或许并非意图预测框架能否正常运行的核心，却可能对设计用户友好且有效的系统至关重要；用户还应能够信任该系统，并因此愿意频繁与之交互。这些问题也确实会影响用户的注视行为，因此应予以考虑。本文将介绍其中一些最相关的问题，以及文献中提出的观点与方法。

5.1 注视的多任务处理《自然》、作为注意代理的注视与双手操作 s

如前所述，我们的大脑在任何时刻都会同时进行多个认知过程。在考虑比抓取或选择物品更复杂的活动时，尤其如此，例如驾驶，或执行由多个步骤组成、且顺序约束较宽松的操作任务。早在Land 等人 [1999] 就已认识到，可以根据四种不同功能（定位、引导、指示和检查）对物体上的注视进行分类；然而，这些

也可能与来自不同子任务的注视点交错出现：因此，即使眼睛正在监控当前执行过程，有些注视也可能提前转向下一个物体以检查其位置，然后再回到当前物体。目前，这种调度仍相当难以捉摸，并且在不同情境、任务和人群之间存在很大差异。

一方面，如果眼睛的注视点能够揭示大量有关信息获取过程的内容，并且通常被视为注意焦点的行为代理指标，那么有时会无端地假设注视区域中的一切都被有意识地察觉。例如，在驾驶时，当我们的内部注意力转向无关的想法，但驾驶和视觉扫描行为仍可按照自动化模式正常进行时，很容易产生时间空白感。另一方面，尤其是在控制多个效应器时，一些监控可能通过周边视觉隐蔽地完成，或者通过有策略地注视某个中间点来实现，因此不会注视相关的兴趣区域。在自然的双手眼协调中，我们通常会更密切地用视觉监控优势效应器，该效应器执行最困难的动作；而对另一只肢体的控制，则依赖隐蔽注意、本体感觉和触觉反馈[Srinivasan 和 Martin 2010]。或者，我们也会以交错的方式在两个目标之间转移视线 [Riek 等 2003]。然而，这类调度同样缺乏充分研究，尤其是在不同人群和不同惯用手之间。就我们所知，目前大多数基于注视的机器人学系统只处理单臂控制，而近期的双手共享控制框架则依赖于直接学习双手的手部运动模式，或者让第二个机械臂自主地与操作员右臂的运动进行协调 [Laghi 等，2018年；Rakita 等 2019]。也就是说，面向多任务处理和多个效应器的基于注视的意图估计仍然是一个尚未解决的问题，尤其是在难以辨别哪些内容曾被隐蔽地感知、哪些内容完全被忽略时。

5.2 对用户注视行为的双向影响

在人机交互中，基于注视的输入设备控制仅在有限范围内得到应用，且大多用于辅助应用场景 [例如Zeng 等人 2020], , 也就是肢体控制能力受损的患者。基于注视的控制器有时确实会带来问题，因为存在“米达斯之触”效应 [Jacob 1991]。事实上，我们并不习惯将眼睛作为效应器来控制，而是以大多自动化的方式用眼睛感知事物，不会有意在物理世界中产生影响。因此，在此类界面中，基于注视的选择可能导致多次误选。系统确实难以区分用于感知的注视和用于执行动作的注视，而这可能使用户做出高度受控且不自然的眼动，进而增加认知负荷。在基于注视的意图估计系统中，这种情况或许不那么明显；但如果用户意识到系统会利用其注视来产生辅助效果，他们可能会改变自然的扫描路径，以主动获得预期效果。因此，扫描模式会偏离无辅助训练期间观察到的行为，进而干扰系统。正如Javdani 等人 [2015], 所述，其基本假设是，用户在执行自己的计划时不会留意系统。然而，如果用户了解到自己的注视会影响机器人运动，就可能主动将注视作为额外的控制输入来引导机器人：这可能是有意为之，也可能只是需要考虑的副作用。例如，在驾驶场景中，如果用户了解到变道时不看后视镜会触发警告，他们可能会主动计划看一眼后视镜，以避免受到干扰，结果反而触发变道辅助。

在遥操作操作场景中，首先需要学习视觉运动任务，即掌握自身运动与远处视觉感觉后果之间的新映射。这意味着在训练达到上限之前，注视模式会随着训练而变化。 Sailer 等人 [2005] 研究了人们在学习使用不熟悉的设备控制屏幕上的光标完成拦截任务时，注视行为如何演变；该设备通过施加等距的

力和扭矩。他们观察到，第一阶段是探索阶段，此时注视大多跟随受控效应器（光标），命中率也相当低。到了第二阶段，眼睛对预期光标位置的预测性增强，同时命中率有所提高。最后，一旦掌握了视觉运动任务，眼睛便会直接跳至目标，无需监视光标运动。虽然这是一项相当陌生的任务，但它或许可以与通过某种输入设备控制机器人肢体相类比，而这种设备与自身的手部运动截然不同。

近期已有研究探讨了控制机械臂时对扫描路径的影响，例如 Aronson 等人 [2018]；Fuchs 和 Belardinelli [2021]；Wang 等人 [2020], 。不过，还需要进一步研究，以深入了解机器人介导的眼手协调可能如何影响健康用户和残障用户协调感知与行动过程。一般而言，这类问题也与用户适应以及对系统的信任程度有关，下一节将对此进行讨论。

5.3 适应、共同适应与对辅助系统的信任

在共享自主中采用先预测后融合方法时，通常会设置一个仲裁参数 α，用于确定机器人辅助相对于用户输入的介入力度。参数值过小可能导致辅助过于谨慎，无法真正提供支持；参数值过大则可能导致辅助过于激进，从而令人厌烦 [Dragan 和 Srinivasa 2013]。当然，辅助程度也应取决于对推断意图的置信度，以免针对错误目标提供过于激进的辅助。为了使仲裁适应用户，并据此找到最佳辅助程度，Oh 等人 [2019] 提出根据用户与系统的交互，通过在线学习来确定最佳仲裁参数。关于一般的人机界面，Gallina 等人 [2015]提出了一种介于目标导向设计（以最大化性能为目标）和以人为中心的设计（以最大化可用性为目标）之间的方法：渐进式共同适应。其理念是，用户在与系统交互的过程中不断适应并提升技能时，系统也应同时监测这些进步并相应地调整参数。正如 Sailer 等人 [2005], 所示，这种学习也会影响注视，因此应将其纳入考量。然而，在大多数人机交互方法中，模型是在目标任务执行过程中通过收集数据来训练的，通常不会根据用户的长期使用和适应情况重新训练。这仍是一个有待解决的问题；一种可能的解决方案是采用在线学习，在数据分布发生变化时调整时间序列预测模型的参数[，例如 Guo 等人 2016]。

尽管 Javdani 等人 [2015] 假设用户对机器人辅助持中立态度，且不会将其纳入自身的控制策略，但最近 Amirshirzad 等人 [2019] 更系统地研究了人类在遥操作与共享控制中的适应方式究竟有何不同。经过初始探索阶段后，与前一种情况相比，后一种情况下的性能显著提升；在此阶段，用户会潜移默化地了解机器人行为并适应它。

最后还需考虑一个方面，正如Haji Fathaliyan 等人 [2018] 和Aronson 和 Admoni[2018], 所指出的，也可以通过注视来评估用户对系统的信任程度、由故障引起的异常或其他问题。事实上，回顾Sailer 等人 [2005], 在遥操作场景中的研究结果，例如，在完成训练且用户熟悉系统后，用户对机器人的注视点会减少，而其注视将更能预测目标物体/位置。实际上，对当前子任务是否成功的简短视觉评估之后，用户会转移注视，以获取下一个子任务所需的信息 [例如参见Fuchs 和 Belardinelli2021]。如果用户一直在监控机器人，以确保机器人完成自己的任务，那么这种情况可能不会发生。另一方面，如果用户信任并使用系统一段时间后，其注视偏离了通常的可预测模式，那么这可能是系统或用户心理状态发生变化的可靠迹象。关于此类假设的

当然，需要通过精心设计的用户研究，并收集客观和主观测量指标，来检验性能、用户满意度和体验。

5.4 能动感与执行速度的权衡：个性化的理由

让用户保有能动感，是大多数人机交互系统的关键所在，即使系统能够透明地与用户共享控制权，并准确读取用户的意图。在半自动驾驶车辆中也是如此：如果人类驾驶员作出决策并执行驾驶操作，系统进行干预可能会严重损害用户对系统的信任。

在 [Pacherie 2008], 将能动感定义为意向性因果关系的感受，它体现为一种主观感知，即先前意图与随后动作相匹配 [Wegner 2005]。这与预测自身动作后果的能力密切相关。如果人工系统干扰了这种预测，并产生了与我们预期不符的动作执行，我们可能会觉得自己并非引发该动作的主体。这有时可能与系统试图提供的协助相冲突。此外，这还会导致眼动模式发生变化，再次变得不那么具有预期性。不过，这种能动感的行为测量指标，尤其是眼动测量指标，一直难以确定 [Schwarz 等人 2018]。

例如，大多数关于共享自主系统的研究都会报告采用客观和主观测量指标的用户研究结果。如果说人机协作通常有助于缩短任务完成时间，那么主观意见呈现的情况就没那么明确了。有时，面对复杂任务，人们更倾向于提高自主程度；而面对其他简单任务，则更倾向于让人类拥有更多控制权。Dragan 和 Srinivasa [2013] 发现，这些偏好仍会与其他因素相互作用，例如仲裁方式（谨慎与激进）以及意图预测准确率（正确与错误）；总的来说，人们在这些偏好上也存在主观差异。

在此背景下，另一些研究得出了不同结果，发现人们更偏好执行速度 [Dragan 和 Srinivasa2013年; Hauser2013] 这表明，参与者与系统互动的方式各不相同，因此有必要对用于建模用户的代价函数进行个性化处理。
Gopinath 等人 [2016] 采纳了这一建议，并指出在辅助情境中，用户的运动能力各不相同（而且这些能力可能会退化，也可能会改善），因此他们可能需要不同程度的辅助。研究提出了一种以最优控制理论为基础的方法，将仲裁参数的调整表述为一个最优控制问题，其中给定用户和机器人的控制信号。用户通过口头指令增加或减少辅助，由此自行完成优化。尽管这让用户还需通过另一种输入模态控制一个额外参数，但这项研究仍为更深入地探讨共享自主系统中的个性化铺平了道路，旨在优化用户满意度与系统效能之间的平衡。

这仍然不只是仲裁的问题，因为扫描路径行为上的个体差异也可能意味着，在意图估计方面同样需要个性化训练 [Wiebel-Herboth 等人 2020]。在拾取与放置这类简单的运动任务中，不同用户的扫描路径模式可能非常相似；而在约束较少、复杂度更高的活动中，眼动行为可能会更明确地反映每个人安排子任务的方式 [Yi 和 Ballard 2009]。

5.5 从三维注视数据中提取更多信息

如前所述，在三维空间中采集高质量的注视数据，并将其定位到现实世界的物体上，仍是一个有待研究的问题（见第 4.2 节）。正如第 2.2 节所述，在某些情况下，也可以根据注视点推断抓取类型和进一步的操作意图

然而，技术系统很少涉及这些方面，因为在没有基准标记的情况下，要识别处于任意姿态和外观的三维物体，同时在线追踪带有头部运动的注视并将其定位到物体的功能部件上，往往在计算上十分困难。这还需要为每个物体定义多个专用 兴趣区域，并将有关预定义物体的任务、抓取和功能知识融入系统。从这个意义上说，基于注视的意图估计若嵌入更先进的认知系统中，便能取得更好的效果；这类系统不仅依赖精准的眼动追踪能力，还具备复杂的场景理解和世界理解能力，正如当前正在开发的社交机器人和辅助机器人所体现的那样。

6 结论

正如 [Vernon 等人所述， 2016], 意图是整合目标与动作的认知桥梁，构成一个前瞻性框架。因此，推断人类意图对于辅助系统和协作系统至关重要。本文综述并总结了基于注视的意图估计方法及其应用，同时梳理了与感知和动作相关的认知科学基础。过去几十年来，眼动追踪技术取得了长足进步，使应用场景从受限且可控的实验室环境扩展到现实世界活动。与此同时，机器学习、机器人学和高级驾驶辅助系统（ADAS）的研究已经发展到一个阶段，使人类与机器能够真正开展整合式协作，并将其应用于日常生活活动、驾驶或人机协作等常见任务。基于注视物体推断人类意图，可以使这种交互更加有效、更加紧密，并让机器人伙伴了解用户想要实现的目标，从而以无缝方式提供协助。在这方面，综述中的研究和框架展现出了令人鼓舞的成果与前景，至少在相对受限的任务和应用中如此。这些研究有力地表明，注视特征对于尽早、准确地推断意图以及评估用户的认知状态至关重要。一方面，正如本文所综述的那样，更有原则地考量目标认知过程及其后续眼动表现，有助于改进此类系统的设计，尤其是选择最合适的特征和预测模型。另一方面，仍需进一步研究并解决一些尚未解决的问题和局限性，才能推动意图估计在认知系统中的广泛应用。

这些技术带来了重要的挑战与机遇，尤其是在物理系统方面。一方面，能够以自然方式远程操作并通过处理低层次运动方面来辅助操作员的机器人，在当前席卷全球的大流行病等危机中将有广泛的应用。同样，能够推断并执行近身空间内动作的辅助系统，可以帮助某些残障人士在其环境中采取行动，并恢复一定程度的独立性。另一方面，基于注视的驾驶辅助系统能够以更智能、更可靠的方式为驾驶员提供支持，还可能为系统学习驾驶员的知觉决策提供基础，并为实现更接近人类的完全自主铺平道路。

参考文献

Daniel Aarno 和 Danica Kragic。2008年。机器人辅助应用中的运动意图识别。《机器人与自主系统》56，8（2008年），692–705。Henny Admoni 和 Brian Scassellati。2017年。人机交互中的社交注视：综述。《人机交互期刊》6，1（2017年），25–63。Henny Admoni 和 Siddhartha Srinivasa。2016年。利用眼睛注视预测用户意图，实现共享自主。载于*2016*年*AAAI* 秋季研讨会系列。C. Ahlstrom、K. Kircher 和 A. Kircher。2013年。基于注视的驾驶员分心警告系统及其对视觉行为的影响。《*IEEE*智能交通系统汇刊》14，2（2013年），965–973。https://doi.org/10.1109/TITS.2013.2247759Negin Amirshirzad、Asiye Kumru 和 Erhan Oztop。2019年。人类对人机共享控制的适应。《*IEEE*人机系统汇刊》49，2（2019年），126–136。

Reuben M Aronson 和 Henny Admoni。2018年。用于人机共享操作中错误检测的注视。载于联合行动基础研讨会，机器人学：科学与系统。Reuben M Aronson 和 Henny Admoni。2019年。面向人机共享操作的语义注视标注。载于第11届ACM眼动追踪研究与应用研讨会论文集。1–9。Reuben M Aronson、NadiaAlmutlak 和 Henny Admoni。2021年。遥操作过程中利用注视推断目标。载于2021年IEEE/RSJ智能机器人与系统国际会议（IROS）。IEEE，7307–7314。Reuben M Aronson、Thiago Santini、Thomas C Kübler、Enkelejda Kasneci、Siddhartha Srinivasa 和 Henny Admoni。2018年。人机共享操作中的眼手行为。载于2018年ACM/IEEE人机交互国际会议论文集。4–13。Thomas Bader、Matthias Vogelgesang 和 Edmund Klaus。2009年。物体操作过程中自然注视行为的多模态整合，用于意图识别。载于2009年多模态界面国际会议论文集（美国马萨诸塞州剑桥）（ICMI-MLMI '09）。计算机协会，纽约，美国纽约州，199–206。Chris L Baker 和 Joshua B Tenenbaum。2014年。运用贝叶斯心智理论对人类计划识别进行建模。《计划、活动与意图识别：理论与实践》（2014年），177–204。Daniel Baldauf 和 Heiner Deubel。2010年。伸手与抓握中的注意力景观。《视觉研究》50，11（2010年），999–1013。Jenay M Beer、Arthur D Fisk 和 Wendy A Rogers。2014年。构建人机交互中机器人自主性等级的框架。《人机交互期刊》3，2（2014年），74–99。Harold Bekkering 和 Sebastiaan FW Neggers。2002年。视觉搜索受到动作意图的调节。《心理科学》13，4（2002年），370–374。Anna Belardinelli、Marissa Barabas、Marc Himmelbach 和 Martin V Butz。2016a。预期性眼动注视揭示工具交互中的工具知识。《实验脑研究》234，8（2016年），2415–2431。Anna Belardinelli、Oliver Herbort 和 Martin V Butz。2015年。物体交互所支持的目标导向注视策略。《视觉研究》106（2015年），47–57。Anna Belardinelli、Anirudh Reddy Kondapally、Dirk Ruiken、Daniel Tanneberg 和 Tomoki Watabe。2022年。基于注视与运动特征的人机共享控制物体操作意图估计。https://doi.org/10.48550/ARXIV.2208.08688 Anna Belardinelli、Madeleine Y Stepper 和 Martin V Butz。2016b。尽在眼中：在执行前规划精确的手部动作。《视觉期刊》16，1（2016年），18–18。Giuseppe Boccignone。2019年。眼动分析与建模的高级统计方法：简明入门。载于《眼动研究》。施普林格，309–405。Jonathan FG Boisvert 和 Neil DB Bruce。2016年。根据眼动预测任务：空间分布、动态特征和图像特征的重要性。《神经计算》207（2016年），653–668。Ali Borji 和 Laurent Itti。2014年。为 Yarbus 辩护：眼动揭示观察者任务。《视觉期刊》14，3（2014年），29–29。Myles Brand。1984年。意图与行动：迈向自然化的行动理论。麻省理工学院出版社（马萨诸塞州）。Michael Bratman 等。1987年。意图、计划与实践理性。第10卷。哈佛大学出版社，剑桥，马萨诸塞州。Anne-Marie Brouwer、Volker H Franz 和 Karl R Gegenfurtner。2009年。抓取物体与观看物体时注视点的差异。《视觉期刊》9，1（2009年），18–18。Daniel N Bub、Michael EJ Masson 和 Terry Lin。2013年。计划中的手部动作特征会影响对可抓握物体的识别。《心理科学》（2013年），0956797612472909。Andreas Bulling、Jamie A Ward、Hans Gellersen 和 Gerhard Troster。2010年。利用眼电图进行活动识别的眼动分析。《IEEE模式分析与机器智能汇刊》33，4（2010年），741–753。Andreas Bulling 和 Thorsten O Zander。2014年。认知感知计算。《IEEE普适计算》13，3（2014年），80–83。Cristiana Cavina-Pratesi 和 Constanze Hesse。2013年。为什么眼睛偏好食指？精确抓握时眼手运动的同步记录。《视觉期刊》13，5（2013年），15–15。Antoine Coutrot、Janet H Hsiao 和 Antoni B Chan。2018年。利用隐马尔可夫模型进行扫描路径建模与分类。《Behavior Research Methods》50，1（2018年），362–379。David Crundall。2005年。驾驶过程中视觉搜索中自上而下与自下而来因素的整合。《眼动引导中的认知过程》（2005年），283–302。Brendan David-John、Candace Peacock、Ting Zhang、T Scott Murdison、Hrvoje Benko 和 Tanya R Jonker。2021年。迈向基于注视的虚拟现实交互意图预测。载于ACM眼动追踪研究与应用研讨会。1–7。Donald Davidson。1980年。《行动与事件论文集：哲学论文第一卷》。第1卷。牛津大学。Daniel Clement Dennett。1989年。《意向立场》。麻省理工学院出版社。Oriane Dermy、François Charpillet 和 Serena Ivaldi。2019年。运用概率运动基元进行多模态意图预测。载于《以人为本的机器人学》。施普林格，181–196。Heiner Deubel、Werner X Schneider 和 Ingo Paprotta。1998年。选择性背侧与腹侧加工：伸手与知觉中存在共同注意机制的证据。《视觉认知》5，1-2（1998年），81–107。A. Doshi 和 M. M. Trivedi。2009年。眼睛注视与头部动态在预测驾驶员变道意图中的作用。《IEEE智能交通系统汇刊》10，3（2009年），453–462。https://doi.org/10.1109/TITS.2009.2026675

Anca D Dragan 和 Siddhartha S Srinivasa。2013年。用于共享控制的策略融合形式体系。《The International Journal of Robotics Research》32，7（2013年），790–805。Nuno Ferreira Duarte、Mirko Raković、Jovica Tasevski、Moreno Ignazio Coco、Aude Billard 和 José Santos-Victor。2018年。动作预期：解读人类与机器人的意图。IEEE Robotics and Automation Letters 3，4（2018年），4132–4139。Wolfgang Einhäuser、Frank Schumann、Stanislavs Bardins、Klaus Bartl、Guido Böning、Erich Schneider 和 Peter König。2007年。自然探索中的人类眼头协调。Network: Computation in Neural Systems 18，3（2007年），267–297。https://doi.org/10.1080/09548980701671094 PMID: 17926195。Nathan J Emery。2000年。眼睛的作用：社会注视的神经行为学、功能与进化。《神经科学与生物行为评论》24，6（2000年），581–604。Alireza Fathi、Yin Li 和 James M Rehg。2012年。学习利用注视识别日常动作。载于《欧洲计算机视觉会议》。施普林格，314–327。Katja Fiehler、Eli Brenner 和 Miriam Spering。2019年。目标导向动作中的预测。《视觉期刊》19，9（2019年），10–10。Karl Friston、Francesco Rigoli、Dimitri Ognibene、Christoph Mathys、Thomas Fitzgerald 和 Giovanni Pezzulo。2015年。主动推断与认知价值。Cognitive neuroscience 6，4（2015年），187–214。Stefan Fuchs 和 Anna Belardinelli。2021年。抓取与放置任务中共享自主的基于注视的意图估计。Frontiers in Neurorobotics 15（2021年），33。https://doi.org/10.3389/fnbot.2021.647930 Paolo Gallina、Nicola Bellotto 和 Massimiliano Di Luca。2015年。人机交互中的渐进式共同适应。载于《2015年第12届国际控制、自动化与机器人信息学会议（ICINCO）》，第2卷。IEEE，362–368。Alvin I. Goldman。1970年。《人类行动理论》。普林斯顿大学出版社。Iván González-Díaz、Jenny Benois-Pineau、Jean-Philippe Domenger、Daniel Cattaert 和 Aymar de Rugy。2019年。用于自我动作预测的感知引导深度神经网络：物体抓取。Pattern Recognition 88（2019年），223–235。Michael A Goodrich、Jacob W Crandall 和 Emilia Barakova。2013年。辅助型人形机器人的遥操作及其拓展。Reviews of Human factors and ergonomics 9，1（2013年），175–226。Deepak Gopinath、Siddarth Jain 和 Brenna D Argall。2016年。辅助机器人技术中共享自主的人在回路优化。IEEE Robotics and Automation Letters 2，1（2016年），247–254。Tian Guo、Zhao Xu、Xin Yao、Haifeng Chen、Karl Aberer 和 Koichi Funaya。2016年。利用循环神经网络进行稳健的在线时间序列预测。载于《2016年IEEE数据科学与高级分析国际会议（DSAA）》。Ieee，816–825。Amin Haji-Abolhassani 和 James J Clark。2014年。逆雅布斯过程：根据眼动模式预测观察者任务。《视觉研究》103（2014年），127–142。Alireza Haji Fathaliyan、Xiaoyu Wang 和 Veronica J Santos。2018年。利用三维注视追踪识别双手操作过程中的动作，以增强人机协作。Frontiers in Robotics and AI 5（2018年），25。Kris Hauser。2013年。自由形式任务辅助遥操作中的识别、预测与规划。Autonomous Robots 35，4（2013年），241–254。Mary Hayhoe 和 Dana Ballard。2014年。眼动任务控制建模。Current Biology 24，13（2014年），R622–R628。Mary M Hayhoe。2017年。视觉与行动。Annual review of vision science 3（2017年），389–413。Mary M Hayhoe、Anurag Shrivastava、Ryan Mruczek 和 Jeff B Pelz。2003年。自然任务中的视觉记忆与运动规划。《视觉期刊》3，1（2003年），6–6。John M Henderson。2017年。作为预测的注视控制。Trends in cognitive sciences 21，1（2017年），15–23。John M Henderson、Svetlana V Shinkareva、Jing Wang、Steven G Luke 和 Jenn Olejarczyk。2013年。根据眼动预测认知状态。PloS one 8，5（2013年），e64937。Kenneth Holmqvist、Marcus Nyström、Richard Andersson、Richard Dewhurst、Halszka Jarodzka 和 Joost Van de Weijer。2011年。《眼动追踪：方法与测量综合指南》。牛津大学出版社。Chien-Ming Huang、Sean Andrist、Allison Sauppé 和 Bilge Mutlu。2015年。利用注视模式预测协作中的任务意图。Frontiers in psychology 6（2015年），1049。Chien-Ming Huang 和 Bilge Mutlu。2016年。用于高效人机协作的预期式机器人控制。载于《2016年第11届ACM/IEEE人机交互国际会议（HRI）》。IEEE，83–90。Laurent Itti 和 Pierre Baldi。2009年。贝叶斯惊奇度吸引人类注意力。《视觉研究》49，10（2009年），1295–1306。Robert JK Jacob。1991年。眼动在人机交互技术中的应用：所见即所得。ACM信息系统汇刊（TOIS）9，2（1991年），152–169。Siddarth Jain 和 Brenna Argall。2018年。共享控制机器人技术中的递归贝叶斯人类意图识别。载于《2018年IEEE/RSJ智能机器人与系统国际会议（IROS）》。IEEE，3905–3912。Siddarth Jain 和 Brenna Argall。2019年。辅助机器人技术中共享自主的概率式人类意图识别。ACM人机交互汇刊（THRI）9，1（2019年），1–23。Shervin Javdani、Henny Admoni、Stefania Pellegrinelli、Siddhartha S Srinivasa 和 J Andrew Bagnell。2018年。通过后见优化实现遥操作与团队协作中的共享自主。《The International Journal of Robotics Research》37，7（2018年），717–742。Shervin Javdani、Siddhartha S Srinivasa 和 J Andrew Bagnell。2015年。通过后见优化实现共享自主。载于《机器人科学与系统：在线论文集》，第2015卷。美国国立卫生研究院公共获取。

Yu-Sian Jiang、Garrett Warnell 和 Peter Stone。2018年。利用车辆中的注视推断用户意图。载于《第20届ACM多模态交互国际会议论文集》。298–306。Roland S Johansson、Göran Westling、Anders Bäckström 和 J Randall Flanagan。2001年。物体操作中的眼手协调。《神经科学杂志》21，17（2001年），6917–6932。Christopher Kanan、Nicholas A Ray、Dina NF Bseiso、Janet H Hsiao 和 Garrison W Cottrell。2014年。利用多次注视模式分析预测观察者任务。载于《眼动追踪研究与应用研讨会论文集》。287–290。Ashima Keshava、Anete Aumeistere、Krzysztof Izdebski 和 Peter Konig。2020年。虚拟现实中的眼动行为任务解码。载于《ACM眼动追踪研究与应用研讨会》（德国斯图加特）（ETRA '20短篇论文）。计算机协会，纽约，美国，文章30，共5页。Kris M Kitani、Takahiro Okabe、Yoichi Sato 和 Akihiro Sugimoto。2011年。针对第一人称体育视频的快速无监督自我动作学习。载于《2011年计算机视觉与模式识别会议（CVPR）》。IEEE，3241–3248。A. Koesdwiady、R. Soua、F. Karray 和 M. S. Kamel。2017年。驾驶员安全监测系统的最新趋势：现状与挑战。IEEE车辆技术汇刊66，6（2017年），4550–4563。https://doi.org/10.1109/TVT.2016.2631604 Yu Kong 和 Yun Fu。2018年。人体动作识别与预测：综述。arXiv预印本 arXiv:1806.11230（2018年）。Iuliia Kotseruba 和 John K Tsotsos。2021年。驾驶员注意力的行为研究与实用模型。arXiv预印本 arXiv:2104.05677（2021年）。Marco Laghi、Michele Maimeri、Mathieu Marchand、Clara Leparoux、Manuel Catalano、Arash Ajoudani 和 Antonio Bicchi。2018年。用于直观双手遥操作的共享自主控制。载于《2018年IEEE-RAS第18届人形机器人会议（Humanoids）》。IEEE，1–9。Michael Land。2009年。视觉、眼动与自然行为。《视觉神经科学》26，01（2009年），51–62。Michael Land、Neil Mennie 和 Jennifer Rusted。1999年。视觉和眼动在日常生活活动控制中的作用。《知觉》28，11（1999年），1311–1328。Michael Land 和 Benjamin Tatler。2009年。《观看与行动：自然行为中的视觉与眼动》。牛津大学出版社。Michael F Land。2006年。眼动与日常生活中的动作控制。《视网膜与眼科学进展》25，3（2006年），296–324。Michael F Land 和 David N Lee。1994年。转向时我们看向何处。《自然》369，6483（1994年），742–744。Gabor Lengyal、Kevin Carlberg、Majed Samad 和 Tanya R Jonker。2021年。利用眼动注视动态中的隐藏结构预测视觉注意。载于《CHI2021：作为认知状态接口的眼动（EMICS）研讨会论文集》。ACM。Firas Lethaus、Martin RK Baumann、Frank Köster 和 Karsten Lemmer。2013年。基于注视数据预测驾驶员意图的若干简单监督学习算法比较。《神经计算》121（2013年），108–130。Songpo Li、Michael Bowman、Hamed Nobarani 和 Xiaoli Zhang。2020年。遥操作中机器人辅助的操作意图推断。《智能与机器人系统杂志》（2020年），1–15。Songpo Li、Xiaoli Zhang 和 Jeremy D Webb。2017年。面向运动障碍人群，通过模仿人类视觉运动功能实现基于三维注视的机器人抓取。《IEEE生物医学工程汇刊》64，12（2017年），2824–2835。Martin Liebner、Christian Ruhhammer、Felix Klanner 和 Christoph Stiller。2013年。基于参数模型的通用驾驶员意图推断。载于《第16届IEEE智能交通系统国际会议（ITSC 2013）》。IEEE，268–275。Meng Liu、Youfu Li 和 Hai Liu。2020年。带自动校准方法的头戴式眼动追踪系统的三维注视估计。《IEEE开放获取》8（2020年），104207–104215。Dylan P Losey、Craig G McDonald、Edoardo Battaglia 和 Marcia K O'Malley。2018年。实体人机交互中共享控制的意图检测、仲裁与沟通方面综述。《应用力学评论》70，1（2018年）。Franck Mars。2008年。通过操控眼动与转向协调来驾驶弯道。《视觉期刊》8，11（2008年），10–10。Joel C McCall 和 Mohan M Trivedi。2007年。智能车辆的驾驶员行为与情境感知制动辅助。《IEEE会刊》95，2（2007年），374–387。Alfred R. Mele。1992年。《行动的源泉：理解意向行为》。牛津大学出版社。RC Miall 和 GZ Reckess。2002年。小脑与协调的眼手追踪时序。《大脑与认知》48，1（2002年），212–226。Wookhee Min、Bradford Mott、Jonathan Rowe、Robert Taylor、Eric Wiebe、Kristy Elizabeth Boyer 和 James Lester。2017年。开放世界数字游戏中的多模态目标识别。载于《第十三届人工智能与交互式数字娱乐会议》。Thi-Hoa-Cuc Nguyen、Jean-Christophe Nebel、Francisco Florez-Revuelta 等。2016年。利用自我中心视觉识别日常生活活动：综述。《传感器》16，1（2016年），72。Keisuke Ogaki、Kris M Kitani、Yusuke Sugano 和 Yoichi Sato。2012年。结合眼动与自运动特征进行第一人称活动识别。载于《2012年IEEE计算机学会计算机视觉与模式识别会议研讨会》。IEEE，1–7。Yoojin Oh、Marc Toussaint 和 Jim Mainprice。2019年。通过事后数据聚合学习共享自主的仲裁机制。arXiv预印本 arXiv:1906.12280（2019年）。N. Oliver 和 A. P. Pentland。2000年。用于SmartCar驾驶员行为识别的图模型。载于《2000年IEEE智能车辆研讨会论文集》（目录号00TH8511）。7–12。https://doi.org/10.1109/IVS.2000.898310 Elisabeth Pacherie。2008年。行动现象学：概念框架。《认知》107，1（2008年），179–217。Elisabeth Pacherie。2012年。行动。载于《剑桥认知科学手册》，Keith Frankish 和 William Editors Ramsey 编。剑桥大学出版社。Jeff B Pelz 和 Roxanne Canosa。2001年。复杂任务中的眼动行为与知觉策略。《视觉研究》41，25-26（2001年），3587–3596。

Jinshuan Peng、Yingshi Guo、Rui Fu、Wei Yuan 和 Chang Wang。2015年。利用神经网络模型预测驾驶员的变道行为。《应用人因工程》50（2015年），207–217。Daniel Rakita、Bilge Mutlu、Michael Gleicher 和 Laura M Hiatt。2019年。基于共享控制的双手机器人操作。《科学机器人》4，30（2019年）。Harish Chaandar Ravichandar、Avnish Kumar 和 Ashwin Dani。2018年。融合注视与运动信息以推断人类意图。《智能机器人与应用国际期刊》2，2（2018年），136–148。Yosef Razin 和 Karen Feigh。2017年。学习在机器人手眼协调过程中根据注视预测意图。载于第三十一届人工智能促进协会人工智能会议。Stephan Riek、James R Tresilian、M Mon-Williams、Vanessa L Coppard 和 Richard G Carson。2003年。双手瞄准与显性注意：双手遵循同一规律。《实验脑研究》153，1（2003年），59–75。Giacomo Rizzolatti、Lucia Riggio、Isabella Dascola 和 Carlo Umiltá。1987年。注意在水平与垂直子午线之间的重新定向：支持注意前运动理论的证据。《神经心理学》25，1（1987年），31–40。Agnes Rosner、Thomas Franke、Frederik Platten 和 Christiane Attig。2019年。车辆控制中的眼动。载于《眼动研究》。施普林格，929–969。Uta Sailer、J Randall Flanagan 和 Roland S Johansson。2005年。学习新视觉运动任务期间的眼手协调。《神经科学杂志》25，39（2005年），8833–8842。Kenji Sakita、Koichi Ogawara、Shinji Murakami、Kentaro Kawamura 和 Katsushi Ikeuchi。2004年。通过解读人类注视信息实现人机灵活协作。载于*2004*年*IEEE/RSJ*智能机器人与系统国际会议，仙台，日本，*2004*年*9*月*28*日至*10*月*2*日。IEEE，846–851。Dario D Salvucci 和 Joseph H Goldberg。2000年。在眼动追踪规程中识别注视与扫视。载于*2000*年眼动追踪研究与应用研讨会论文集。71–78。Dario D Salvucci 和 Andrew Liu。2002年。变道的时间进程：驾驶员控制与眼动行为。《交通研究*F*辑：交通心理学与行为》5，2（2002年），123–132。Malte Schilling、Stefan Kopp、Sven Wachsmuth、Britta Wrede、Helge Ritter、Thomas Brox、Bernhard Nebel 和 Wolfram Burgard。2016年。迈向共享自主的多维视角。载于*2016*年*AAAI* 秋季研讨会系列。Katharina A Schwarz、Sebastian Burger、David Dignath、Wilfried Kunde 和 Roland Pfister。2018年。动作-效应绑定与能动性。《意识与认知》65（2018年），304–309。Paul Schydlo、Mirko Rakovic、Lorenzo Jamone 和 José Santos-Victor。2018年。人机协作中的预期：一种用于预测多个动作序列的循环神经网络方法。载于*2018*年*IEEE*机器人与自动化国际会议（*ICRA*）。IEEE，1–6。John R. Searle。1983年。《意向性：心灵哲学论文集》。剑桥大学出版社。Ali Shafti、Pavel Orlov 和 A Aldo Faisal。2019年。用于辅助伸手与抓取的基于注视的情境感知机器人系统。载于*2019*年机器人与自动化国际会议（*ICRA*）。IEEE，863–869。Ronal Singh、Tim Miller、Joshua Newn、Eduardo Velloso、Frank Vetere 和 Liz Sonenberg。2020年。结合注视与人工智能规划，实现在线人类意图识别。《人工智能》（2020年），103275。Divya Srinivasan 和 Bernard J Martin。对称双手伸手任务中的眼手协调：时间特征。《实验脑研究》203，2（2010年），391–405。Namrata Srivastava、Joshua Newn 和 Eduardo Velloso。2018年。结合低层与中层注视特征进行桌面活动识别。《*ACM*交互式、移动、可穿戴与泛在技术汇刊》2，4（2018年），1–27。Bernard Marius 't Hart 和 Wolfgang Einhäuser。2012年。注意脚下：隐性任务对现实情境中注视分配时眼动与头部运动的互补影响。《实验脑研究》223（2012年），233–249。Enkelejda Tafaj、Gjergji Kasneci、Wolfgang Rosenstiel 和 Martin Bogdan。2012年。眼动数据的贝叶斯在线聚类。载于眼动追踪研究与应用研讨会论文集。285–288。Karim A Tahboub。2006年。基于动态贝叶斯网络概率意图识别的智能人机交互。《智能与机器人系统杂志》45，1（2006年），31–52。Ryo Takahashi、Hiromasa Suzuki、Jouh Yeong Chew、Yutaka Ohtake、Yukie Nagai 和 Koichi Ohtomi。2018年。使用眼动追踪眼镜进行三维注视点分析的系统。《计算设计与工程杂志》5，4（2018年），449–457。Ajay Kumar Tanwani 和 Sylvain Calinon。2017年。用于遥操作中意图识别与操作辅助的生成模型。载于*2017*年*IEEE/RSJ*智能机器人与系统国际会议（*IROS*）。IEEE，43–50。Michael Tomasello、Malinda Carpenter、Josep Call、Tanya Behne 和 Henrike Moll。2005年。理解并共享意图：文化认知的起源。《行为与脑科学》28，5（2005年），675–691。John K Tsotsos、Iuliia Kotseruba、Alexander Andreopoulos 和 Yulong Wu。2019年。数据驱动方法优于理论驱动的计算机视觉的一个可能原因。*arXiv* 预印本 *arXiv:1908.10933*（2019年）。Ralf Van der Lans、Michel Wedel 和 Rik Pieters。2011年。跨个体与任务界定眼注视序列：双眼个体阈值（BIT）算法。*Behavior Research Methods*43，1（2011年），239–257。David Vernon、Serge Thill 和 Tom Ziemke。2016年。意图在认知机器人学中的作用。载于《迈向具有社会可信行为的机器人系统》第一卷。施普林格，15–27。

Xiaoyu Wang、Alireza Haji Fathaliyan 和 Veronica J Santos。2020年。迈向适用于人机系统的共享自主控制方案：利用眼动注视特征进行动作基元识别。*Frontiers in Neurorobotics* 14（2020年），66。Jeremy D Webb、Songpo Li 和 Xiaoli Zhang。2016年。利用视觉运动倾向提升遥操作的控制性能。载于*2016*年美国控制会议（*ACC*）。IEEE，7110–7116。Sascha Weber、Rebekka S Schubert、Stefan Vogt、Boris M Velichkovsky 和 Sebastian Pannasch。2018年。Gaze3DFix：利用椭球边界体检测三维注视点。*Behavior Research Methods* 50，5（2018年），2004–2015。Daniel M. Wegner。2005年。谁是受控过程的控制者？载于*TheNew Unconscious. Oxford Seriesin Social Cognition andSocial Neuroscience*，Ran R. Hassin、James S. Uleman 和 John A. Bargh（编）。牛津大学出版社，19–36。Julia Werneke 和 Mark Vollrath。2012年。驾驶员在看什么？交叉口特征对注意分配和驾驶行为的影响。*AccidentAnalysis &Prevention* 45（2012年），610–619。Christiane B. Wiebel-Herboth、Matti Krüger 和 Martina Hasenjäger。2020年。个体间与个体内效应对注视行为的交互影响。载于第*28*届*ACM*用户建模、自适应与个性化会议附属出版物（意大利热那亚）（*UMAP '20* 附属会议）。计算机协会，纽约，美国，35–40。M. Wu、T. Louw、M. Lahijanian、W. Ruan、X. Huang、N. Merat 和 M. Kwiatkowska。2019年。半自动驾驶车辆驾驶操作中的基于注视的意图预测。载于*2019*年*IEEE/RSJ*智能机器人与系统国际会议（*IROS*）。6210–6216。https://doi.org/10.1109/IROS40897.2019.8967779Y. Xing、C. Lv、H. Wang、H. Wang、Y. Ai、D. Cao、E. Velenis 和 F. Wang。2019年。智能车辆驾驶员变道意图推断：框架、综述与挑战。*IEEE*车辆技术汇刊68，5（2019年），4377–4390。https://doi.org/10.1109/TVT.2019.2903299Alfred L Yarbus。1967年。感知复杂物体时的眼动。载于《眼球运动与视觉》。施普林格，波士顿，马萨诸塞州，171–211。Weilie Yi 和 Dana Ballard。2009年。识别手眼协调模式中的行为。国际仿人机器人学杂志6，03（2009年），337–359。Wentao Yu、Redwan Alqasemi、Rajiv Dubey 和 Norali Pernalete。2005年。基于运动意图识别的遥操作辅助。载于*2005*年*IEEE*机器人与自动化国际会议论文集。IEEE，1121–1126。Raimondas Zemblys、Diederick C Niehorster、Oleg Komogortsev 和 Kenneth Holmqvist。2018年。利用机器学习检测眼动追踪数据中的事件。*Behavior Research Methods* 50，1（2018年），160–181。Hong Zeng、Yitao Shen、Xuhui Hu、Aiguo Song、Baoguo Xu、Huijun Li、Yanxin Wang 和 Pengcheng Wen。2020年。利用混合式眼动—脑机接口进行半自主机械臂伸达。*Frontiers in Neurorobotics* 13（2020年），111。Ruohan Zhang、Akanksha Saran、Bo Liu、Yifeng Zhu、Sihang Guo、Scott Niekum、Dana Ballard 和 Mary Hayhoe。2020年。人类注视辅助人工智能：综述。载于*IJCAI*：会议论文集，第2020卷。美国国立卫生研究院公共获取，4951。